

BUKU PEGANGAN IMPLEMENTASI:

Artificial Intelligence, Machine Learning dan Big Data

UNTUK KOPERASI DAN UMKM

Onno W. Purbo

2024

**Institut Teknologi Tangerang Selatan
Kementrian Koperasi dan UMKM**

Buku Pegangan Implementasi Artificial Intelligence, Machine Learning dan Big Data: Untuk Koperasi dan UMKM

Onno W. Purbo

onno@indo.net.id

X: @onnowpurbo

Institut Teknologi Tangerang Selatan (ITTS)
Kementerian Koperasi dan Usaha Kecil Menengah (KemenkopUKM)

DAFTAR ISI

- DAFTAR ISI..... 2**
- KATA PENGANTAR..... 7**
- BAGIAN 1: PENDAHULUAN..... 8**
 - BAB 1: Kecerdasan Buatan untuk Indonesia..... 8
 - Strategi Nasional Kecerdasan Buatan Indonesia..... 8
 - Kunci Keberhasilan AI di Indonesia..... 9
 - Hambatan Mindset SDM..... 9
 - Tujuan Buku..... 10
 - Gambaran Umum Buku..... 11
 - BAB 2: Sejarah dan Definisi Singkat..... 16
 - Sejarah singkat AI, Machine Learning dan Big Data..... 16
 - Definisi singkat AI, Machine Learning dan Big Data..... 17
 - Manfaat dan Potensi AI..... 19
 - Tantangan dan Risiko AI..... 21
 - BAB 3: AI untuk Koperasi & UKM..... 22
 - Efisiensi dan Efektivitas Layanan..... 22
 - Pengambilan Keputusan berdasarkan Data..... 23
 - Transparansi dan Akuntabilitas..... 24
 - Memperkuat demokrasi dan partisipasi publik..... 24
- BAGIAN 2: POLA PENERAPAN AI, MACHINE LEARNING & BIG DATA. 26**
 - BAB 4: Pola AI dan Transformasi Digital..... 26
 - Transformasi Digital dan AI..... 26
 - Contoh Penerapan AI di Dunia Nyata..... 28
 - Persamaan Linier: Cara Berpikir Sederhana AI..... 33
 - Pengelolaan data dan informasi..... 38
 - Pengambilan keputusan dan kebijakan..... 40
 - Tata kelola dan akuntabilitas..... 42
 - BAB 5: Aplikasi Orange Data Mining..... 43
 - Instalasi Orange Data Mining..... 43
 - Menjalankan Orange Data Mining..... 46
 - Contoh Penggunaan Orange..... 51
 - BAB 6: Contoh Prediksi Sederhana..... 58
 - Coba Prediksi Anggaran..... 58

Coba Prediksi Anggaran dengan Deviasi.....	66
Coba Prediksi Anggaran dengan Multi Fitur.....	68
BAB 7: Explorasi Contoh Bawaan Orange.....	85
Menganalisa Model dari Data.....	85
Contoh Pengambilan Keputusan Pembelian Mobil.....	90
Contoh Klasifikasi Binatang.....	98
BAB 8: Prediksi Data & Situasi Kompleks.....	113
Pola pikir prediksi.....	113
Contoh Sederhana Klasifikasi Kemiskinan.....	114
Contoh Sederhana Klasifikasi Stunting.....	123
Contoh Klasifikasi Stunting dengan Korelasi Kemiskinan.....	126
Contoh Prediksi Program Pengentasan Kemiskinan.....	128
Evaluasi fitur yang dominan mengubah target.....	128
Logika berfikir untuk prediksi.....	139
Percobaan Menaikan Pendapatan 10%.....	146
Percobaan Menaikan Pendapatan 10 kali lipat.....	147
Percobaan menambahkan lama pendidikan 6 tahun.....	149
Percobaan menambahkan 1 rumah.....	150
Contoh Prediksi Program Peningkatan Penghasilan.....	151
Contoh Sederhana Prediksi Cuaca.....	158
Contoh Analisa untuk SATU Indonesia Awards.....	163
BAB 9: Analisis Data Koperasi dan UKM.....	168
Data Koperasi dan UMKM.....	168
Orange Workflows untuk analisa.....	169
Analisa OMZET.....	171
Analisa Metoda Pemasaran.....	177
BAB 10: AI dan Digital Marketing.....	182
Penerapan AI dalam Berbagai Aspek Digital Marketing.....	182
Keuntungan dan Tantangan Penggunaan AI dalam Digital Marketing.....	186
BAB 11: Process Mining untuk Transformasi Digital.....	189
Peran Process Mining dalam Transformasi Digital.....	190
Langkah Implementasi Process Mining.....	191
Tool dan Teknologi Process Mining.....	193
Contoh Implementasi Process Mining.....	195
PM4PY: Penyiapan data.....	196

PM4PY: Instalasi.....	198
PM4PY: Contoh Source Code.....	198
BAB 12: ChatGPT dan Generative AI.....	203
Ollama Untuk Generative AI Lokal.....	204
Ollama: Instalasi dan penggunaan yang mudah.....	205
Retrieval Augmented Generation (RAG) memperkaya Ollama lokal	208
BAB 13: Sepintas Tensorflow untuk Neural Network.....	211
Sepintas Machine Learning dengan Neural Network.....	211
Neural Network sebagai Ekspansi Regresi Linier.....	214
Contoh Neural Network Ekspansi dari Regresi Linier dalam Python..	215
Pengantar Tensorflow untuk Pemula.....	218
Set up TensorFlow.....	218
Load dataset.....	219
Built machine learning model.....	219
Loss Function.....	221
Train & evaluate model.....	222
BAGIAN 3: TANTANGAN DAN RISIKO IMPLEMENTASI AI, MACHINE	
LEARNING DAN BIG DATA.....	224
BAB 14: Etika dan Privasi Data.....	224
Perlindungan Data Pribadi dan Hak Individu.....	224
Algoritma yang Adil dan Transparan.....	224
Akuntabilitas dan Tanggung Jawab.....	225
Implikasi dan Tantangan.....	225
BAB 15: Kesenjangan Digital dan Keterampilan.....	227
BAB 16: Kerangka Kerja dan Regulasi.....	229
Strategi Nasional Kecerdasan Buatan (AI) untuk Koperasi dan	
UMKM.....	229
Tata kelola dan standar data.....	230
Keamanan siber dan ketahanan data.....	231
Ancaman Keamanan Siber pada Operasional AI.....	231
Peran Pemerintah dalam Memproteksi Data Koperasi dan	
UMKM.....	232
BAGIAN 4: MASA DEPAN AI, MACHINE LEARNING DAN BIG DATA.....	
235	
BAB 17: Strategi Big Data Efektif dan Hemat Biaya:.....	235

BAB 18: Trend dan Perkembangan Terbaru.....	237
Dari Laboratorium ke Dunia Nyata: Penerapan AI di Indonesia..	237
Indonesia di Era Big Data: Tantangan dan Peluang.....	238
Indonesia Digital: Tantangan dan Solusi dalam Era Transformasi Teknologi.....	239
Transformasi Digital: Integrasi Blockchain dan Process Mining..	240
BAB 19: Rekomendasi dan Strategi untuk Implementasi yang Efektif....	242
Membangun kepemimpinan dan komitmen.....	242
Implementasi Strategi yang Konkrit.....	242
Potensi Besar AI untuk Koperasi dan UMKM.....	243
Mengembangkan infrastruktur dan kapasitas.....	243
Memastikan etika dan privasi data.....	245
Peraturan Privasi Data.....	245
Mengapa Etika dan Privasi Penting bagi Koperasi dan UMKM?.	245
Prinsip-Prinsip Etika AI untuk Koperasi dan UMKM.....	246
Langkah-langkah Praktis.....	246
Manfaat Mengimplementasikan AI yang Bertanggung Jawab....	246
Kolaborasi AI untuk Memperkuat Koperasi dan UMKM.....	247
BAGIAN 5: SIAP LEPAS LANDAS.....	250
BAB 20: Ringkasan dan Pesan Utama.....	250
Potensi AI/ML/Big Data untuk Koperasi dan UMKM.....	250
Menjinakkan AI: Strategi Mengatasi Tantangan dan Risiko Implementasi.....	251
Dari Konsep ke Kenyataan: Strategi Implementasi AI.....	252
Dilema Digital: Konflik Antara Antusiasme dan Realitas dalam Implementasi AI.....	253
REFERENSI:.....	255
LAMPIRAN: MSE, RMSE, MAPE, MAR, dan R2.....	256
Mean Squared Error (MSE).....	256
Root Mean Squared Error (RMSE).....	256
Mean Absolute Error (MAR).....	256
Mean Absolute Percentage Error (MAPE).....	257
Coefficient of Determination (R2).....	257
Perbedaan Utama:.....	257

Pemilihan Metrik:.....	257
Tentang Penulis.....	259
Sinopsis.....	260

KATA PENGANTAR

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada BPKP, BPK, Dirjen Anggaran KEMENKEU, Dirjen Perimbangan Keuangan KEMENKEU, KS Wapres, MABES TNI, Kemenkop UKM dan seluruh institusi yang telah memberikan kesempatan kepada penulis untuk berbagi pengetahuan. Buku ini merupakan dokumentasi dari berbagai workshop, kuliah, maupun penelitian yang pernah penulis jalankan.

Melalui buku ini, penulis berharap dapat berkontribusi dalam pengembangan sumber daya manusia Indonesia, khususnya di bidang teknologi artificial intelligence (AI) dan machine learning (ML) yang sedang berkembang pesat.

Dengan memanfaatkan tool open source, buku ini diharapkan dapat mendorong lebih banyak lagi masyarakat Indonesia untuk terlibat dalam pengembangan dan pemanfaatan teknologi AI. Penulis yakin bahwa dengan kolaborasi yang kuat, kita dapat membangun Indonesia yang lebih maju dan berdaya saing melalui pemanfaatan teknologi AI secara optimal.

Jakarta, November 2024
Penulis

BAGIAN 1: PENDAHULUAN

BAB 1: Kecerdasan Buatan untuk Indonesia

Strategi Nasional Kecerdasan Buatan Indonesia

Pandemi COVID-19 telah menjadi katalisator bagi transformasi digital yang pesat di seluruh dunia. Digitalisasi telah merombak berbagai sektor, termasuk bisnis. Di tengah disrupsi ini, banyak usaha besar yang kesulitan beradaptasi, sementara pelaku usaha kecil dan menengah (UKM) justru menemukan peluang baru. Dengan memanfaatkan platform digital dan media sosial, UKM dapat menjangkau pasar yang lebih luas dan bersaing dengan pemain besar. Transformasi digital ini didukung oleh perkembangan teknologi kecerdasan buatan (AI) yang semakin canggih.

Pemerintah Indonesia telah merumuskan Strategi Nasional Kecerdasan Buatan (Stranas KA) sebagai upaya untuk mendorong pengembangan dan pemanfaatan AI di berbagai sektor, termasuk UMKM. Strategi nasional kecerdasan buatan (Stranas KA) dapat di download dari:

<https://ai-innovation.id/images/gallery/ebook/stranas-ka.pdf>

Stranas KA bertujuan untuk menciptakan ekosistem yang kondusif bagi pertumbuhan bisnis berbasis AI, serta meningkatkan daya saing produk dalam negeri. Melalui Stranas KA, diharapkan UMKM dapat memanfaatkan teknologi AI untuk meningkatkan efisiensi operasional, mengembangkan produk baru, dan menjangkau pasar global. Beberapa sektor prioritas yang menjadi fokus Stranas KA, seperti pertanian, manufaktur, dan pariwisata, memiliki potensi besar untuk memberdayakan UMKM.

Peningkatan minat dan aktivitas di bidang AI di Indonesia membuka peluang besar bagi UMKM untuk memanfaatkan teknologi ini. AI dapat membantu UMKM dalam berbagai aspek, mulai dari pemasaran dan penjualan hingga pengelolaan rantai pasok. Misalnya, AI dapat digunakan untuk menganalisis data pelanggan, mempersonalisasi pemasaran, dan mengoptimalkan produksi. Selain itu, AI juga dapat membantu UMKM

dalam mengakses pembiayaan dan mengembangkan produk baru. Dengan dukungan pemerintah, akademisi, dan industri, diharapkan UMKM di Indonesia dapat memanfaatkan potensi AI secara maksimal untuk tumbuh dan berkembang.

Kunci Keberhasilan AI di Indonesia

Implementasi kecerdasan buatan (AI) di Indonesia, khususnya pada sektor koperasi dan UMKM, membutuhkan **sumber daya manusia (SDM)** yang kompeten. Tidak hanya para ahli data dan programmer, namun juga para pelaku bisnis di koperasi dan UMKM perlu memiliki pemahaman dasar tentang AI. Ini penting agar mereka dapat mengidentifikasi peluang penggunaan AI dalam bisnis mereka, seperti dalam analisis data pelanggan, optimasi produksi, atau pengembangan produk baru. Selain itu, pemerintah perlu menyediakan pelatihan dan edukasi yang memadai agar SDM di sektor ini dapat meningkatkan kapasitasnya dalam memanfaatkan teknologi AI.

Ketersediaan data yang berkualitas merupakan tantangan besar dalam pengembangan AI di koperasi dan UMKM. Banyak koperasi dan UMKM di Indonesia yang belum memiliki sistem pencatatan data yang baik, sehingga data yang ada seringkali tidak terstruktur dan sulit diolah. Padahal, data yang berkualitas sangat penting untuk melatih model AI agar dapat memberikan hasil yang akurat. Pemerintah perlu mendorong koperasi dan UMKM untuk melakukan digitalisasi dan mengintegrasikan sistem pencatatan data mereka. Selain itu, perlu dibangun infrastruktur data yang memadai untuk mendukung pengembangan AI di sektor ini. Dengan demikian, koperasi dan UMKM dapat memanfaatkan potensi AI untuk meningkatkan produktivitas, efisiensi, dan daya saing mereka.

Hambatan Mindset SDM

Antusiasme Aparatur Sipil Negara (ASN) terhadap kecerdasan buatan (AI) pada awalnya begitu tinggi. Mereka melihat potensi AI untuk meningkatkan efisiensi dan akurasi dalam pekerjaan sehari-hari. Namun, ketika diskusi beralih ke implementasi yang lebih konkret, seperti pengumpulan dan analisis data, antusiasme ini seringkali memudar. Banyak ASN yang merasa kurang nyaman dengan aspek teknis AI dan ragu untuk menerapkannya secara langsung.

Kurangnya pemahaman teknis menjadi salah satu hambatan utama dalam adopsi AI oleh ASN. Proses pengumpulan data dalam skala besar dan penggunaan tool analisis seperti ORANGE terasa kompleks bagi banyak ASN. Selain itu, kekhawatiran akan keamanan data dan perubahan signifikan dalam proses kerja juga menghalangi penerapan AI. Akibatnya, ASN seringkali lebih memilih untuk menyerahkan implementasi AI kepada pihak eksternal, mempersempit peluang untuk mengembangkan keahlian internal.

Penerapan AI tidak hanya membutuhkan keahlian teknis, tetapi juga perubahan budaya organisasi yang mendasar. AI menuntut pergeseran dari cara kerja konvensional ke pendekatan yang lebih berbasis data. Hal ini memerlukan perubahan mindset, keterampilan, dan struktur organisasi yang signifikan. Untuk mengatasi tantangan ini, diperlukan upaya yang komprehensif, seperti peningkatan literasi digital ASN, pelatihan intensif, dan dukungan berkelanjutan. Dengan demikian, ASN dapat lebih percaya diri dan aktif terlibat dalam proses transformasi digital di instansi pemerintah, koperasi, dan UMKM.

Tujuan Buku

Buku ini dirancang sebagai pendamping praktis bagi mereka yang telah mempelajari Strategi Nasional Kecerdasan Artifisial. Dengan pendekatan yang lebih aplikatif, buku ini menyajikan contoh-contoh konkret yang dapat langsung diterapkan dalam berbagai bidang yang relevan dengan kebijakan Kementerian Koperasi dan UKM (KemenkopUKM). Salah satu keunggulan utama buku ini adalah penggunaan aplikasi data mining ORANGE yang gratis dan mudah digunakan. Hal ini memungkinkan para pembaca, termasuk mereka yang tidak memiliki latar belakang teknis yang kuat, untuk secara aktif belajar dan menerapkan konsep kecerdasan buatan.

Fokus utama buku ini adalah penerapan kecerdasan buatan dalam konteks KemenkopUKM. Pembahasan teknis akan lebih berorientasi pada kebutuhan praktis seperti analisis kebijakan dan program-program yang relevan dengan UMKM di Indonesia. Dengan menggunakan ORANGE, pembaca dapat memperoleh pemahaman mendalam tentang bagaimana teknologi kecerdasan buatan dapat digunakan untuk mendukung pengambilan keputusan yang lebih baik dalam lingkungan

KemenkopUKM. Misalnya, analisis data dengan ORANGE dapat membantu mengidentifikasi tren, pola, dan peluang baru dalam pengembangan UMKM di Indonesia.

Kemampuan untuk melakukan analisis data dengan menggunakan tool seperti ORANGE tidak hanya terbatas pada sektor pemerintahan. Keterampilan ini juga sangat bermanfaat bagi dunia usaha, terutama dalam meningkatkan kinerja dan daya saing UMKM. Dengan memanfaatkan analisis data, UMKM dapat menyusun strategi pemasaran yang lebih efektif, meningkatkan pengalaman pelanggan, dan mengoptimalkan proses bisnis. Oleh karena itu, buku ini tidak hanya relevan bagi para pembuat kebijakan di KemenkopUKM, tetapi juga bagi para pelaku usaha UMKM yang ingin memanfaatkan teknologi kecerdasan buatan untuk mengembangkan bisnisnya.

Gambaran Umum Buku

Buku ini akan membawa kita dalam perjalanan mendalam mengenai peran kecerdasan buatan, *machine learning*, dan *big data* dalam transformasi KemenkopUKM / Pemerintahan / corporate pada BAGIAN 1. Setelah membahas dasar-dasar teknologi ini, kita telah mengeksplorasi berbagai pola penerapannya pada BAGIAN 2, tantangan yang dihadapi pada BAGIAN 2, serta potensi masa depan yang menjanjikan pada BAGIAN 4. Pada BAGIAN 5, Bab terakhir ini akan merangkum temuan-temuan penting dari seluruh bab sebelumnya, menyoroti pesan-pesan kunci, dan memberikan rekomendasi untuk implementasi yang efektif.

BAB 1: Kecerdasan Buatan untuk Indonesia, kita akan memulai perjalanan untuk memahami Kecerdasan Buatan di Indonesia. Bab ini akan mengulas secara mendalam Strategi Nasional Kecerdasan Buatan sebagai kerangka kerja pengembangan AI di Indonesia. Selain itu, kita juga akan mendapatkan gambaran umum tentang seluruh isi buku, sehingga pembaca dapat mengetahui topik-topik yang akan dibahas dan bagaimana topik-topik tersebut saling berkaitan.

Kecerdasan Buatan adalah pisau bermata dua. Di satu sisi, teknologi ini menawarkan berbagai manfaat yang dapat mengubah kehidupan kita. Namun, disisi lain, AI juga membawa sejumlah tantangan dan risiko yang

perlu kita waspadai. BAB 2 akan membahas secara rinci mengenai tantangan-tantangan tersebut, serta bagaimana kita dapat memanfaatkan AI secara bertanggung jawab untuk memaksimalkan manfaatnya sambil meminimalkan risikonya.

Penerapan AI tidak hanya terbatas pada sektor swasta. Bab 3 akan mengulas bagaimana AI dapat menjadi tool yang ampuh bagi Kementerian Koperasi dan UKM untuk memperkuat demokrasi. Dengan meningkatkan transparansi, akuntabilitas, dan akses informasi publik, AI berpotensi meningkatkan partisipasi masyarakat dalam proses pemerintahan.

Tahukah Anda bahwa di balik rekomendasi musik yang tepat dan mobil self-driving yang canggih, terdapat teknologi Kecerdasan Buatan (AI)? BAB 4 akan membawa Anda menjelajahi berbagai aplikasi AI dalam kehidupan sehari-hari. Anda akan memahami bagaimana AI bekerja secara mendasar dan bagaimana teknologi ini dapat mengambil keputusan yang kompleks. Selain itu, kita juga akan membahas pentingnya memastikan bahwa pengembangan AI dilakukan secara bertanggung jawab dan etis.

Jika Anda ingin belajar cara mengolah data dan menemukan insight berharga di dalamnya, BAB 5: Aplikasi Orange Data Mining adalah jawabannya. Bab ini akan mengajarkan Anda cara menggunakan Orange untuk membangun model data mining yang akurat dan efisien. Dengan panduan langkah demi langkah, Anda akan dapat menerapkan pengetahuan ini dalam proyek-proyek data mining Anda.

Bab 6 memberikan panduan awal bagi kita untuk melakukan prediksi anggaran menggunakan perangkat lunak Orange. Dengan dasar pemahaman yang ada, kita akan mencoba menerapkan teknik prediksi sederhana. Selanjutnya, kita akan mengeksplorasi bagaimana model prediksi dapat ditingkatkan dengan melibatkan lebih banyak variabel atau fitur. Tujuan akhir dari pembahasan ini adalah untuk memperoleh prediksi anggaran yang lebih akurat dan komprehensif dengan mempertimbangkan berbagai faktor yang relevan.

Ingin lebih memahami cara kerja Orange? BAB 7 mengajak Anda untuk berlatih langsung dengan dataset bawaan. Dengan membangun dan mengevaluasi model, Anda akan belajar bagaimana menangani masalah-masalah klasik dalam data mining seperti klasifikasi, regresi, dan pengelompokan data.

Prediksi data memiliki dampak yang signifikan terhadap keputusan yang kita ambil. BAB 8 akan menunjukkan bagaimana kita dapat menggunakan prediksi untuk mengatasi masalah kompleks seperti kemiskinan dan perencanaan pembangunan. Kita juga akan belajar bagaimana menilai kualitas prediksi dan mengidentifikasi faktor-faktor kunci yang perlu diperhatikan.

BAB 9 membahas analisis mendalam terhadap data koperasi dan UMKM di Indonesia menggunakan perangkat lunak Orange. Di contohkan teknik mengolah data yang melimpah ini memberikan gambaran komprehensif tentang kondisi UMKM di Indonesia.

BAB 10 membahas Kecerdasan Buatan (AI) yang merevolusi dunia pemasaran digital. AI memungkinkan bisnis, terutama koperasi dan UMKM, untuk lebih memahami pelanggan mereka dan memberikan penawaran yang lebih personal. Dengan AI, tugas-tugas rutin dapat diotomatisasi, analisis data menjadi lebih mendalam, dan keputusan bisnis pun lebih cerdas. AI dapat digunakan dalam berbagai aspek pemasaran, seperti personalisasi, otomatisasi, analisis data, dan periklanan.

Dalam upaya untuk mencapai efisiensi dan efektivitas yang lebih tinggi, banyak organisasi beralih ke *process mining*. BAB 11: Process Mining untuk Transformasi Digital akan memperkenalkan kita pada *process mining* sebagai tool yang ampuh untuk menganalisis dan mengoptimalkan proses bisnis. Kita akan membahas langkah-langkah implementasi *process mining*, serta tool dan teknologi yang umum digunakan, seperti PM4Py. Melalui contoh-contoh kasus, kita akan melihat bagaimana *process mining* dapat diterapkan dalam berbagai konteks bisnis.

BAB 12 akan membahas secara mendalam mengenai Ollama, sebuah model bahasa besar (LLM) sumber terbuka yang mendukung bahasa Indonesia. Kita akan menjelajahi berbagai kemampuan Ollama dalam

menghasilkan konten teks, terjemahan bahasa, dan menjawab pertanyaan. Selain itu, bab ini juga akan mengkaji integrasi teknologi Retrieval Augmented Generation (RAG) untuk meningkatkan kemampuan Ollama dalam mengakses dan memanfaatkan informasi yang relevan.

BAB 13 akan memberikan pengantar mendalam tentang TensorFlow, sebuah library yang populer untuk membangun dan melatih berbagai macam model neural network. Materi yang akan dibahas mencakup proses instalasi TensorFlow, teknik loading dataset, arsitektur model neural network, serta evaluasi kinerja model. Selain itu, bab ini juga akan membahas konsep loss function dan optimasi sebagai dasar dalam meningkatkan akurasi model.

Kemajuan pesat kecerdasan buatan telah menghadirkan dilema etis yang kompleks. Meskipun AI menjanjikan peningkatan signifikan dalam kualitas hidup manusia, namun penerapannya juga memunculkan berbagai permasalahan, terutama terkait dengan etika dan perlindungan data pribadi. BAB 14 akan secara mendalam mengkaji tantangan-tantangan tersebut dan menawarkan solusi yang komprehensif untuk memastikan pengembangan dan pemanfaatan AI yang bertanggung jawab.

Untuk mewujudkan kemajuan yang merata, kita perlu mengatasi permasalahan kesenjangan digital. BAB 15 akan mengulas berbagai solusi dan strategi untuk meningkatkan akses teknologi, keterampilan digital masyarakat, serta mengidentifikasi pihak-pihak yang berperan penting dalam mengatasi tantangan ini.

BAB 16 akan mengupas tuntas tantangan keamanan siber yang menyertai kemajuan pesat teknologi kecerdasan buatan (AI) dan big data. Ancaman seperti kebocoran data dan gangguan operasional sistem AI akan menjadi fokus utama pembahasan. Dalam konteks ini, peran pemerintah, khususnya KemenkopUKM, dalam merumuskan kebijakan dan regulasi yang komprehensif untuk melindungi data dan infrastruktur digital akan menjadi sorotan penting. Selain itu, bab ini juga akan menekankan urgensi penerapan standar keamanan siber yang ketat dalam setiap tahap pengembangan dan pemanfaatan teknologi AI.

Big Data memiliki potensi yang sangat besar untuk meningkatkan kinerja bisnis. Namun, tanpa pendekatan yang tepat, data ini hanya akan menjadi sekumpulan informasi mentah yang tidak bernilai. BAB 17 akan mengungkap berbagai strategi untuk mengubah data besar menjadi aset yang berharga. Bab ini akan membahas secara mendalam mengenai teknik-teknik pengolahan, analisis, dan visualisasi data yang efektif. Tujuannya adalah untuk menghasilkan wawasan yang dapat diandalkan dan dapat segera diimplementasikan untuk mencapai tujuan bisnis.

Perkembangan pesat AI, machine learning, dan big data telah membawa perubahan besar dalam berbagai aspek kehidupan kita. BAB 18 akan mengkaji lebih lanjut bagaimana teknologi-teknologi ini telah membentuk ulang lanskap bisnis dan masyarakat. Kita akan melihat bagaimana AI dapat meningkatkan efisiensi dan produktivitas, serta membuka peluang baru. Namun, disisi lain, kita juga perlu memperhatikan tantangan yang muncul, seperti masalah privasi data dan etika dalam pengembangan AI.

Implementasi kecerdasan buatan (AI) membutuhkan upaya kolaboratif lintas sektor untuk menciptakan ekosistem inovasi yang mendukung pengembangan dan penerapan solusi AI yang efektif, sebagaimana yang akan dibahas lebih lanjut pada BAB 19.

BAB 2: Sejarah dan Definisi Singkat

Sejarah singkat AI, Machine Learning dan Big Data

Sejarah Singkat Kecerdasan Buatan - Artificial Intelligence (AI)

- 1956: Konferensi Dartmouth menandai awal mula penelitian AI modern. Para peneliti mulai mengeksplorasi kemungkinan menciptakan mesin yang dapat berpikir dan bertindak seperti manusia.
- 1960-an: Perkembangan pesat terjadi, termasuk penciptaan program ELIZA yang mampu mensimulasikan percakapan dengan manusia.
- 1970-an & 1980-an: Periode "AI winter" diwarnai dengan kekecewaan karena kemajuan yang lambat.
- 1990-an: Kebangkitan AI dengan algoritma baru dan peningkatan daya komputasi.
- 2000-an: Perkembangan deep learning dan pembelajaran mesin yang mendalam, memungkinkan AI untuk mencapai kemajuan besar dalam berbagai bidang.
- 2020-an: AI menjadi semakin terintegrasi dalam kehidupan sehari-hari, dengan aplikasi dalam berbagai bidang seperti kesehatan, keuangan, dan transportasi.

Sejarah Singkat Machine Learning

- 1952: Arthur Samuel mengembangkan program komputer yang dapat belajar bermain catur.
- 1960-an: Algoritma machine learning mulai digunakan untuk menyelesaikan berbagai masalah, seperti pengenalan pola dan klasifikasi data.
- 1970-an & 1980-an: Perkembangan statistik Bayesian dan Artificial Neural Networks (ANN).
- 1990-an: Algoritma seperti SVM dan decision tree menjadi populer.
- 2000-an: Kebangkitan deep learning dengan algoritma seperti convolutional neural networks (CNNs) dan recurrent neural networks (RNNs).
- 2020-an: Machine learning menjadi tool yang penting untuk berbagai bidang, termasuk ilmu data, robotika, dan computer vision.

Sejarah Singkat Big Data

- 1940-an: Mesin pemrosesan data elektronik pertama muncul.
- 1960-an: Perkembangan database dan sistem informasi.
- 1970-an & 1980-an: Munculnya istilah "big data" dan pengembangan tool-tool untuk menganalisisnya.
- 1990-an: Perkembangan internet dan e-commerce, yang menghasilkan ledakan data.
- 2000-an: Perkembangan Hadoop dan NoSQL, yang memungkinkan penyimpanan dan analisis data dalam skala besar.
- 2020-an: Big data menjadi sumber daya yang penting untuk berbagai bidang, termasuk bisnis, pemerintahan, dan ilmu pengetahuan.

Hubungan Antara AI, Machine Learning, dan Big Data

- AI adalah sebuah konsep yang luas, sedangkan machine learning adalah salah satu teknik untuk mencapai AI.
- Big data menyediakan bahan bakar untuk machine learning, memungkinkan algoritma AI untuk belajar dari data yang besar dan kompleks.
- Ketiga bidang ini saling terkait dan berkembang bersama, mendorong kemajuan teknologi dan membuka peluang baru di berbagai bidang.

Kemajuan pesat dalam bidang kecerdasan buatan (AI), *machine learning*, dan pengolahan data masif telah membawa dampak yang signifikan terhadap kehidupan manusia. Ketiga teknologi ini saling mendukung dan membuka peluang baru yang tak terbatas, mengubah cara kita bekerja, berkomunikasi, dan bahkan berpikir.

Definisi singkat AI, Machine Learning dan Big Data

Definisi Artificial Intelligence (AI) / Kecerdasan Buatan / Kecerdasan Artifisial :

Kecerdasan Buatan (AI) adalah bidang ilmu komputer yang fokus pada pengembangan mesin yang dapat meniru kecerdasan manusia untuk melaksanakan tugas-tugas kompleks seperti

berpikir, belajar, dan mengambil keputusan. Melalui penggunaan algoritma dan model matematika, AI memungkinkan mesin untuk belajar dari data, mengenali pola, dan memecahkan masalah tanpa campur tangan manusia. Berbagai pendekatan, termasuk pembelajaran mesin, jaringan saraf tiruan, dan pemrosesan bahasa alami, digunakan untuk mencapai kemampuan ini, yang bertujuan untuk mengotomatisasi proses dan meningkatkan efisiensi dalam berbagai aplikasi.

Definisi Machine Learning (ML):

Machine learning adalah subbidang kecerdasan buatan yang berfokus pada pengembangan algoritma dan model statistik. Tujuannya adalah agar sistem komputer dapat belajar secara otomatis dari data, mengidentifikasi pola, dan membuat prediksi atau keputusan tanpa perlu diprogram secara eksplisit untuk setiap tugas.

Ada beberapa jenis utama dari machine learning, termasuk:

- **Supervised Learning:** Dalam metode ini, model dilatih dengan data yang sudah dipasangkan antara input dan output yang benar. Model berusaha menemukan pola hubungan antara input dan output ini. Tujuan akhirnya adalah agar model mampu memberikan output yang tepat ketika diberikan input baru yang belum pernah ditemukan sebelumnya.
- **Unsupervised Learning:** Pendekatan di mana model secara mandiri mencari-cari keteraturan tersembunyi dalam data yang tidak berlabel. Ini mirip dengan cara kita sebagai manusia mengelompokkan objek berdasarkan kesamaan ciri.
- **Reinforcement Learning:** Teknik di mana agen (model) belajar untuk mengambil keputusan optimal dalam suatu lingkungan. Agen akan belajar dari interaksi dengan lingkungannya, di mana setiap tindakan akan menghasilkan suatu hasil (*reward* atau *punishment*). Tujuan utama dari

pembelajaran ini adalah memaksimalkan total reward yang diperoleh dalam jangka panjang.

Machine Learning, dengan kemampuannya dalam mengekstraksi pola dari dataset yang besar dan kompleks, telah menjadi kekuatan pendorong di balik berbagai aplikasi seperti visi komputer, pemrosesan bahasa alami, dan sistem pengambilan keputusan. Teknologi ini memungkinkan kita untuk memperoleh wawasan yang mendalam dan membuat prediksi yang akurat.

Definisi Big Data:

merujuk pada volume besar data yang dihasilkan oleh organisasi dan sistem di seluruh dunia. Data ini memiliki karakteristik tersendiri yang dikenal sebagai **"3V"**: volume (jumlah data yang besar), velocity (kecepatan di mana data tersebut diproduksi dan diproses), dan variety (beragam jenis data, termasuk data terstruktur dan tidak terstruktur). Terkadang, karakteristik tambahan seperti variabilitas (perubahan dalam format atau nilai data) dan veracity (kepastian atau keandalan data) juga dimasukkan ke dalam definisi big data.

Big data merujuk pada sekumpulan data yang sangat besar dan kompleks, mencakup data terstruktur, semi-terstruktur, dan tidak terstruktur. Data ini berasal dari berbagai sumber, seperti transaksi bisnis, sensor IoT, dan media sosial. Pertumbuhannya yang pesat memunculkan tantangan dalam hal pengelolaan dan analisis data.

Organisasi mengadopsi beragam solusi teknologi untuk mengatasi tantangan dalam mengolah big data. Di antaranya adalah sistem manajemen basis data yang dirancang khusus untuk data berskala besar, teknik pemrosesan data yang terdistribusi di banyak komputer, metode analisis data tingkat lanjut seperti machine learning dan data mining, serta infrastruktur komputasi yang dapat disesuaikan dengan kebutuhan yang terus berubah, seperti cloud computing.

Manfaat dan Potensi AI

Koperasi dan UMKM dapat meraih manfaat besar dari integrasi kecerdasan buatan, pembelajaran mesin, dan big data. Kombinasi

teknologi ini memungkinkan analisis mendalam terhadap data bisnis yang sangat besar dan kompleks. Melalui pembelajaran mesin, koperasi dan UMKM dapat mengidentifikasi pola-pola tersembunyi dalam data, seperti tren pasar, perilaku konsumen, dan kinerja produk. Dengan begitu, keputusan bisnis dapat diambil secara lebih akurat dan efektif. Misalnya, koperasi dapat memprediksi permintaan produk di masa depan, sehingga dapat mengoptimalkan produksi dan distribusi.

Salah satu keunggulan utama dari integrasi teknologi ini adalah kemampuan untuk mempersonalisasi layanan dan produk. Dengan memanfaatkan data pelanggan, koperasi dan UMKM dapat memahami preferensi dan kebutuhan individu secara lebih baik. Melalui pembelajaran mesin, sistem dapat memberikan rekomendasi produk yang relevan, penawaran khusus yang sesuai, serta layanan pelanggan yang lebih personal. Hal ini tidak hanya meningkatkan kepuasan pelanggan, tetapi juga mendorong peningkatan penjualan. Selain itu, data yang dikumpulkan dapat digunakan untuk mengembangkan produk dan layanan baru yang lebih sesuai dengan kebutuhan pasar.

Penerapan kecerdasan buatan, pembelajaran mesin, dan big data dapat mendorong koperasi dan UMKM untuk menjadi lebih inovatif dan kompetitif. Dengan memanfaatkan teknologi ini, koperasi dan UMKM dapat mengidentifikasi peluang bisnis baru, mengoptimalkan proses produksi, dan meningkatkan efisiensi operasional. Selain itu, teknologi ini juga dapat membantu koperasi dan UMKM dalam memasuki pasar baru dan bersaing dengan bisnis yang lebih besar. Dengan demikian, integrasi teknologi ini dapat menjadi kunci keberhasilan koperasi dan UMKM dalam menghadapi tantangan bisnis di era digital.

Intinya: Integrasi teknologi kecerdasan buatan, pembelajaran mesin, dan big data menawarkan potensi besar bagi koperasi dan UMKM untuk meningkatkan efisiensi, personalisasi layanan, inovasi, dan daya saing. Dengan memanfaatkan teknologi ini, koperasi dan UMKM dapat membuat keputusan bisnis yang lebih baik, meningkatkan kepuasan pelanggan, dan membuka peluang bisnis baru.

Tantangan dan Risiko AI

Penerapan teknologi kecerdasan buatan (AI), pembelajaran mesin (ML), dan data besar (Big Data) di dunia bisnis, termasuk koperasi dan UMKM, membawa angin segar namun juga tantangan serius. Salah satu tantangan utama adalah **keamanan dan privasi data**. Koperasi dan UMKM yang mengelola data pelanggan atau anggota rentan terhadap pelanggaran data yang dapat berdampak buruk pada reputasi dan kepercayaan. Selain itu, **bias dalam algoritma AI** dapat menghasilkan keputusan yang tidak adil, misalnya dalam hal pemberian kredit atau pemilihan produk. Hal ini dapat merugikan koperasi dan UMKM, terutama yang menasar segmen masyarakat tertentu.

Ketergantungan yang berlebihan pada teknologi AI juga menjadi risiko. Jika sistem AI mengalami gangguan atau memberikan hasil yang tidak akurat, koperasi dan UMKM dapat mengalami kerugian finansial dan operasional. Selain itu, muncul pertanyaan **etika terkait penggunaan AI** dalam pengambilan keputusan bisnis. Misalnya, bagaimana jika algoritma AI merekomendasikan kebijakan yang merugikan anggota koperasi? Pertanyaan-pertanyaan seperti ini perlu dijawab dengan jelas untuk memastikan bahwa penggunaan AI tidak bertentangan dengan nilai-nilai koperasi yang menjunjung tinggi keadilan dan kesejahteraan bersama.

Untuk mengatasi tantangan-tantangan tersebut, koperasi dan UMKM perlu mempersiapkan diri dengan baik. Salah satunya adalah dengan **meningkatkan kapasitas sumber daya manusia** agar mampu memahami dan mengelola teknologi AI. Selain itu, diperlukan kerja sama yang erat antara koperasi, UMKM, pemerintah, dan akademisi untuk mengembangkan standar **etika penggunaan AI** yang sesuai dengan konteks Indonesia, khususnya dalam konteks koperasi dan UMKM. Dengan demikian, potensi positif dari teknologi AI dapat dimanfaatkan secara optimal untuk meningkatkan kinerja dan daya saing koperasi dan UMKM.

BAB 3: AI untuk Koperasi & UKM

Efisiensi dan Efektivitas Layanan

Teknologi kecerdasan buatan (AI), pembelajaran mesin (*machine learning*), dan *big data* telah membuka peluang besar untuk meningkatkan kualitas layanan yang diberikan oleh pemerintah, termasuk kepada koperasi dan UMKM. Dengan adanya chatbot dan asisten virtual, koperasi dan UMKM dapat mengakses informasi dan layanan publik secara lebih mudah dan cepat. Selain itu, analisis data memungkinkan pemerintah untuk memahami kebutuhan spesifik dari koperasi dan UMKM sehingga dapat menyusun program dan kebijakan yang lebih relevan dan efektif. Hal ini akan berdampak positif pada pertumbuhan dan perkembangan koperasi dan UMKM.

Penggunaan teknologi AI, *machine learning*, dan *big data* juga dapat membantu koperasi dan UMKM dalam pengambilan keputusan bisnis. Dengan menganalisis data pasar, tren industri, dan perilaku konsumen, koperasi dan UMKM dapat mengidentifikasi peluang bisnis baru, mengoptimalkan strategi pemasaran, serta mengelola risiko dengan lebih baik. Selain itu, teknologi ini dapat digunakan untuk memprediksi permintaan pasar sehingga koperasi dan UMKM dapat mengatur produksi dan persediaan barang dengan lebih efisien.

Otomasi proses bisnis melalui teknologi AI dan *machine learning* dapat membantu koperasi dan UMKM meningkatkan efisiensi operasional. Tugas-tugas rutin dan berulang seperti pembukuan, pengelolaan inventori, dan layanan pelanggan dapat dilakukan secara otomatis, sehingga sumber daya manusia dapat difokuskan pada tugas-tugas yang lebih strategis. Selain itu, analisis data dapat membantu koperasi dan UMKM mengidentifikasi inefisiensi dalam proses produksi dan distribusi, sehingga dapat dilakukan perbaikan untuk mengurangi biaya dan meningkatkan produktivitas.

Meskipun menawarkan banyak manfaat, pemanfaatan teknologi AI, *machine learning*, dan *big data* juga dihadapkan pada beberapa tantangan, seperti ketersediaan data yang berkualitas, keamanan data, dan kesenjangan digital. Namun, dengan mengatasi tantangan-tantangan

tersebut, koperasi dan UMKM dapat memanfaatkan teknologi ini untuk meningkatkan daya saing dan memberikan kontribusi yang lebih besar bagi perekonomian nasional. Pemerintah juga perlu berperan aktif dalam mendukung adopsi teknologi ini di kalangan koperasi dan UMKM, misalnya dengan menyediakan pelatihan dan infrastruktur yang memadai.

Pengambilan Keputusan berdasarkan Data

Koperasi dan UMKM dapat sangat terbantu dengan penerapan teknologi kecerdasan buatan (AI) dan pembelajaran mesin (ML) dalam pengambilan keputusan. AI dan ML mampu menganalisis data dalam jumlah besar dan kompleks untuk menghasilkan prediksi yang akurat. Misalnya, dengan memprediksi tren pasar, Koperasi dan UMKM dapat menyesuaikan produksi atau layanan mereka. Selain itu, rekomendasi yang dipersonalisasi berdasarkan data pelanggan dapat meningkatkan penjualan dan kepuasan pelanggan. Dengan kata lain, AI dan ML dapat menjadi tool yang sangat berharga untuk membuat keputusan bisnis yang lebih cerdas dan efektif.

Salah satu manfaat besar dari AI dan ML adalah kemampuannya untuk mengoptimalkan berbagai proses. Dalam konteks Koperasi dan UMKM, optimasi ini dapat diterapkan dalam berbagai bidang, mulai dari manajemen rantai pasok hingga penentuan harga produk. Misalnya, AI dapat membantu mengoptimalkan rute pengiriman barang, sehingga mengurangi biaya operasional. Selain itu, AI juga dapat digunakan untuk mendeteksi pola konsumsi pelanggan dan mengidentifikasi peluang bisnis baru. Dengan demikian, Koperasi dan UMKM dapat menjadi lebih efisien dan kompetitif.

Meskipun menawarkan banyak manfaat, penerapan AI dan ML dalam Koperasi dan UMKM juga menghadapi sejumlah tantangan. Salah satu tantangan utama adalah ketersediaan data yang berkualitas dan relevan. Selain itu, diperlukan sumber daya manusia yang memiliki keahlian di bidang AI dan ML untuk mengembangkan dan mengimplementasikan solusi berbasis AI. Tantangan lainnya adalah terkait dengan privasi data dan keamanan informasi. Oleh karena itu, penting bagi Koperasi dan UMKM untuk memiliki strategi yang jelas dalam mengelola data dan memastikan bahwa penggunaan AI sesuai dengan etika dan regulasi yang berlaku.

Transparansi dan Akuntabilitas

Koperasi dan UMKM kini semakin merasakan manfaat dari penerapan teknologi kecerdasan buatan (AI) dan pembelajaran mesin. Dengan memanfaatkan analisis data yang mendalam, pemerintah dan lembaga terkait seperti KemenkopUKM dapat mendeteksi potensi penyimpangan atau ketidaksesuaian dalam pengelolaan dana koperasi dan UMKM. Hal ini memungkinkan pengawasan yang lebih efektif dan transparan, sehingga dapat mencegah terjadinya penyalahgunaan dana dan meningkatkan kepercayaan masyarakat terhadap sektor koperasi dan UMKM.

Selain itu, pemantauan kinerja koperasi dan UMKM juga dapat dilakukan secara otomatis dengan bantuan AI. Misalnya, dalam memantau perkembangan proyek-proyek yang didanai oleh koperasi atau UMKM. Dengan adanya data yang akurat dan terkini, pemerintah dan pemangku kepentingan lainnya dapat memberikan dukungan yang lebih tepat sasaran kepada koperasi dan UMKM. Hal ini akan mendorong pertumbuhan yang lebih berkelanjutan dan meningkatkan kontribusi sektor koperasi dan UMKM terhadap perekonomian nasional.

Koperasi dan UMKM juga dapat memanfaatkan AI untuk mendengarkan suara para anggotanya atau pelanggannya. Melalui analisis sentimen pada media sosial, koperasi dan UMKM dapat memahami lebih baik kebutuhan dan harapan para anggotanya atau pelanggannya. Dengan demikian, koperasi dan UMKM dapat menyusun strategi bisnis yang lebih relevan dan responsif. Hal ini tidak hanya akan meningkatkan kepuasan anggota atau pelanggan, tetapi juga memperkuat posisi koperasi dan UMKM dalam persaingan bisnis.

Memperkuat demokrasi dan partisipasi publik

Kecerdasan Buatan (AI) dan pembelajaran mesin memiliki potensi besar untuk merevolusi cara Koperasi dan UMKM berinteraksi dengan pemerintah. Dengan mengotomatiskan penyediaan informasi melalui chatbot AI, misalnya, UMKM dapat dengan mudah mendapatkan jawaban atas pertanyaan mereka tentang kebijakan atau program pemerintah yang mendukung usaha mereka. Akses informasi yang lebih mudah ini akan memberdayakan para pelaku UMKM untuk membuat keputusan bisnis

yang lebih baik dan berpartisipasi aktif dalam ekosistem ekonomi yang lebih luas.

Selain itu, analisis data menggunakan pembelajaran mesin dapat memberikan wawasan berharga bagi Koperasi dan UMKM. Data yang kompleks dapat diolah menjadi informasi yang mudah dipahami, seperti infografis atau visualisasi data, yang memungkinkan UMKM untuk mengidentifikasi tren pasar, peluang bisnis baru, dan tantangan yang perlu diatasi. Dengan pemahaman yang lebih baik tentang lingkungan bisnis mereka, UMKM dapat membuat strategi yang lebih efektif untuk meningkatkan daya saing dan pertumbuhan.

AI juga dapat mendorong partisipasi UMKM dalam proses pembuatan kebijakan. Platform online yang didukung AI dapat memfasilitasi dialog antara pemerintah dan UMKM, memungkinkan UMKM untuk memberikan masukan dan umpan balik secara langsung terhadap kebijakan yang berdampak pada usaha mereka. Dengan melibatkan UMKM dalam proses pengambilan keputusan, pemerintah dapat memastikan bahwa kebijakan yang dibuat relevan dan bermanfaat bagi pelaku usaha di lapangan. Hal ini akan memperkuat hubungan antara pemerintah dan UMKM, serta menciptakan lingkungan bisnis yang lebih kondusif bagi pertumbuhan ekonomi.

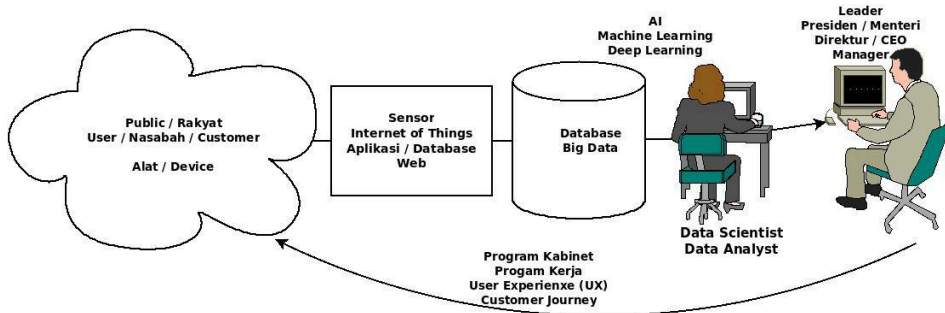
BAGIAN 2: POLA PENERAPAN AI, MACHINE LEARNING & BIG DATA

BAB 4: Pola AI dan Transformasi Digital

Transformasi Digital dan AI

Transformasi digital bagi koperasi dan UMKM bukan sekadar mengganti cara kerja konvensional dengan yang digital. Ini adalah perubahan mendasar yang membutuhkan pemahaman yang dalam. Penting bagi koperasi dan UMKM untuk tidak hanya sekedar mendigitalisasi proses-proses yang sudah ada, tetapi juga mengubah cara berpikir dan bekerja secara fundamental. Dengan demikian, transformasi digital tidak hanya sekedar mengubah format, tetapi juga meningkatkan efisiensi dan efektivitas. Lebih dari itu, koperasi dan UMKM Indonesia harus memiliki kemandirian dalam mengadopsi teknologi digital agar tidak hanya menjadi pengguna pasif, tetapi juga menjadi pelaku aktif dalam pengembangan teknologi.

Data merupakan aset berharga dalam era digital. Koperasi dan UMKM yang memiliki data yang lengkap dan akurat dapat memanfaatkannya untuk membuat keputusan bisnis yang lebih baik. Dengan bantuan teknologi kecerdasan buatan (AI), koperasi dan UMKM dapat menganalisis data dengan cepat dan akurat, sehingga dapat mengidentifikasi peluang pasar yang baru, meningkatkan efisiensi operasional, dan memberikan pelayanan yang lebih baik kepada anggota atau pelanggan. Selain itu, data juga dapat digunakan untuk mengembangkan produk dan layanan yang lebih personal dan sesuai dengan kebutuhan pasar. Dengan demikian, data dapat menjadi pendorong utama bagi pertumbuhan dan keberhasilan koperasi dan UMKM di era digital.



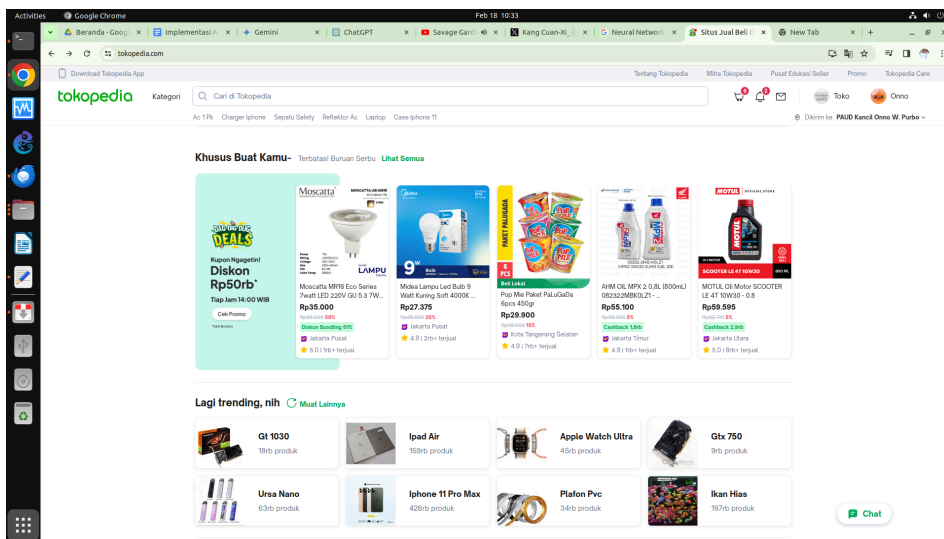
Gambar 4.1. Bisnis Proses AI / Machine Learning

Dalam praktiknya, big data dapat dikumpulkan dari berbagai sumber seperti laporan keuangan, komentar media sosial, berita, laporan kementerian, lembaga, pemerintah kota, dan pemerintah daerah. Data yang terkumpul kemudian diolah oleh para data scientist menggunakan berbagai tool seperti AI, machine learning, dan deep learning. Hasil analisis ini memberikan informasi yang berguna bagi para pemimpin, baik itu presiden, menteri, direktur, CEO, Board of Director (BoD) maupun manajer, untuk merancang program atau kegiatan yang lebih personal dan sesuai dengan kebutuhan masyarakat atau pengguna, dengan tujuan meningkatkan pengalaman (*user experience*) dan perjalanan pengguna (*customer journey*) secara keseluruhan. Namun, penting untuk diingat bahwa keberhasilan proses ini sangat tergantung pada kemampuan untuk memantau situasi dengan baik di lapangan, tanpa melanggar hak privasi individu atau berpotensi menjadi proses pengawasan yang tidak etis.

Kecerdasan buatan (AI) memiliki potensi besar untuk membantu Koperasi dan UMKM dalam pengambilan keputusan. Namun, keberhasilan penerapan AI sangat bergantung pada kualitas data yang digunakan untuk melatih modelnya. **Penggunaan data dari luar negeri yang tidak relevan** dengan kondisi Koperasi dan UMKM di Indonesia dapat menghasilkan prediksi yang tidak akurat. Oleh karena itu, sangat penting bagi Koperasi dan UMKM untuk memastikan bahwa data yang digunakan dalam sistem AI mereka berasal dari sumber yang terpercaya dan mencerminkan kondisi bisnis lokal. Dengan demikian, AI dapat memberikan rekomendasi yang lebih tepat dan membantu Koperasi serta UMKM dalam meningkatkan kinerja bisnis mereka.

Contoh Penerapan AI di Dunia Nyata

Kecerdasan Buatan (AI) dan Pembelajaran Mesin (*Machine Learning*) dapat menjadi tool yang sangat berharga bagi koperasi dan UMKM dalam meningkatkan penjualan dan pengalaman pelanggan. Dengan menerapkan teknologi ini, koperasi dan UMKM dapat mempersonalisasi pemasaran mereka, sama seperti platform toko online besar. Melalui analisis data pelanggan, iklan yang ditampilkan dapat disesuaikan dengan minat dan preferensi masing-masing individu. Hal ini tidak hanya meningkatkan relevansi iklan, tetapi juga menciptakan pengalaman belanja yang lebih personal dan memuaskan bagi anggota koperasi atau pelanggan UMKM. Dengan demikian, kemungkinan anggota atau pelanggan untuk melakukan pembelian berulang dan merekomendasikan produk atau jasa kepada orang lain akan semakin besar.



Gambar 4.2. Halaman Depan Tokopedia.

Platform e-commerce seperti Tokopedia, dengan dominasi iklannya, telah memanfaatkan algoritma untuk memahami preferensi pengguna secara mendalam. **Data riwayat belanja, pencarian, dan aktivitas pengguna lainnya dikumpulkan dan dianalisis untuk menyajikan iklan yang relevan.** Mekanisme ini, meskipun efektif untuk bisnis besar, perlu dipertimbangkan dampaknya bagi koperasi dan UMKM. Algoritma yang canggih ini dapat memberikan keuntungan bagi perusahaan besar dengan sumber daya yang lebih besar, sementara koperasi dan UMKM mungkin

kesulitan bersaing dalam hal visibilitas iklan. Pertanyaan mendasar yang muncul adalah bagaimana platform dapat memastikan bahwa koperasi dan UMKM juga memiliki kesempatan yang sama untuk menjangkau konsumen potensial melalui platform e-commerce.

Teknologi kecerdasan buatan (AI) dan pembelajaran mesin (machine learning) telah membuka peluang besar bagi koperasi dan UMKM untuk meningkatkan penjualan dan pelayanan mereka melalui platform e-commerce. Dengan melacak aktivitas pengguna secara cermat, platform dapat menciptakan profil individu yang unik berdasarkan preferensi belanja masing-masing. Profil ini kemudian digunakan untuk menyajikan rekomendasi produk dan iklan yang lebih relevan, sehingga meningkatkan kemungkinan terjadinya transaksi. Pendekatan personalisasi ini tidak hanya membuat pengalaman belanja menjadi lebih menyenangkan bagi konsumen, tetapi juga membantu koperasi dan UMKM menjangkau target pasar yang lebih spesifik.

Proses personalisasi yang didukung AI ini melibatkan pengumpulan dan analisis data pengguna dalam skala besar. Kualitas data yang digunakan dalam proses pelatihan model AI sangat krusial, karena data yang buruk dapat menghasilkan rekomendasi yang tidak akurat. Koperasi dan UMKM perlu berhati-hati dalam mengelola data pelanggan dan memastikan bahwa data tersebut digunakan secara etis dan sesuai dengan peraturan yang berlaku. Selain itu, penting untuk terus melakukan evaluasi dan perbaikan terhadap model AI agar tetap relevan dan efektif dalam menghadapi perubahan perilaku konsumen.

Untuk memperlihatkan kemampuan AI untuk melakukan forecasting atau sering kali saya sebut sebagai perdukunan. Paling gampang jika kita dapat menggunakan Google Maps untuk visualisasikan proses *perdukunan* di sebuah platform.

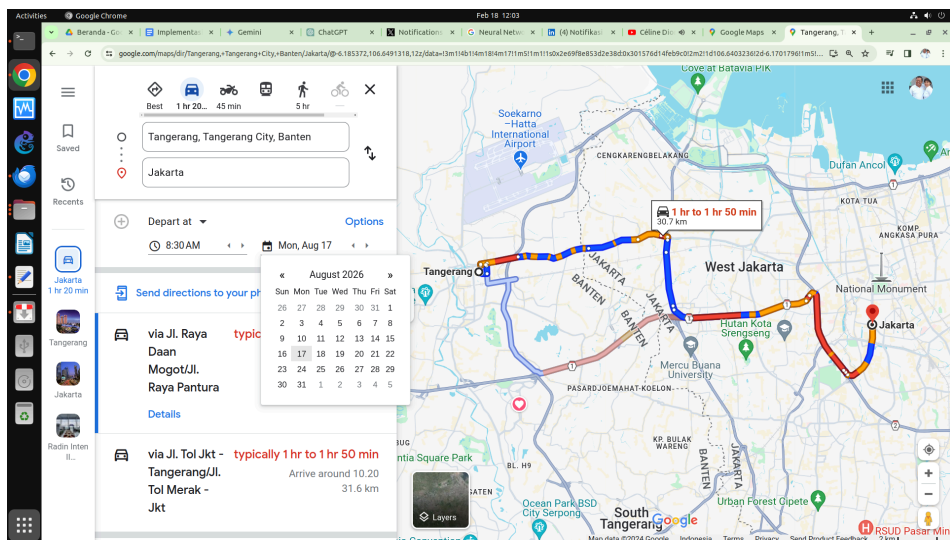
Langkah yang perlu dilakukan adalah sebagai berikut,

- Buka web Google Maps <https://www.google.com/maps/>
- Jika kita ingin menganalisis perjalanan dari Tangerang Ke Jakarta, masukan di text box search kata kunci "Tangerang"
- Klik button DIRECTIONS

- Isi pada Text Box “Choose Destination” - kata kunci “Jakarta”.

Pada titik ini, kita akan memperoleh informasi jalur yang bisa di tempuh untuk perjalanan Tangerang ke Jakarta lengkap dengan tempat macet yang akan terjadi. Untuk melihat kemampuan Google Maps melakukan perdukunan, kita dapat melakukan hal yang cukup ekstrim seperti,

- Klik pilihan “Leave Now”
- Ubah menjadi “Depart at”
- Langkah ekstrim yang perlu dilakukan adalah ubah tanggal ke tanggal yang akan datang, dapat contoh / gambar dipilih tanggal 17 Agustus 2026 jam 8:30 pagi.



Gambar 4.3. Saran Perjalanan Tangerang Jakarta 17 Agustus 2026

Dalam gambar yang ditampilkan, terdapat saran jalur yang dapat diambil untuk perjalanan dari Tangerang menuju Jakarta pada tanggal 17 Agustus 2026, yang masih dua tahun lagi dari saat naskah ini ditulis. Dari gambar tersebut, kita dapat melihat titik-titik kemacetan yang diperkirakan akan terjadi pada tanggal tersebut. Penting untuk dicatat bahwa perjalanan ini baru akan dilakukan di masa yang akan datang, mengingat buku ini ditulis pada tahun 2024.

Kemacetan lalu lintas di kota besar seperti Jakarta sangat dinamis dan sering berubah-ubah. Jam sibuk di pagi hari, misalnya, biasanya menyebabkan kemacetan parah dari Tangerang menuju Kebon Jeruk. Selain itu, kondisi lalu lintas pada akhir pekan juga jauh berbeda dengan hari kerja. Faktor-faktor seperti waktu dan hari dalam seminggu sangat mempengaruhi di mana dan kapan kemacetan terjadi.

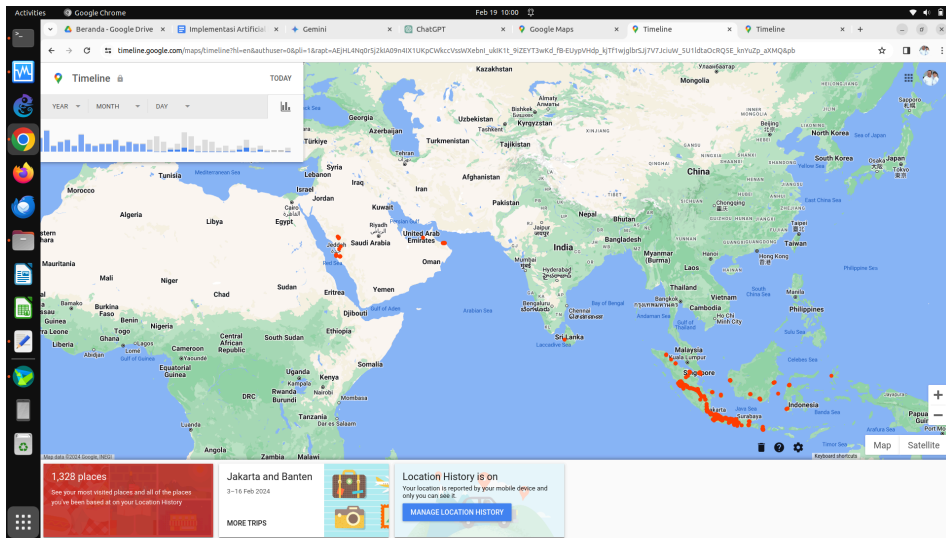
Banyak orang penasaran bagaimana Google Maps bisa tahu persis di mana terjadi kemacetan pada waktu tertentu. Beberapa orang mungkin berpikir Google menggunakan satelit atau sensor khusus di jalan. Namun, sebenarnya, data utama yang digunakan Google berasal dari ponsel pintar pengguna. Setiap kali kita menggunakan Google Maps, terutama saat navigasi, ponsel kita mengirimkan data lokasi secara berkala ke server Google.

Dengan mengumpulkan data lokasi dari jutaan pengguna Android di seluruh dunia selama bertahun-tahun, Google dapat menganalisis pola lalu lintas. Data ini kemudian diolah oleh algoritma canggih untuk memprediksi dimana kemacetan akan terjadi di masa depan. Singkatnya, Google Maps dapat memberikan informasi lalu lintas yang akurat karena memanfaatkan data lokasi dari ponsel kita. Semakin banyak orang yang menggunakan Google Maps, semakin akurat pula prediksi kemacetannya.

Intinya: Google Maps bisa memprediksi kemacetan dengan sangat akurat karena mengumpulkan data lokasi dari jutaan ponsel Android. Data ini kemudian diolah untuk melihat pola lalu lintas dan memprediksi kemacetan di masa depan.

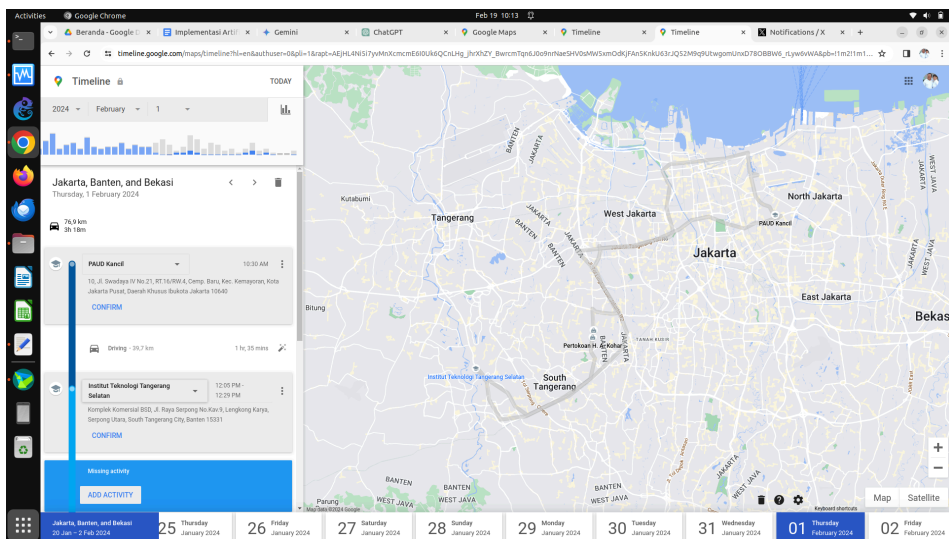
Untuk memverifikasi bahwa kita atau pengguna ponsel Android di monitor / dimata-matai oleh Google, langkah yang paling mudah dilakukan adalah:

- Buka Google Maps melalui peramban web di komputer.
- Klik pada menu Google Maps, biasanya terletak di sudut kiri atas.
- Pilih "Your Timeline."



Gambar 4.4. Lokasi / Kota yang pernah didatangi penulis.

Penulis terlihat telah melakukan banyak perjalanan ke berbagai kota di Indonesia, serta ke Arab, terutama Mekkah dan Madinah. Untuk melihat rute perjalanan secara lebih detail, kita dapat mengklik tanggal yang tertera di bagian kiri atas. Hasilnya akan ditampilkan dalam gambar selanjutnya.



Gambar 4.5. Tampilan detail rute perjalanan.

Pada gambar di atas, terlihat perjalanan penulis dalam satu hari dari Jakarta ke Tangerang Selatan dan sebaliknya. Jika kita memperluas skala pemantauan, misalnya dengan membayangkan Google Maps melacak seluruh ponsel di dunia selama bertahun-tahun, maka kita dapat melihat potensi bahaya dari kecerdasan buatan. Bayangkan saja jika data ini digunakan untuk melacak pergerakan seluruh anggota aparat atau TNI Indonesia. Kemampuan AI untuk memantau aktivitas kita secara massal dan mendalam seperti ini menunjukkan betapa mudahnya privasi kita terancam, bahkan dalam aktivitas yang kita anggap biasa dan aman.

Persamaan Linier: Cara Berpikir Sederhana AI

Kecerdasan buatan (AI) bergantung pada pemodelan data untuk melakukan tugas seperti klasifikasi dan prediksi. Para ahli seringkali menggambarkan cara kerja AI dengan analogi persamaan linear sederhana. Dalam banyak kasus, AI menganalisis data dengan menggunakan persamaan linear yang memiliki bentuk umum tertentu. Dengan kata lain, AI mencoba menemukan hubungan linier antara data yang diberikan untuk membuat kesimpulan atau prediksi yang akurat. Konsep ini menjadi dasar dari banyak algoritma pembelajaran mesin yang digunakan dalam AI.

$$A x + B = y$$

Persamaan linear dalam kecerdasan buatan adalah tool yang digunakan untuk mengidentifikasi pola dalam data. Parameter A dan B dalam persamaan ini berperan penting dalam memetakan hubungan antara variabel input (x) dan output (y). Variabel input dapat berupa berbagai jenis data, mulai dari angka hingga data non-numerik seperti gambar atau suara. Meskipun data input bisa multidimensional, tujuan utama seringkali adalah untuk memprediksi atau mengklasifikasikan suatu variabel target tunggal. Dengan memahami persamaan dasar ini, kita bisa melihat bagaimana AI mampu mengubah data mentah menjadi informasi bermakna yang dapat kita gunakan untuk membuat keputusan.

Untuk memberikan gambaran yang lebih jelas, mari kita lihat contoh data training sederhana untuk persamaan linear. Misalkan kita memiliki pasangan data x dan y sebagai berikut:

Tabel 4.1. Data Training Sederhana

X	Y
1	7
2	8
3	9

Data yang telah kita kumpulkan ini berperan sebagai 'guru' bagi model kecerdasan buatan kita. Dalam dunia pembelajaran mesin, data ini dikenal sebagai data latih. Tujuannya adalah untuk 'mengajar' model persamaan linear $Ax+B=y$ sehingga model tersebut mampu memprediksi nilai y dengan akurat. Proses pembelajaran ini melibatkan penentuan nilai A dan B yang paling tepat. Dengan memanfaatkan konsep matematika dasar yang kita pelajari di sekolah menengah, kita dapat menggunakan metode substitusi untuk menghitung nilai A dan B dari data latih yang ada.

$$A + B = 7 \quad (4.1)$$

$$2 A + B = 8 \quad (4.2)$$

$$3 A + B = 9 \quad (4.3)$$

Misalnya, dengan mengurangkan Persamaan (4.2) dengan Persamaan (4.1), kita mendapatkan:

$$A = 1 \quad (4.4)$$

$$B = 6 \quad (4.5)$$

Dalam dunia kecerdasan buatan, model A dan B yang kita cari dari data pelatihan seringkali tidak dapat ditemukan dengan metode substitusi sederhana karena data real-world cenderung lebih kompleks. Oleh karena itu, berbagai algoritma dari statistik (seperti Regresi Linier) hingga kecerdasan buatan (seperti Neural Network) dikembangkan untuk menemukan model yang paling sesuai. Pilihan algoritma sangat bergantung pada jenis data dan masalah yang ingin diselesaikan. Dalam buku ini, kita akan fokus pada contoh-contoh sederhana menggunakan perangkat lunak Orange Data Mining untuk memperkenalkan beberapa algoritma yang umum digunakan.

Setelah melatih model A dan B menggunakan data yang ada, kita bisa langsung menggunakan model-model tersebut untuk membuat ramalan. Misalnya, jika kita ingin mengetahui nilai y ketika x bernilai 10 atau 20, kita hanya perlu memasukkan nilai x ke dalam rumus yang telah dihasilkan oleh model. Dengan cara ini, kita akan mendapatkan hasil prediksi untuk nilai y yang kita inginkan.

$$y = 16 \text{ untuk } x = 10 \quad (4.6)$$

$$y = 26 \text{ untuk } x = 20 \quad (4.7)$$

Selanjutnya kita coba mengganti training data dengan informasi yang bukan angka, misalnya seperti tabel berikut,

x	y
Dia sangat pintar	1
Saya sakit perut	0
Jakarta kota keren	1

Dengan data training di tabel di atas, kita akan memperoleh naluri bahwa jika kita diberikan kalimat

"Yogya kota yang asik"

Secara insting, kita akan tahu bahwa nilai y untuk kalimat tersebut adalah

1

Dalam hal ini, kita tidak perlu memeriksa nilai A dan B dari persamaan $Ax + B = y$. Secara sederhana, y adalah 1 jika kalimat memiliki konotasi positif, dan y adalah 0 jika kalimat memiliki konotasi negatif.

Naluri yang diperoleh dalam persamaan di atas, dalam Artificial Intelligence atau Machine Learning dilakukan dengan melakukan modeling A dan B dari persamaan linear $Ax + B = y$. Oleh karenanya jika data training yang diberikan kurang banyak kemungkinan hasil prediksi y yang diperoleh kurang presisi.

Untuk memperdalam pemahaman kita, mari kita gunakan contoh data cuaca yang sederhana. Bayangkan kita memiliki kumpulan data yang berisi informasi suhu dan kelembaban. Dengan data ini, kita akan mencoba memprediksi jenis cuaca yang akan terjadi. Perlu diingat, lembaga seperti BMKG menggunakan data yang jauh lebih kompleks dan menyeluruh dalam membuat prakiraan cuaca mereka. Namun, contoh sederhana ini akan membantu kita memahami dasar-dasar proses peramalan cuaca.

Suhu	Kelembaban	Cuaca
23C	30%	Cerah
22C	35%	Cerah
16C	90%	Hujan
15C	95%	Hujan

Berdasarkan data pelatihan yang disajikan, kita dapat membuat estimasi kasar mengenai kondisi cuaca di masa mendatang. Data ini terdiri dari dua variabel utama, yaitu suhu dan kelembaban. Dengan menganalisis pola hubungan antara kedua variabel ini dengan kondisi cuaca (hujan atau cerah), kita dapat membuat perkiraan awal tentang kapan suatu wilayah cenderung mengalami hujan atau cuaca cerah.

- Jika suhu 17C kelembaban 95% maka kemungkinan besar cuaca akan hujan.
- Jika suhu 24C kelembaban 20% maka kemungkinan besar cuaca akan cerah.

Kemampuan untuk mengira-ngira oleh manusia dimodelkan dengan parameter A dan B persamaan linear $A \times B = y$ tersebut. Tentunya A dan B menjadi lebih rumit.

Tapi bayangkan jika kita mempunyai data yang berbeda sama sekali dengan data training, maka aplikasi akan bingung. Contoh,

- Jika suhu 22C kelembaban 100% maka aplikasi akan bingung apakah cerah atau hujan.
- Jika suhu 15C kelembaban 10% maka aplikasi akan bingung apakah cerah atau hujan.

Masalah yang sering ditemui dalam pembelajaran mesin adalah bias pada data pelatihan. Bias ini terjadi ketika data yang digunakan untuk melatih model tidak cukup representatif terhadap semua kemungkinan yang bisa terjadi di dunia nyata. Akibatnya, model yang dihasilkan cenderung membuat prediksi atau klasifikasi yang tidak akurat, terutama ketika dihadapkan pada data yang berbeda dari data pelatihannya. Model tersebut seolah-olah "bingung" dalam menentukan keputusan yang tepat karena kurangnya paparan terhadap variasi data yang lebih luas.

Tentu saja, akan lebih baik jika data pelatihan yang diberikan mencakup banyak data sehingga semua kemungkinan dapat tercakup dengan baik. Kemungkinan untuk bias yang salah dalam prediksi menjadi lebih kecil.

Kecerdasan buatan sangat bergantung pada kualitas data yang digunakan untuk melatihnya. Jika data yang digunakan tidak mewakili populasi yang ingin kita teliti, maka model yang dihasilkan akan cenderung bias. Misalnya, melatih model AI untuk memahami perilaku masyarakat Indonesia dengan data dari Amerika Serikat akan menghasilkan hasil yang tidak akurat karena perbedaan budaya yang signifikan. Oleh karena itu, pemilihan data pelatihan yang relevan dan representatif adalah kunci untuk mendapatkan hasil yang akurat dan dapat diandalkan.

Di Indonesia, tantangan dalam pengembangan kecerdasan buatan semakin kompleks karena keterbatasan data pelatihan yang berkualitas, terutama dalam sektor pemerintahan. Kurangnya data digital membuat sulit untuk membangun model AI yang andal. Selain itu, kemampuan manusia dalam memahami data dan merumuskan pertanyaan yang tepat sangat penting sebelum menggunakan AI. Manusia harus mampu menentukan jenis analisis yang ingin dilakukan (misalnya, klasifikasi atau peramalan) berdasarkan data yang tersedia. AI kemudian dapat digunakan untuk meningkatkan akurasi hasil analisis tersebut. Singkatnya, kombinasi antara data berkualitas, kemampuan manusia dalam

memahami data, dan kekuatan komputasi AI adalah kunci untuk pengembangan kecerdasan buatan yang efektif di Indonesia.

Pengelolaan data dan informasi

Koperasi dan UMKM di Indonesia seringkali terkendala oleh pengelolaan data yang tidak efisien. Perbedaan format data, pergantian pengelola program, dan arsip kertas yang belum terdigitalisasi menjadi hambatan dalam mengelola informasi penting bagi pengembangan usaha. Akibatnya, data yang dihasilkan seringkali tidak komprehensif dan sulit dianalisis. Misalnya, data kemiskinan yang hanya menyajikan jumlah orang miskin tanpa rincian lebih lanjut menghambat upaya pengentasan kemiskinan yang tepat sasaran. **Pengelolaan data yang baik sangat krusial bagi Koperasi dan UMKM untuk meningkatkan efektivitas dan efisiensi dalam pengambilan keputusan serta pengembangan usaha.** Dengan data yang akurat dan terintegrasi, Koperasi dan UMKM dapat mengidentifikasi peluang pasar, mengoptimalkan sumber daya, serta menyusun strategi bisnis yang lebih efektif.

Untuk memberikan gambaran secara umum tentang pengelolaan data, berikut adalah tahapan proses pengelolaan data yang baik:

- **Perencanaan:**
 - Tentukan tujuan pengelolaan data.
 - Identifikasi jenis data yang dibutuhkan.
 - Buatlah rencana untuk pengumpulan, penyimpanan, analisis, dan pemanfaatan data.
- **Pengumpulan Data:**
 - Kumpulkan data dari berbagai sumber yang akurat dan terpercaya.
 - Pastikan data yang dikumpulkan lengkap, konsisten, dan up-to-date.
 - Gunakan metode pengumpulan data yang sesuai dengan jenis data dan sumbernya.
- **Penyimpanan Data:**
 - Simpan data di tempat yang aman dan terjamin.

- Gunakan infrastruktur penyimpanan data yang handal dan scalable.
- Terapkan kontrol akses untuk melindungi data dari unauthorized access.
- **Pembersihan dan Pengolahan Data:**
 - Bersihkan data dari kesalahan, duplikasi, dan inkonsistensi.
 - Proses data sesuai dengan kebutuhan analisis.
 - Gunakan tools dan teknik yang tepat untuk pembersihan dan pengolahan data.
- **Analisis Data:**
 - Analisis data untuk menemukan pola, trend, dan insight.
 - Gunakan teknik analisis data yang sesuai dengan jenis data dan tujuan analisis.
 - Visualisasikan data untuk memudahkan pemahaman.
- **Pemanfaatan Data:**
 - Gunakan hasil analisis data untuk mendukung pengambilan keputusan.
 - Kembangkan produk dan layanan yang berbasis data.
 - Tingkatkan efisiensi dan efektivitas operasi.
- **Pemeliharaan Data:**
 - Perbarui data secara berkala.
 - Pantau kualitas data.
 - Sesuaikan proses pengelolaan data sesuai dengan kebutuhan.
- **Tool yang Diperlukan:**
 - **Perangkat lunak:** Microsoft Excel, Tableau, Power BI, SPSS, R, Python, dll.
 - **Infrastruktur:** Server, storage, network, dll.
 - **Sumber daya manusia:** Data scientist, data analyst, data engineer, dll.
- **Hal Penting yang Harus Diperhatikan:**

- **Keamanan data:** Lindungi data dari *unauthorized access*, *modification*, dan *destruction*.
- **Kualitas data:** Pastikan data akurat, lengkap, konsisten, dan *up-to-date*.
- **Etika data:** Gunakan data secara bertanggung jawab dan sesuai dengan peraturan yang berlaku.
- **Keterbukaan data:** Berikan akses data kepada publik sesuai dengan prinsip open data.
- Budaya data: Ciptakan budaya organisasi yang menghargai dan memanfaatkan data.

Proses pengelolaan data yang baik akan memberikan manfaat signifikan bagi Kementerian Koperasi dan UKM, pemerintah, serta dunia usaha di Indonesia. Dengan penerapannya, diharapkan dapat meningkatkan kinerja dan produktivitas dalam berbagai kegiatan, meningkatkan kualitas layanan yang diberikan kepada masyarakat, mendorong lahirnya ide-ide kreatif dan inovatif, serta membangun pondasi ekonomi yang lebih kuat dan berbasis data.

Pengambilan keputusan dan kebijakan

Dengan mengadopsi pendekatan Data-Driven Decision Making (DDDM), organisasi dapat meningkatkan efisiensi operasi, mengoptimalkan alokasi sumber daya, dan membuat keputusan yang lebih cepat dan tepat. Hal ini memungkinkan organisasi untuk lebih adaptif terhadap perubahan pasar dan persaingan yang semakin ketat. Berikut tahapannya:

- **Menentukan Tujuan:**
 - Pastikan tujuan yang ingin dicapai jelas dan terukur.
 - Contoh: Mencari program yang cocok untuk kebutuhan masyarakat, Mengurangi tingkat kemiskinan, Mengurangi tingkat pengangguran, Meningkatkan kepuasan pelanggan, mengurangi tingkat turnover karyawan, dll.
- **Mengidentifikasi dan Mengumpulkan Data:**
 - Pilih data yang relevan dengan tujuan yang telah ditetapkan.
 - Data dapat berasal dari berbagai sumber internal maupun eksternal.

- Pastikan data akurat, lengkap, dan terpercaya.
- **Mengolah dan Menganalisis Data:**
 - Gunakan tool analisis data yang sesuai untuk mengolah dan menganalisis data yang telah dikumpulkan.
 - Contoh tool: Microsoft Excel, Tableau, Power BI, dll.
 - Lakukan analisis untuk menemukan pola, trend, dan wawasan dari data.
- **Menarik Kesimpulan dan Memutuskan Kebijakan:**
 - Berdasarkan hasil analisis, tarik kesimpulan yang mendukung pencapaian tujuan.
 - Gunakan kesimpulan tersebut untuk membuat keputusan dan merumuskan kebijakan.
 - Pastikan keputusan dan kebijakan terukur, realistis, dan dapat diimplementasikan.
- **Menerapkan dan Memantau Kebijakan:**
 - Implementasikan kebijakan yang telah dibuat dengan jelas dan terstruktur.
 - Lakukan monitoring dan evaluasi untuk mengukur efektivitas kebijakan.
 - Lakukan penyesuaian kebijakan jika diperlukan berdasarkan hasil monitoring.
- **Pertimbangan Penting:**
 - **Kualitas Data:** Pastikan data yang digunakan berkualitas tinggi untuk menghasilkan analisis yang akurat.
 - **Keterampilan Analisis:** Memiliki tim yang terampil dalam menganalisis data sangatlah penting.
 - **Komunikasi:** Komunikasikan hasil analisis dan keputusan kepada seluruh pemangku kepentingan.
 - **Budaya Berbasis Data:** Ciptakan budaya organisasi yang menghargai dan memanfaatkan data dalam pengambilan keputusan.

Penerapan DDDM dapat membantu organisasi dalam meningkatkan efektivitas dan efisiensi, meningkatkan profitabilitas, dan mencapai tujuan strategis dengan lebih optimal.

Tata kelola dan akuntabilitas

Teknologi kecerdasan buatan (AI) dan pembelajaran mesin (ML) menawarkan peluang besar bagi koperasi dan UMKM di Indonesia. Dengan otomatisasi tugas-tugas rutin, seperti pembukuan dan pengelolaan inventori, AI dapat membantu koperasi dan UMKM meningkatkan efisiensi operasional. Selain itu, analisis data yang mendalam melalui AI memungkinkan identifikasi peluang pasar baru, optimalisasi rantai pasok, serta pengambilan keputusan bisnis yang lebih baik. Misalnya, AI dapat membantu UMKM memprediksi permintaan pasar, sehingga dapat mengelola produksi dan persediaan dengan lebih efektif.

Penerapan AI juga dapat meningkatkan transparansi dan akuntabilitas dalam pengelolaan koperasi. Dengan menggunakan AI untuk melacak transaksi dan kinerja keuangan, anggota koperasi dapat dengan mudah memantau penggunaan dana dan memastikan pengelolaan yang baik. Selain itu, AI dapat membantu mendeteksi potensi penyalahgunaan dana atau tindakan yang tidak sesuai dengan aturan koperasi. Hal ini akan memperkuat kepercayaan anggota terhadap pengurus koperasi.

Meskipun menawarkan banyak manfaat, penerapan AI di koperasi dan UMKM juga menghadapi sejumlah tantangan. Keterbatasan akses terhadap teknologi dan data, serta kurangnya sumber daya manusia yang kompeten dalam bidang AI, menjadi hambatan utama. Namun, dengan dukungan pemerintah, pengembangan infrastruktur digital, dan pelatihan yang tepat, tantangan ini dapat diatasi. Pemerintah perlu mendorong adopsi AI di kalangan koperasi dan UMKM melalui program-program inkubasi, fasilitasi akses pendanaan, serta penyediaan platform digital yang terintegrasi. Dengan demikian, AI dapat menjadi katalisator pertumbuhan dan pengembangan koperasi dan UMKM di Indonesia.

BAB 5: Aplikasi Orange Data Mining

Instalasi Orange Data Mining

Bagi koperasi dan UMKM yang ingin memulai analisis data namun memiliki keterbatasan teknis, terdapat sejumlah tool analisis data yang user-friendly. Beberapa contohnya adalah Rapidminer, Microsoft Power BI, dan Tableau. Meskipun tool ini menawarkan kemudahan dalam penggunaan, sayangnya mereka umumnya memerlukan lisensi berbayar. Untuk alternatif yang lebih terjangkau dan fleksibel, koperasi serta UMKM dapat mempertimbangkan penggunaan bahasa pemrograman seperti Python atau R, yang dilengkapi dengan library seperti TensorFlow dan Keras. Namun, penggunaan tool ini membutuhkan pemahaman yang lebih mendalam tentang konsep machine learning dan artificial intelligence.

Salah satu tool favorit penulis yang sifatnya open source, free, sederhana dapat digunakan dengan mudah adalah

Orange Data Mining

Dalam buku ini juga akan digunakan tool Process Mining for Python PM4PY untuk melakukan analisa bisnis proses.

Pada kesempatan ini kita akan lebih banyak menggunakan tool Orange untuk melakukan analisa data. Orange tersedia untuk sistem operasi,

- Windows
- MacOS
- Linux

Instalasi di Windows dan Mac OS sangat sederhana sekali, langkahnya adalah,

- Buka alamat URL <https://orangedatamining.com/download/>
- Download file *exe (Windows) atau *dmg (Mac OS). Pada saat buku ini ditulis file exe Windows adalah Orange versi 3.36.2 sebagai berikut,

Orange3-3.36.2-Miniconda-x86_64.exe

- Jalankan file tersebut untuk menginstalasi aplikasi Orange Data Mining.

Alternatif cara instalasi di Windows yang mungkin lebih mudah adalah sebagai berikut,

- Buka alamat URL <https://orangedatamining.com/download/>
- Download file *zip (Windows)
- Buka file *zip
- Orange dapat di jalankan tanpa di install.

Untuk menjalankan orange di Windows atau di Mac OS relatif sederhana, cukup dengan mengklik ICON aplikasi orange yang ada di desktop.

Bagi anda yang menggunakan Linux, cara menginstalnya tidak se-manusiawi Windows. Sebetulnya ada beberapa cara, akan tetapi cara yang manjur untuk Linux yang biasa penulis lakukan adalah sebagai berikut,

Pertama2 kita perlu menginstal Instal Pip dan Virtualenv

```
sudo apt update
sudo apt install build-essential python3-dev \

python3-pip python3-virtualenv
```

Selanjutnya membuat lingkungan virtual Python dan mengaktifkan lingkungan virtual-nya sebagai berikut,

```
mkdir ~/orangenv
cd ~/orangenv
virtualenv --python=python3 venv
source venv/bin/activate
```

Setelah lingkungan virtual selesai diinstalasi kita bisa menginstal Orange dan aplikasi pendukungnya melalui perintah berikut,

```
sudo apt install '^libxcb.*-dev' libx11-xcb-dev \
libglu1-mesa-dev libxrender-dev libxi-dev \
libxkbcommon-dev libxkbcommon-x11-dev
```

```
sudo apt install libxcb-cursor0 xcb libxcb-xinerama0
pip install pyqt5 pyqt6 pyside6 pywayland
pip install orange3
```

Untuk menjalankan orange, jika virtual environment masih aktif cukup dengan menulis,

```
orange-canvas
```

untuk keluar dan menonaktifkan virtual environment dapat dilakukan dengan mengklik,

```
deactivate
```

Untuk menjalankan setiap kali harus mengaktifkan lingkungan virtual melalui perintah berikut,

```
cd orangenv
source venv/bin/activate
```

Walaupun mungkin tidak reliabel seperti sebelumnya, kita dapat menggunakan alternatif instalasi Orange Data Mining di Linux di bantu oleh anaconda melalui perintah berikut,

Disini kita menggunakan Anaconda untuk mendukung operasi Orange Data Mining. Pertama download Anaconda dari <https://www.anaconda.com/download/> Sebagai user biasa, ketik,

```
cd ~/Downloads/
bash Anaconda3-2023.09-0-Linux-x86_64.sh
```

Setelah selesai menjalankan script anaconda, jalankan

```
source ~/.bashrc
```

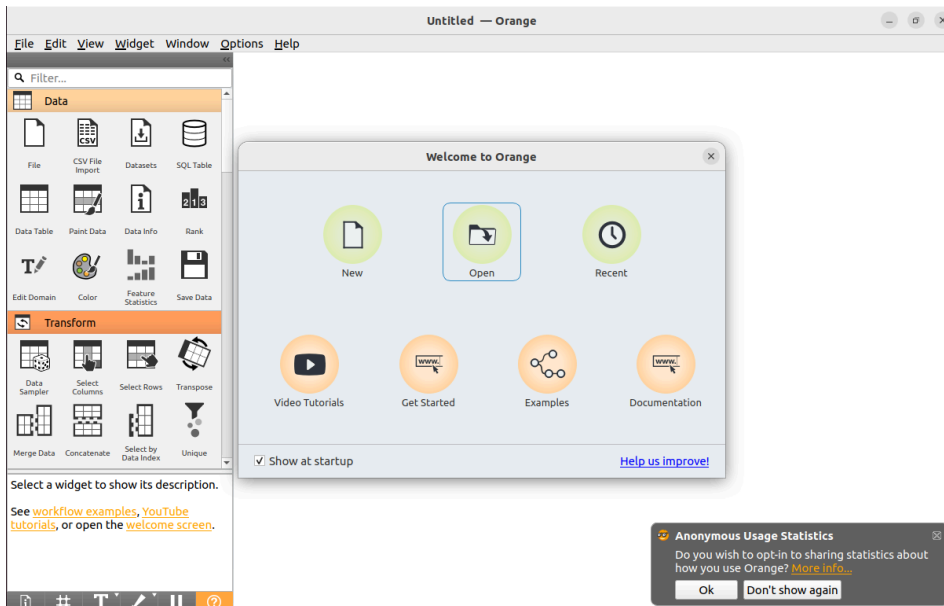
Setelah anaconda selesai di instalasi, selanjutnya kita dapat menginstal orange melalui perintah berikut,

```
conda config --add channels conda-forge
conda config --set channel_priority strict
conda create python=3 --yes --name orange3
conda activate orange3
conda install orange3
conda update -n base -c defaults conda
```

Penggunaan anaconda tampaknya akan menyederhanakan proses instalasi dan operasi orange data mining di Linux. Akan tetapi pengalaman penulis menunjukan bahwa penggunaan proses instalasi manual tampaknya lebih reliable.

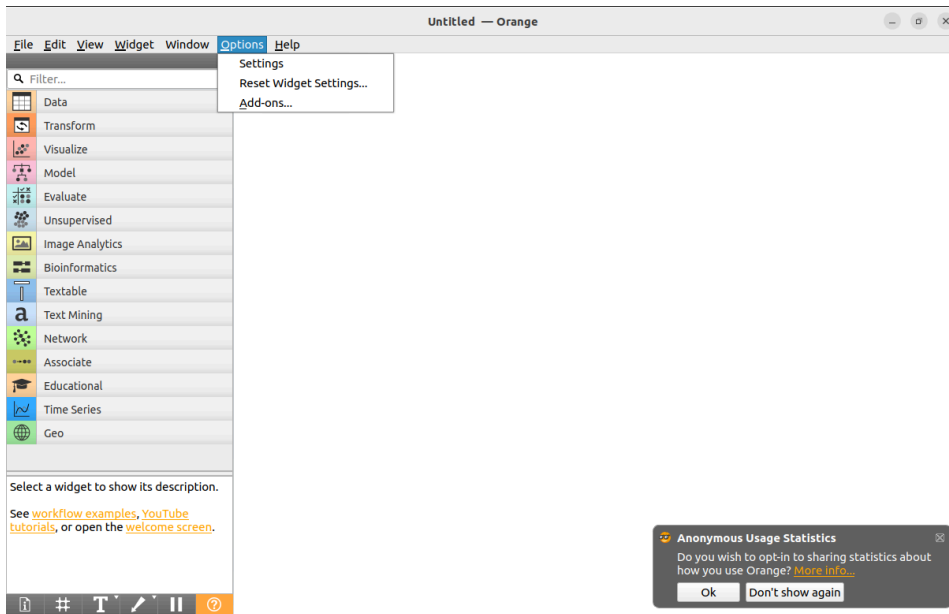
Menjalankan Orange Data Mining

Ketika kita menjalankan Orange Data Mining, antarmuka pengguna yang ditampilkan seperti pada Gambar 5.1. Panel sebelah kiri menyajikan beragam widget yang siap digunakan untuk berbagai tahapan analisis data. Kelompok widget ini umumnya mencakup data input, transformasi data, visualisasi, pemodelan, evaluasi, dan analisis tanpa pengawasan. Fleksibilitas widget-widget ini memungkinkan kita untuk menjalankan berbagai teknik analisis data yang umum digunakan dalam ranah kecerdasan buatan, pembelajaran mesin, dan bahkan beberapa aspek pembelajaran mendalam, baik untuk masalah supervised maupun unsupervised learning.



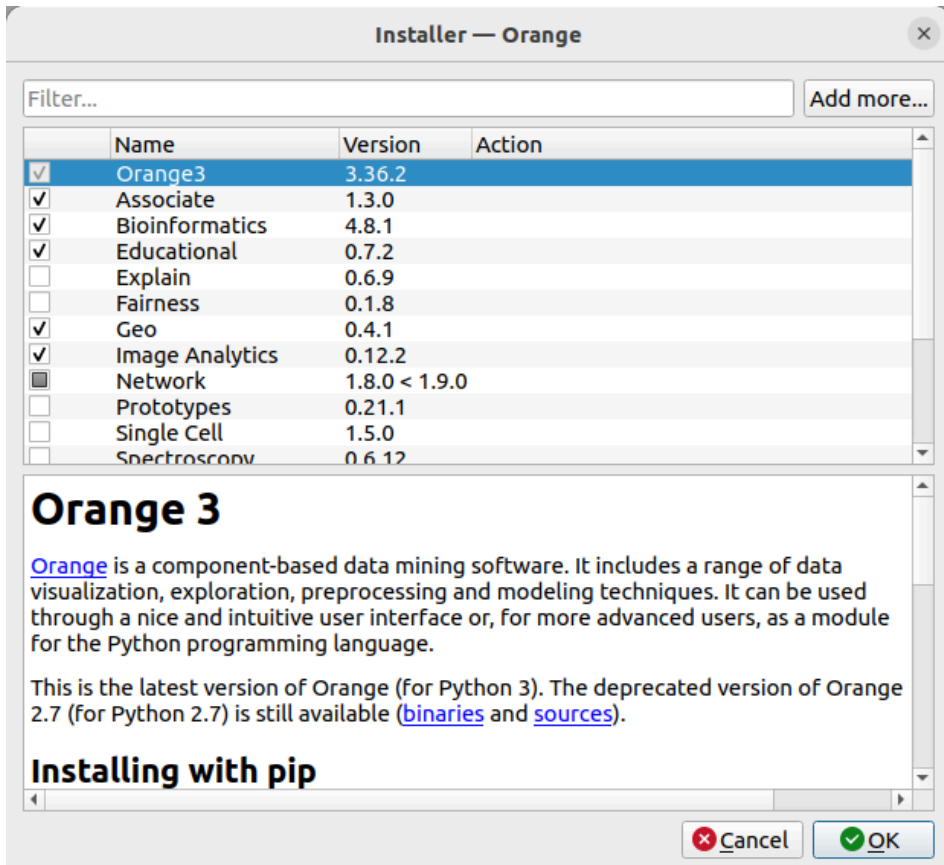
Gambar 5.1. Tampilan Orange Saat Jalan Pertama Kali.

Antarmuka Orange Data Mining menyediakan menu-menu administratif di bagian atas seperti File, Edit, dan Help. Meskipun menu-menu ini penting untuk pengaturan umum, pengguna cenderung lebih sering berinteraksi dengan menu Options saat melakukan analisis data. Sementara itu, area 'Welcome to Orange' di bagian tengah, khususnya menu Examples, menyajikan berbagai contoh kasus yang sangat berharga bagi pengguna baru untuk memulai eksplorasi fitur-fitur Orange.



Gambar 5.2. Tampilan saat mengakses menu Options untuk Add-ons

Gambar memperlihatkan saat kita memilih menu Options untuk menambahkan Add-ons. Ini sangat dibutuhkan bagi anda yang ingin memaksimalkan penggunaan orange data mining untuk menganalisa lebih dalam berbagai jenis data. Melalui menu Add-ons kita dapat menambahkan berbagai aplikasi tambahan untuk menganalisa berbagai hal yang menarik.

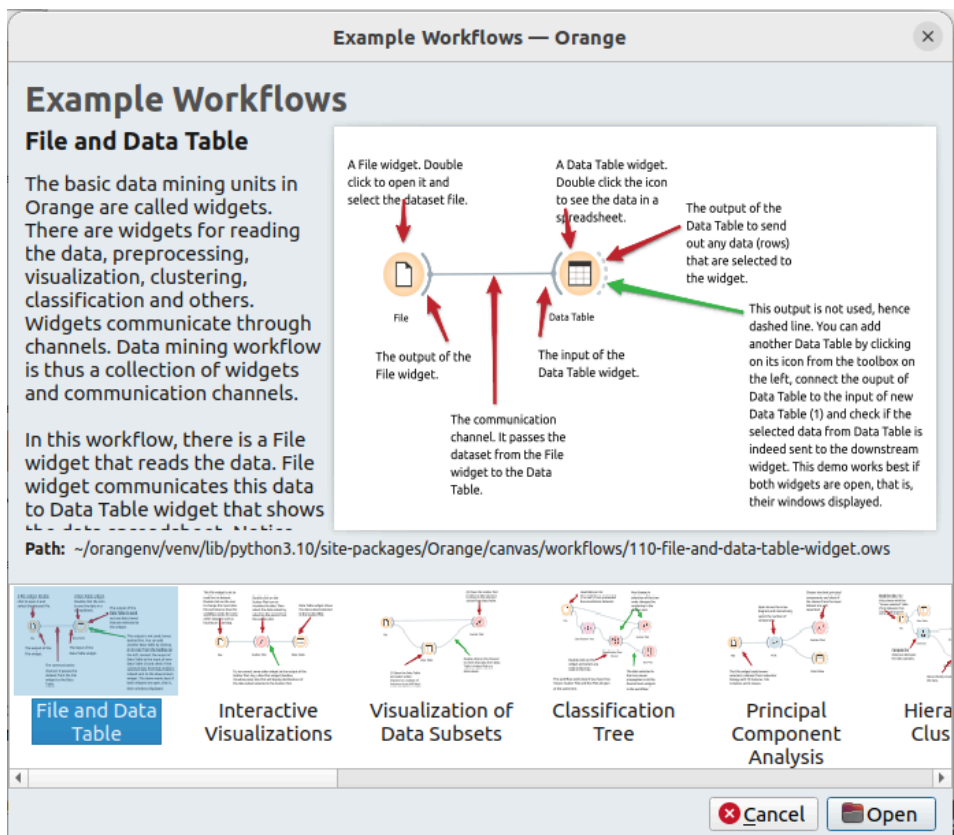


Gambar 5.3. Tampilan Menu Add-ons

Gambar memperlihatkan tampilan menu add-ons. Disini kita bisa melihat nama dari add-ons dan keterangan dari berbagai add-ons di bawahnya. Ada cukup banyak add-ons yang dapat di tambahkan di orange data mining, penulis biasanya akan menginstalasi tambahan add-ons berikut,

- Bioinformatics - untuk memberikan akses ke data-data bioinformatics seperti GEO data set, GO, dan KEGG.
- Educational - untuk menambahkan fasilitas untuk pemula dalam mendalami data mining. Juga berguna bagi pengajar di kelas.
- Geo - add-ons yang akan membantu dalam bidang geography, geolocation termasuk visualisasinya di layar.
- Image Analytics - add-ons untuk me-load gambar, hingga mengirimkan gambar untuk analisa deep-learning.

- Network - add-ons untuk analisis dan visualisasi jaringan data. Ini biasanya menggunakan graph.
- Text - add-ons untuk memudahkan akses ke text yang tersedia bebas di Internet seperti NY Time, Twitter, Wikipedia. Disamping itu juga di siapkan beberapa tool untuk melakukan vektorisasi text hingga visualisasi text dalam word cloud atau geo map.
- Textable - add-ons melakukan analisa yang lebih advanced lagi terhadap text yang di oleh menu text.



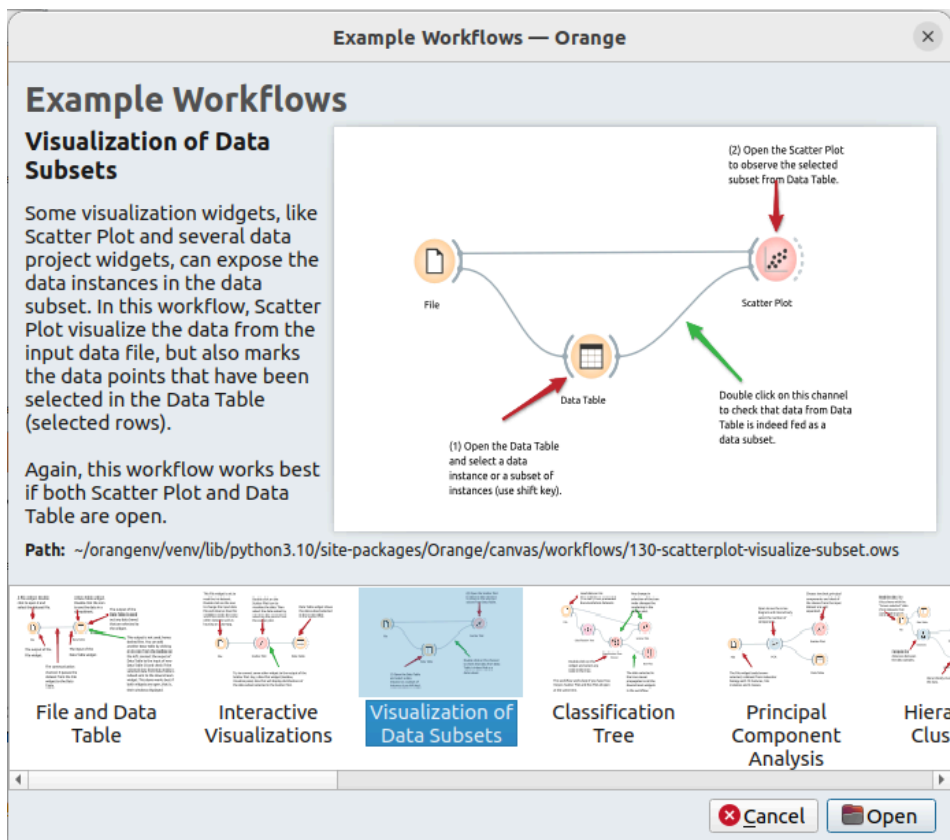
Gambar 5.4. Menu Example Workflows di “Welcome to Orange”

Gambar diatas memperlihatkan tampilan menu Example Workflows yang diakses dari menu Welcome to Orange saat orange data mining pertama kali dijalankan. Slider di bawah akan terlihat banyak pilihan contoh aplikasi yang dalam dicoba untuk belajar orange data mining. Aplikasi dalam orange biasanya dikenal sebagai Orange Workflow. Pada gambar di atas adalah workflow paling sederhana yang berisi dua (2) widget yaitu untuk

membaca file data dan menampilkan data dalam bentuk tabel. Dalam contoh aplikasi ini tidak ada sama sekali proses analisa data. modelling data atau proses yang lebih rumit.

Contoh Penggunaan Orange

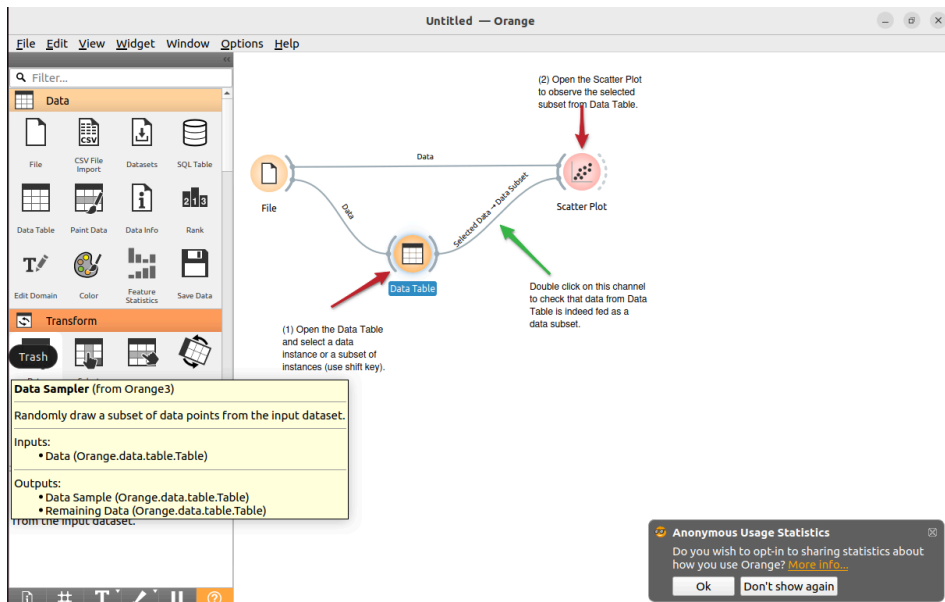
Pada gambar di atas di perlihatkan beberapa contoh aplikasi orange, seperti, File and Data Table, Interactive Visualizations, Visualization of Data Subsets, Classification Tree dan masih banyak lagi.



Gambar 5.5. Aplikasi Visualization of Data Subsets.

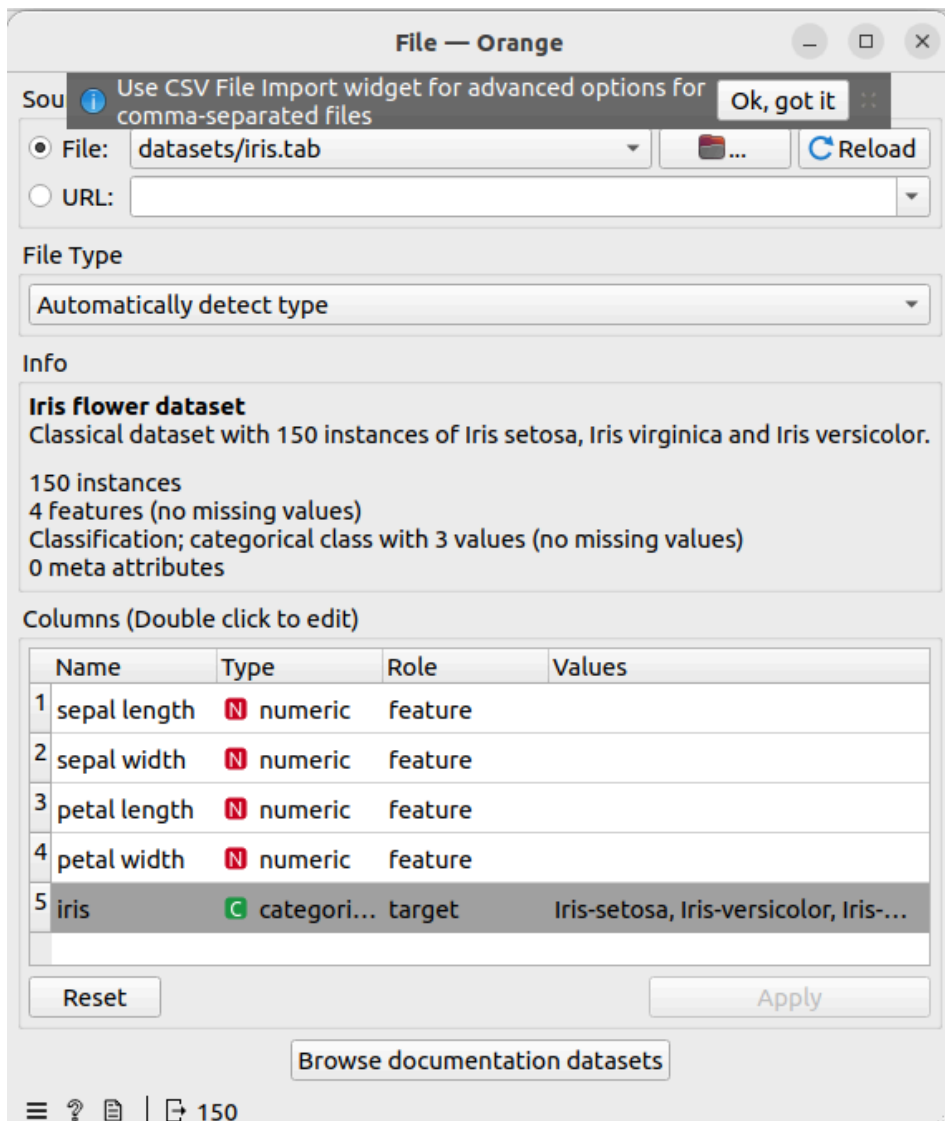
Untuk memberikan gambaran dalam menjalankan contoh orange workflow, kita akan menjalankan contoh Virtualization of Data Subsets. Kira-kira kisah yang akan dikerjakan oleh workflow tersebut adalah sebagai berikut, workflow ini pada dasarnya berusaha menampilkan dalam bentuk gambar scatter plot dari data-data yang kita pilih (select) di

widget data table. Inti dari workflow ini menjelaskan bahwa antar widget bisa saling interaksi dengan baik.



Gambar 5.6. Saat Workflow Visualization of Data Subsets di load ke orange.

Gambar 5.6. memperlihatkan saat workflow visualization of data subset di load / dimasukan ke dalam orange. Pada saat di load, semua widget sebetulnya akan langsung menjalankan tugas-nya. Misalnya, widget file akan membaca file yang di set untuk dibaca. Widget Data Table akan meload data yang dibaca oleh widget file. Sementara, Widget Scatter Plot akan meload semua data dalam bentuk scatter plot dan menunggu pilihan data yang dipilih oleh pengguna dari widget data table. Yang perlu kita lakukan hanya mengklik widget yang kita inginkan untuk lebih mengaktifkan lagi workflow yang sudah operasional ini.



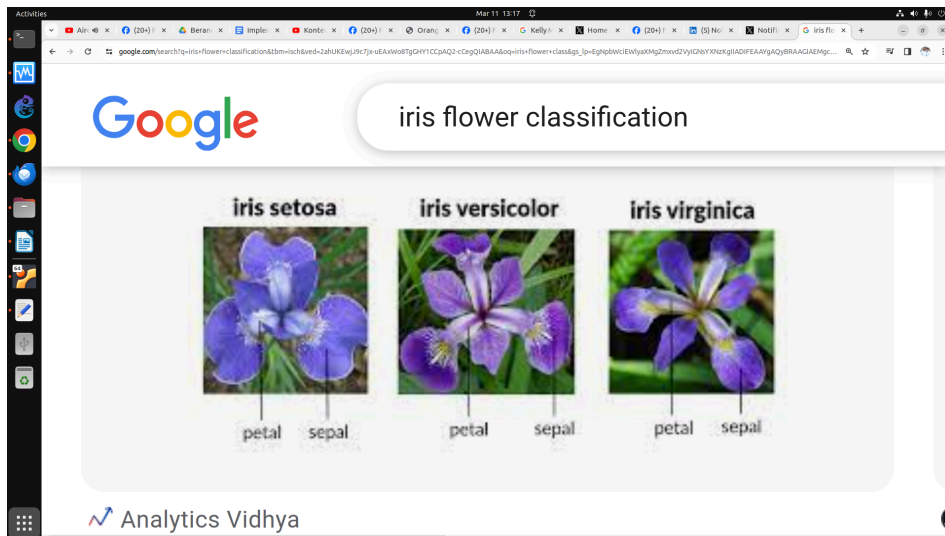
Gambar 5.7. Tampilan Widget File.

Gambar 5.7. memperlihatkan tampilan widget file sesudah kita klik. Terlihat di bagian paling atas bahwa file yang diakses adalah

datasets/iris.tab

Isi file tersebut adalah data-data hasil pengukuran bunga iris ada dua komponen dari bunga Iris, yaitu, (1) petal, dan (2) sepal, kita ukur panjang

dan lebarnya. Ukuran panjang dan lebar dari petal dan sepal akan menentukan klasifikasi bunga iris yang terdiri dari tiga (3) jenis bunga, yaitu, iris setosa, iris versicolor dan iris virginica. Tampilan bunga iris dapat dilihat pada gambar 5.8.



Gambar 5.8. Tampilan Bunga Iris Setosa, Iris Versicolor dan Iris Virginica. Terlihat jelas Sepal dan Petal dari masing-masing bunga.

File iris.tab berisi ukuran panjang, lebar dari sepal dan petal beserta jenis bunga iris-nya. Data dalam bentuk kolom di pisahkan dengan spasi, sehingga nama file yang digunakan menggunakan kode .tab. Setelah data di baca oleh Widget File, maka data bisa dikirimkan ke Widget yang lain untuk diproses lebih lanjut. Salah satu proses yang paling sederhana adalah dengan mengirimkan ke Widget Table untuk melihat data yang ada, seperti ditampilkan pada Gambar 5.9.

	iris	sepal length	sepal width	petal length	petal w
1	Iris-setosa	5.1	3.5	1.4	
2	Iris-setosa	4.9	3.0	1.4	
3	Iris-setosa	4.7	3.2	1.3	
4	Iris-setosa	4.6	3.1	1.5	
5	Iris-setosa	5.0	3.6	1.4	
6	Iris-setosa	5.4	3.9	1.7	
7	Iris-setosa	4.6	3.4	1.4	
8	Iris-setosa	5.0	3.4	1.5	
9	Iris-setosa	4.4	2.9	1.4	
10	Iris-setosa	4.9	3.1	1.5	
11	Iris-setosa	5.4	3.7	1.5	
12	Iris-setosa	4.8	3.4	1.6	
13	Iris-setosa	4.8	3.0	1.4	
14	Iris-setosa	4.3	3.0	1.1	
15	Iris-setosa	5.8	4.0	1.2	
16	Iris-setosa	5.7	4.4	1.5	
17	Iris-setosa	5.4	3.9	1.3	
18	Iris-setosa	5.1	3.5	1.4	
19	Iris-setosa	5.7	3.8	1.7	
20	Iris-setosa	5.1	3.8	1.5	

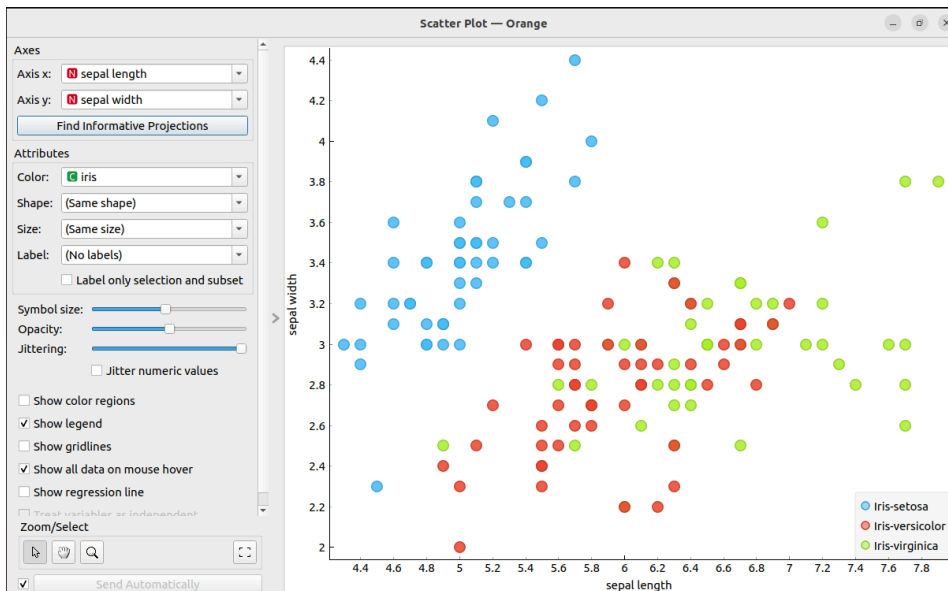
Gambar 5.9. Tampilan Widget Data Table.

Pada Widget Data Table di perlihatkan informasi lebih lengkap dengan data yang kita baca, beberapa ringkasan dari data yang kita load dari file, seperti,

- Ada 150 data / instances yang dalam file tersebut. Dan tidak ada data yang kosong.
- Ada empat (4) fitur dalam data, dalam hal ini panjang sepal, lebar sepal, panjang petal, dan lebar petal.
- Ada tiga (3) target dalam data, ini merupakan klasifikasi masing-masing data, yaitu iris setosa, iris versicolor dan iris virginica.

Kita perlu ingat bahwa data yang sudah diklasifikasikan biasanya digunakan untuk jenis pembelajaran yang disebut *supervised learning*. Dalam metode ini, kita sebagai manusia yang memberikan tanda atau label pada data pelatihan untuk menunjukkan kategori masing-masing data. Karena adanya peran manusia dalam memberikan label inilah maka disebut *supervised* atau "diawasi".

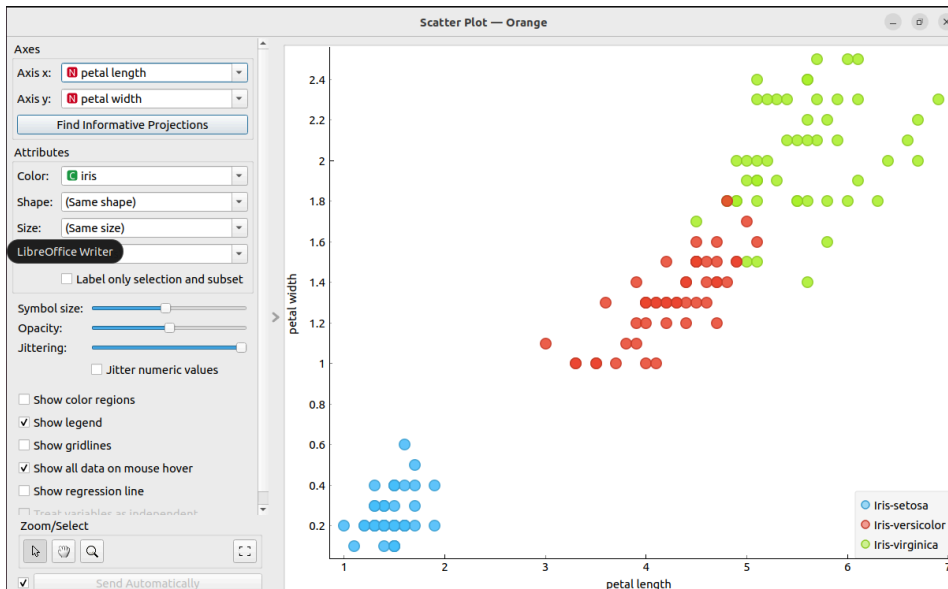
Berbeda halnya dengan *unsupervised learning*. Pada jenis pembelajaran ini, data pelatihan tidak memiliki label kategori. Komputer akan secara otomatis mengelompokkan data berdasarkan kesamaan fitur yang dimiliki. Meskipun begitu, peran manusia tetap penting untuk memeriksa hasil pengelompokan yang dilakukan oleh komputer. Hal ini dikarenakan tidak ada jaminan bahwa hasil pengelompokan yang dilakukan komputer selalu akurat.



Gambar 5.10. Scatter Plot data iris.

Dengan menghubungkan data dari Widget File ke Widget Scatter Plot, kita bisa melihat data dalam bentuk grafik titik-titik (scatter plot). Contohnya, Gambar 5.10 menampilkan data bunga iris. Setiap jenis bunga memiliki warna yang berbeda: biru untuk setosa, merah untuk versicolor, dan hijau untuk virginica. Kita bisa memilih data mana yang akan ditampilkan di sumbu X dan Y. Pada gambar tersebut, sumbu X menunjukkan panjang sepal, sedangkan sumbu Y menunjukkan lebar sepal.

Dari Gambar 5.10, kita bisa lihat dengan jelas bahwa bunga iris setosa mudah dibedakan dari dua jenis lainnya hanya dengan melihat panjang dan lebar sepalnya. Namun, bunga iris versicolor dan virginica sulit dibedakan karena titik-titik datanya saling tumpang tindih..



Gambar 5.11. Scatter Plot data Iris menggunakan Data Petal sebagai sumbu x dan y.

Ketika kita mengganti data yang digunakan untuk membandingkan jenis bunga iris dari ukuran sepal menjadi ukuran petal (seperti yang ditunjukkan pada Gambar 5.11), kita melihat bahwa bunga iris Setosa menjadi sangat mudah dibedakan dari jenis bunga iris lainnya. Yang menarik, bunga iris versicolor dan iris virginica juga bisa dibedakan dengan cukup baik, meskipun ada beberapa yang terlihat mirip. Namun, secara umum, ketiga jenis bunga iris ini dapat dibedakan dengan mudah menggunakan data ukuran petal.

Dengan mengamati data secara visual, kita bisa dengan mudah memahami bagaimana membedakan jenis-jenis bunga iris. Namun, dalam machine learning, komputerlah yang akan melakukan tugas pengelompokan ini tanpa bantuan kita. Jika manusia saja kesulitan membedakan suatu hal, maka komputer pun akan kesulitan. Sebaliknya, jika manusia mudah membedakannya, komputer juga cenderung akan mudah. Pemahaman ini sangat penting untuk kita yang ingin mempelajari bidang kecerdasan buatan, baik itu machine learning maupun deep learning.

BAB 6: Contoh Prediksi Sederhana

Seringkali, manusia terdorong untuk melakukan praktik tanpa terlebih dahulu mendalami landasan teoritisnya. Dalam rangka menjawab ketertarikan tersebut, bagian ini akan memberikan ilustrasi sederhana tentang bagaimana meramalkan anggaran masa depan dengan memanfaatkan data anggaran sebelumnya.

Coba Prediksi Anggaran

Misalkan kita mempunyai file berisi history anggaran sebuah kementerian / lembaga yang sangat sederhana sekali seperti tampak pada tabel berikut, misalnya,

Tabel 6.1. Contoh History Anggaran

Tahun	Anggaran
2018	100
2019	200
2020	300
2021	400
2022	500
2023	600

Dari tabel di atas, secara sederhana tanpa menggunakan program apapun, kita dapat menebak bahwa anggaran tahun 2024 adalah,

700

Mari kita lakukan sesuatu yang sangat sederhana dengan ORANGE data mining agar proses perdukunan ini bisa dilakukan menggunakan komputer. Pertama-tama data contoh history anggaran kita masukan ke semua file Excel sebagai berikut,

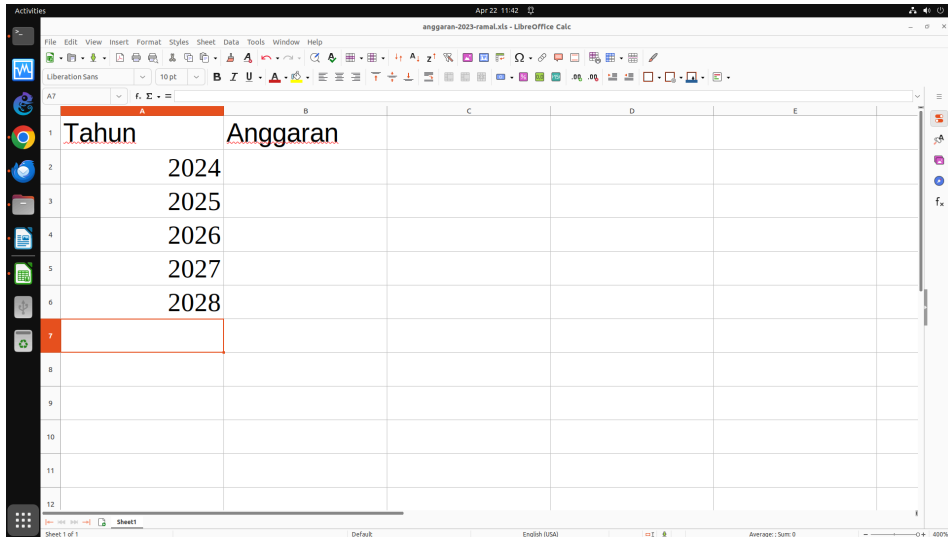
Tahun	Anggaran
2018	100
2019	200
2020	300
2021	400
2022	500
2023	600

Gambar 6.1. File anggaran-2023-history.xls

Pada gambar di atas adalah screenshot file anggaran-2023-history.xls isinya persis seperti tabel di atas.

Untuk melakukan peramalan anggaran pada tahun 2024, 2025, dan seterusnya, kita perlu menyiapkan sebuah file data baru. File ini akan berfungsi sebagai data masukan untuk model peramalan. Format file ini serupa dengan file-file data historis yang telah kita gunakan sebelumnya. Bedanya, pada file baru ini, kita akan mengisi kolom tahun dengan rentang tahun yang ingin kita ramalkan anggarannya.

Kolom anggaran pada file baru ini memiliki dua opsi pengisian. Pertama, kita dapat mengisi kolom anggaran dengan nilai perkiraan anggaran berdasarkan analisis atau pengalaman kita. Kedua, kita dapat membiarkan kolom anggaran kosong. Jika kita memilih opsi kedua, maka sistem peramalan akan secara otomatis mencoba memprediksi nilai anggaran untuk tahun-tahun tersebut berdasarkan pola data historis yang telah ada. Sebagai contoh, kita dapat menamai file baru ini dengan nama "anggaran-2023-ramal.xls".

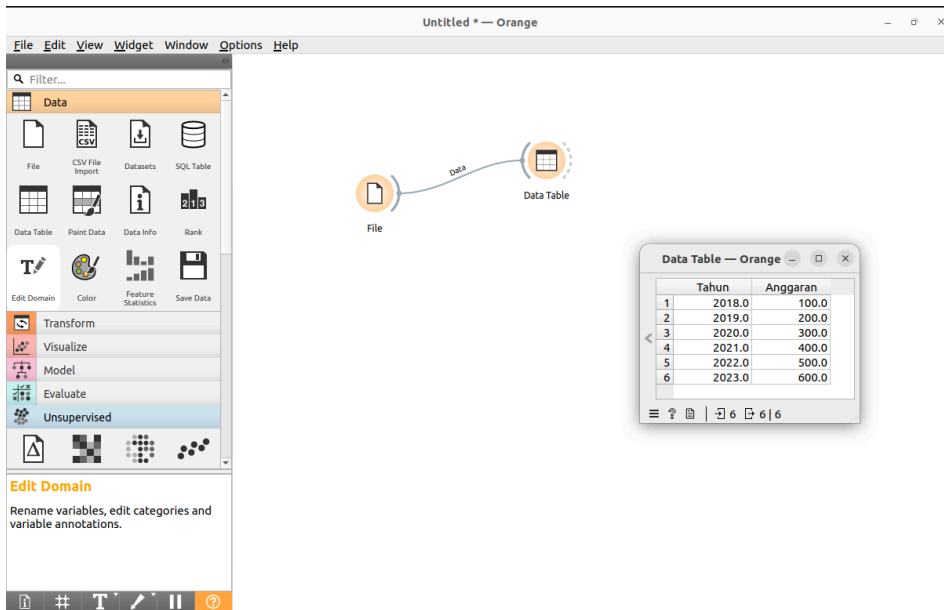


Gambar 6.2. File anggaran-2023-ramal.xls

Dalam gambar di atas dibuat file anggaran-2023-ramal.xls yang berisi tahun 2024 s/d 2028 dengan anggaran yang dikosongkan. Jadi kita mempunyai dua (2) file,

- anggaran-2023-history.xls
- anggaran-2023-ramal.xls

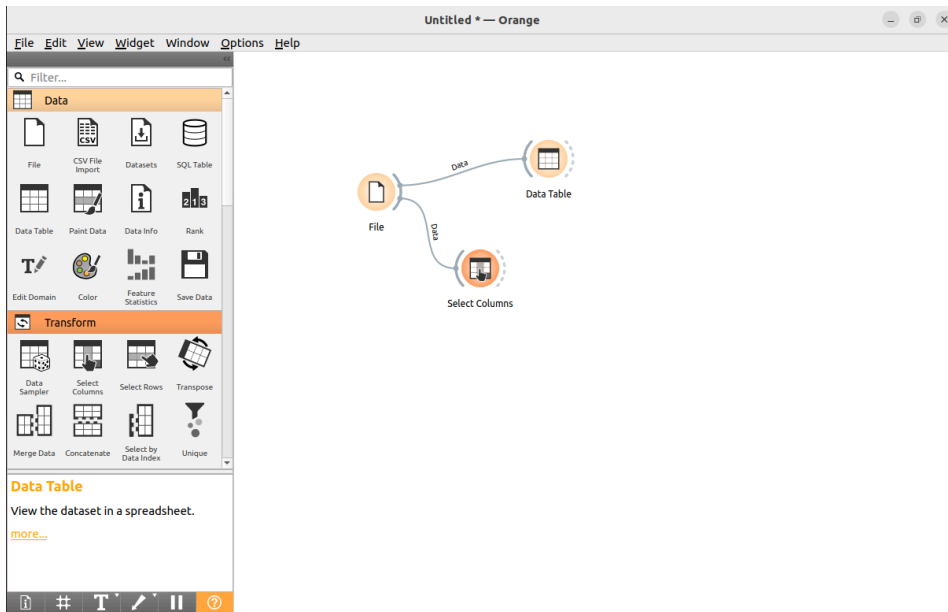
Mari jalankan ORANGE data mining dan load anggaran-2023-history.xls di dalam ORANGE. Adapun program yang kita buat untuk hal tersebut tampak pada gambar berikut,



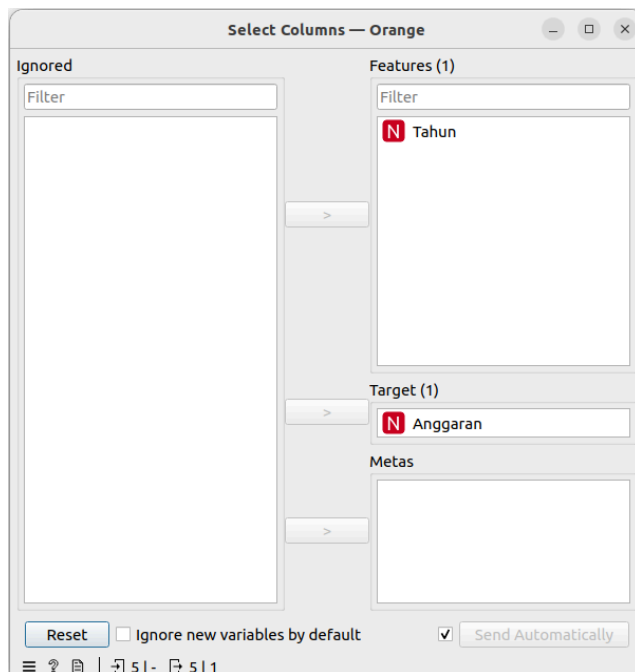
Gambar 6.3. file anggaran-2023-history.xls dibaca ORANGE.

Kita mulai dengan mengimpor data anggaran dari file Excel 'anggaran-2023-history.xls' ke dalam Orange menggunakan widget File. Untuk memastikan data terbaca dengan benar, kita langsung tampilkan dalam bentuk tabel menggunakan widget Data Table. Hasilnya, kita bisa melihat data anggaran kita sudah siap untuk dianalisis di Orange.

Sekarang, kita perlu memberitahu Orange mana data yang ingin kita jadikan sebagai variabel yang akan kita prediksi (target) dan mana yang akan kita gunakan sebagai dasar prediksi (fitur). Dalam kasus ini, kita ingin memprediksi 'Anggaran' berdasarkan 'Tahun'. Jadi, kita gunakan widget Select Columns untuk memilih kedua kolom ini dan menghubungkannya dengan data yang sudah kita impor.



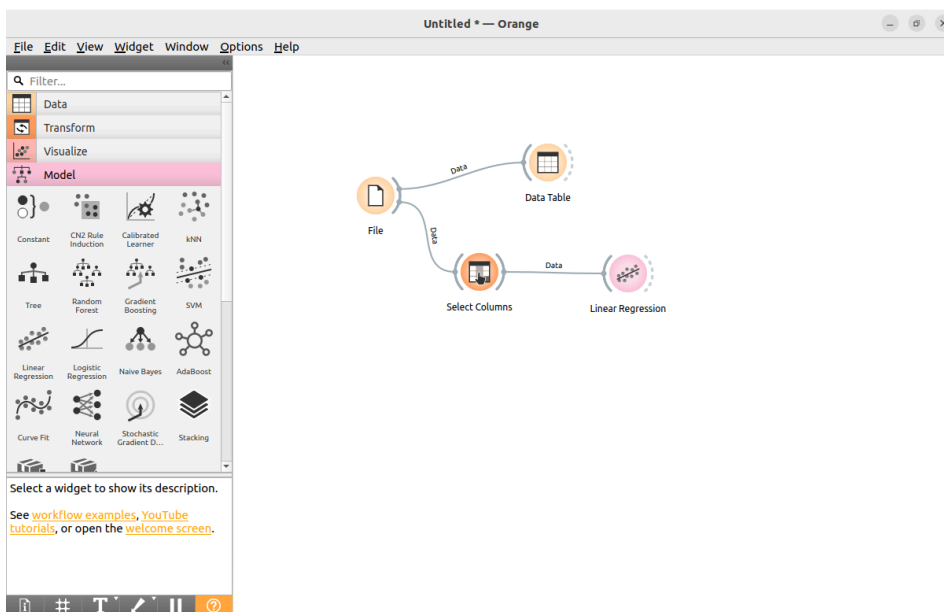
Gambar 6.4. Widget File di sambungkan ke Widget Select Columns



Gambar 6.5. Tampilan Widget Select Columns.

Setelah kita menghubungkan data kita dari widget "File" ke widget "Select Columns", kita bisa mulai memilih-milih kolom mana saja yang akan kita gunakan untuk analisis. Kolom-kolom ini disebut sebagai fitur. Selain itu, kita juga perlu menentukan kolom mana yang akan kita jadikan sebagai target prediksi kita. Dalam contoh ini, kita ingin memindahkan kolom "Anggaran" dari bagian fitur menjadi bagian target.

Untuk membuat model prediksi berdasarkan data historis yang sudah kita siapkan, kita bisa menghubungkan widget "Select Columns" dengan widget "Linear Regression". Kita memilih metode Linear Regression karena data historis yang kita miliki cukup sederhana dan mengikuti pola garis lurus (linear). Namun, jika data kita ternyata tidak terlalu linear, kita bisa mencoba algoritma lain yang lebih cocok dan menghasilkan kesalahan prediksi yang lebih kecil.



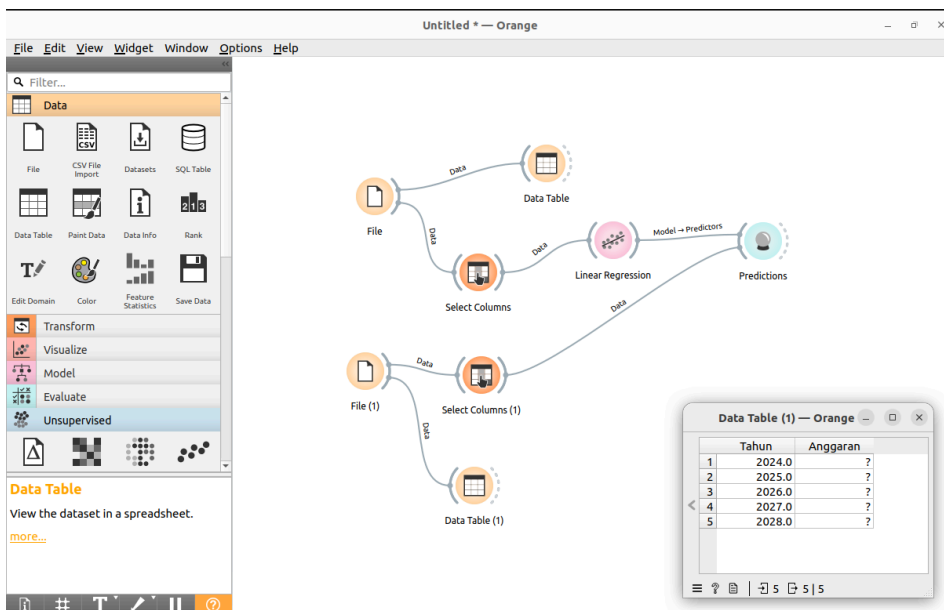
Gambar 6.6. Widget Linear Regression yang di sambung ke Widget Select Columns.

Dari keluaran Widget Linear Regression kita akan memperoleh model dari data history yang kita masukan ke widget Linear Regression. Model tersebut dapat kita gunakan untuk berbagai hal, baik itu untuk melakukan klasifikasi maupun prediksi.

Selanjutnya, setelah memiliki model dari data history dari file anggaran-2023-history.xls. Kita dapat meramalkan (memprediksi) anggaran yang diperlukan untuk tahun-tahun selanjutnya. Untuk bisa meramal, yang perlu kita tidak banyak, yaitu:

1. Sambungkan input widget predicts ke output dari widget linear regression agar model yang diperoleh masuk ke widget predicts.
2. Masukkan data dari anggaran-2023-ramal.xls ke input widget predicts setelah melalui widget select columns agar fitur dan target yang digunakan sama persis dengan data data anggaran-2023-history.xls.

Penampakan workflow yang menyatukan data dari anggaran-2023-ramal.xls ke widget predict di perlihatkan di gambar di bawah ini. Pada widget data table yang di sambungkan ke widget file yang membaca file excel ramal, jika kita klik terlihat data dengan kolom anggaran berisi tanda tanya “?” yang artinya kosong.

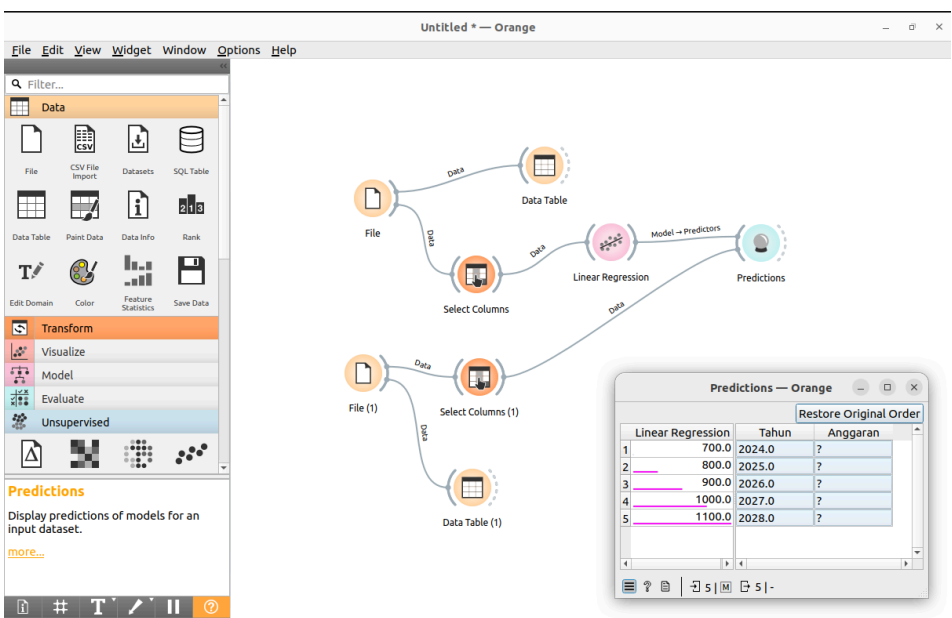


Gambar 6.7. Workflow untuk meramal anggaran.

Setelah data ramal di masukan ke widget select columns, kita dapat memilih kolom tahun sebagai fitur dan kolom anggaran sebagai target. Kita masukan keluaran dari widget select columns ke widget predictions.

Dengan masukan model dari widget linear regression dan data ramal dari widget select columns, maka widget predictions siap untuk “meramal” kebutuhan anggaran di tahun 2024 dst ..

Hasil ramalan terlihat pada gambar berikutnya, yang dapat diperoleh setelah semua input data dimasukan dengan baik, dan kita klik widget predictions maka akan di tampaknya tabel berisi beberapa kolom yang menarik.



Gambar 6.8. Penampakan Hasil Ramalan Anggaran dari Widget Predictions.

Pada gambar di atas terlihat sebuah tabel dengan tiga (3) kolom, tabel tersebut ditampilkan saat kita klik widget Predictions. Informasi dari masing-masing kolom adalah sebagai berikut,

- Kolom 1: hasil “Ramalan” dari model yang diperoleh dari history data untuk fitur yang ada di kolom 2.

- Kolom 2: kolom tahun, yang merupakan fitur untuk dilakukan ramalan.
- Kolom 3: kolom anggaran, tepatnya perkiraan anggaran untuk di bandingkan dengan hasil “ramalan” di kolom 1.

Hasil ramalan di kolom satu untuk setiap tahun yang ada di kolom tahun diperlihatkan pada tabel berikut. Dari sisi kita sebagai manusia dengan data yang bisa diprediksi dengan mudah, kita tahu bahwa prediksi anggaran tersebut berhasil dilakukan dengan baik. Tentunya hasil prediksi akan seru jika data yang diberikan tidak benar-benar linier.

Tabel 6.2. Hasil Ramalan Anggaran

Tahun	Anggaran
2024	700
2025	800
2026	900
2027	1000
2028	1100

Coba Prediksi Anggaran dengan Deviasi

Percobaan awal untuk meramal anggaran, kita menggunakan data yang sangat baik dan sangat mudah untuk di ramal. Mari kita coba ubah salah satu data history yang kita gunakan, dan kita lihat efeknya terhadap ramalan yang akan dihasilkan oleh workflow tersebut.

Perubahan yang dilakukan sangat sederhana, yaitu, anggaran pada tahun 2021 kita ubah menjadi 350 seperti pada tabel berikut,

Tabel 6.3. Contoh History Anggaran

Tahun	Anggaran
2018	100
2019	200

2020	300
2021	350
2022	500
2023	600

Dengan menggunakan workflow orange yang sama dengan sebelumnya kita akan memperoleh hasil ramalan anggaran yang berbeda lumayan signifikan.

	Linear Regression	Anggaran	Tahun
1	686.7	?	2024.0
2	785.2	?	2025.0
3	883.8	?	2026.0
4	982.4	?	2027.0
5	1081.0	?	2028.0

Gambar 6.9. Hasil Widget Prediksi.

Perkiraan kita menggunakan metode regresi linier ternyata tidak sesuai dengan data sebelumnya. Untuk memperbaiki perkiraan ini, kita bisa menambahkan informasi lain seperti tingkat bunga, inflasi, dan kondisi ekonomi negara. Dengan kata lain, kita akan membuat perkiraan yang lebih lengkap dengan mempertimbangkan banyak faktor, bukan hanya tahun saja. Nanti kita akan bahas cara memasukkan semua faktor ini ke dalam model perkiraan kita.

Coba Prediksi Anggaran dengan Multi Fitur

Selanjutnya kita akan mencoba memprediksi anggaran yang mempunyai banyak fitur yang saling tergantung satu sama lain. Dalam kesempatan ini anggaran yang akan dianalisa, mempunyai fitur, Tahun, Anggaran, Populasi Pendidikan, Jumlah Tank, Jumlah APC (Armoured Personnel Carrier), Jumlah Pesawat Tempur F35, Jumlah Seadrone, Anggaran Riset dan Development (RnD), Suku Buku, Inflasi.

Jujur data tersebut waktu itu penulis gunakan untuk memberikan demo prediksi anggaran di MABES TNI. Tidak heran kolom yang digunakan untuk prediksi adalah jumlah tank, APC, pesawat tempur, Seadrone dll. Kolom-kolom ini bisa kita ubah sesuai dengan kebutuhan kita. Tentunya ini semua bukan data yang sebenarnya, hanya data untuk kita belajar saja.

Kita pada dasarnya akan membutuhkan dua (2) file untuk bisa melakukan analisa menggunakan orange data mining, yaitu,

- File excel berisi history data atau training data dengan kolom yang kita inginkan.
- File excel berisi data untuk di ramal dengan kolom yang sama dengan kolom yang ada di training data.

Penampakan file excel berisi history data / training data yang akan dianalisis adalah sebagai berikut.

Tahun	Anggaran	Populasi	Tank	APC	F35	SEADRON	RnD	Suku Bunga	Inflasi
1990	1000	100	0	1	0	0	10	10	3
1991	2000	200	0	1	0	0	10	11	3
1992	3000	300	0	1	0	0	10	10	3
1993	4000	400	0	0	0	0	20	10	3
1994	5000	500	0	1	0	0	10	10	3
1995	6000	600	0	2	0	0	30	10	4
1996	6700	700	0	1	0	0	10	10	5
1997	8000	800	0	2	0	0	10	11	6
1998	9000	900	0	3	0	0	20	12	7
1999	10000	1000	0	4	0	1	30	13	8
2000	10500	1100	1	5	0	2	40	7	8
2001	12000	1200	2	6	0	4	50	6	9
2002	13000	1300	3	7	1	5	20	5	10
2003	14000	1400	5	5	2	6	20	4	11

Gambar 6.10. Penampakan Training Data Anggaran Multi Fitur.

Tabel data kita berisi informasi tahunan mulai dari 1990 hingga 2003, termasuk anggaran, populasi, pertoolan militer, dan indikator ekonomi. Data ini akan digunakan untuk melatih model peramalan yang akan memprediksi tren di masa depan (mulai tahun 2003). Angka-angka selain tahun dibuat khusus untuk simulasi peramalan multi-variabel.

Tahun	Anggaran	Populasi	Tank	APC	F35	SEADRON	RnD	Suku Bunga	Inflasi
2002	13000	1300	2	5	1	2	20	5	10
2003	14000	1400	2	5	1	2	20	5	10
2004	15000	1500	2	5	1	2	20	5	10
2005	16000	1600	2	5	1	2	20	5	10
2006		2000	2	5	1	2	20	5	10
2007		1800	2	5	1	2	20	5	10
2008		1900	2	5	1	2	20	5	10
2009		2000	2	5	1	2	20	5	10
2010		3000	2	5	1	2	20	5	10
2020		6000	2	5	1	2	20	5	10
2021		7000	2	5	1	2	20	5	10
2022		15000	2	5	1	2	20	5	10
2023		21000	2	5	1	2	20	5	10
2024		22000	2	5	1	2	20	5	10
2025		20000	2	5	1	2	20	5	10

Gambar 6.11. Penampakan Forecasting Data Anggaran Multi-Fitur.

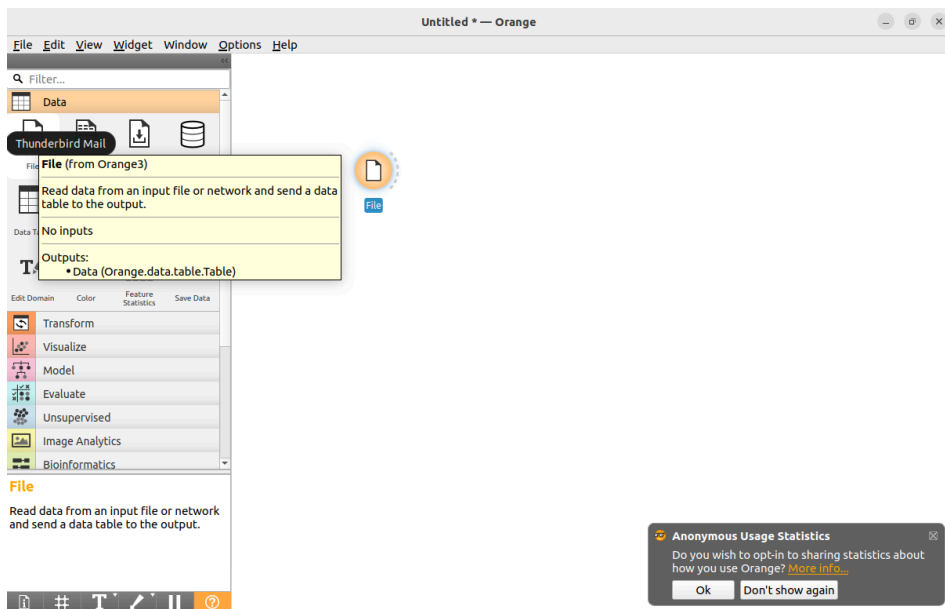
Format kolom dalam file untuk forecasting data tidak berbeda jauh dengan file sebelumnya. Beberapa hal yang berbeda adalah,

- Kolom tahun berisi angka tahun yang akan kita forecast.
- Kolom anggaran, boleh berisi data, boleh juga data hanya diisi pada sebagian tahun saja.
- Kolom lainnya di isi angka data estimasi pada tahun-tahun tersebut.

Konsekuensi dari pengisian tersebut adalah,

- Pada kolom target (disini kolom Anggaran) yang akan kita prediksi dapat saja kita kosongkan. Jika dikosongkan, maka kita akan mengandalkan Orange untuk menebak Anggaran pada tahun tersebut.
- Kolom lainnya, yang merupakan fitur tidak boleh kosong harus berisi. Jika ada yang kita tidak ketahui angkanya, maka kita harus mengestimasi angka pada kolom fitur tersebut.

Untuk meload data anggaran tersebut kita bisa mengambil widget File seperti tampak pada gambar di bawah ini.

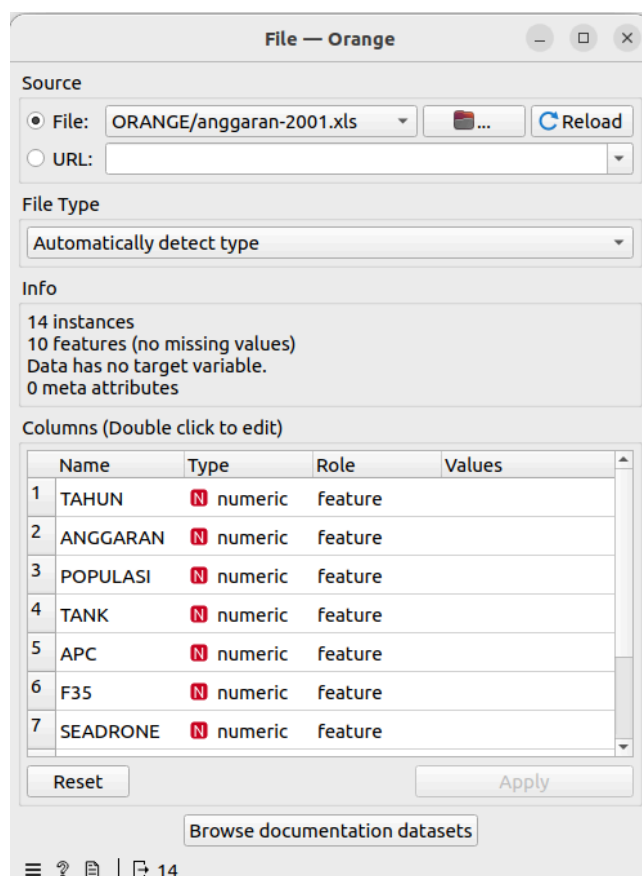


Gambar 6.12. Widget File untuk me-Load data.

Dalam contoh ini, anggaran yang kompleks dengan multi-Fitur berada pada file anggaran-2001.xls. Pada saat di load kita dapat langsung melihat rangkuman data yang kita load dalam widget File. Beberapa informasi penting akan ditampilkan adalah,

- Jumlah instance, jumlah data.
- Jumlah fitur dalam data, jika kita mengacu pada persamaan $Ax+B=Y$, maka fitur adalah X.
- Informasi tentang nilai yang kosong, jika ada nilai kosong maka kita harus mengisi nilai kosong tersebut.
- Variabel yang menjadi target, jika kita mengacu pada persamaan $Ax+B=Y$, maka target adalah Y.
- Daftar kolom, jenis data, fungsinya sebagai fitur atau target.

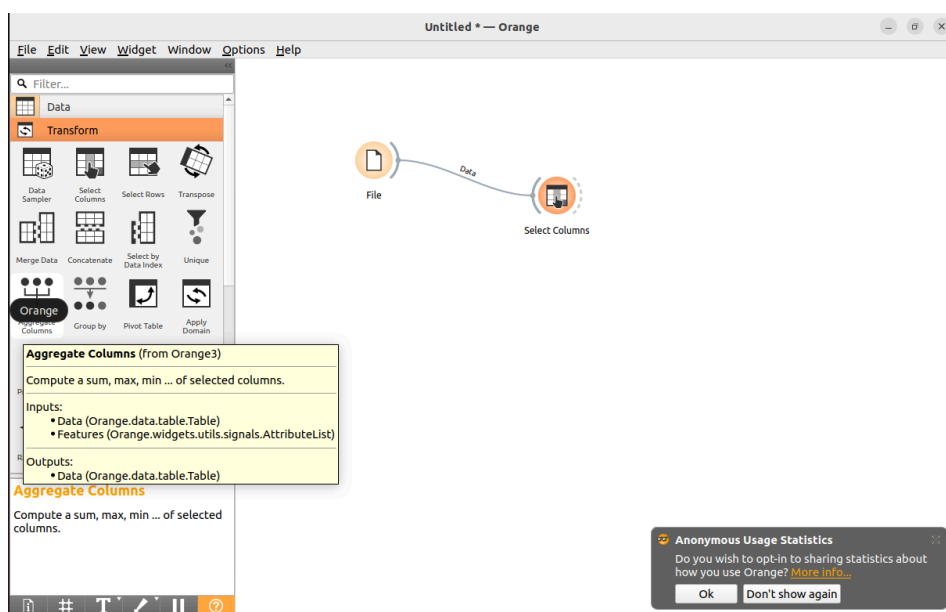
Untuk lebih jelaskan kita melihat rangkuman data yang kita lihat pada gambar berikut.



Gambar 6.13. Rangkuman informasi data

Yang menjadi masalah adalah file yang di load tidak ada kolom yang menjadi target. Artinya, semua kolom yang ada di file yang di load adalah fitur semua tidak ada kolom target.

Untuk mengubah fungsi (role) sebuah kolom dapat dilakukan menggunakan Widget Select Column. Untuk itu, keluaran widget File kita sambungkan ke Widget Select Column seperti tampak pada gambar di bawah ini.



Gambar 6.14. Keluaran Widget File di sambungkan ke Widget Select Column.

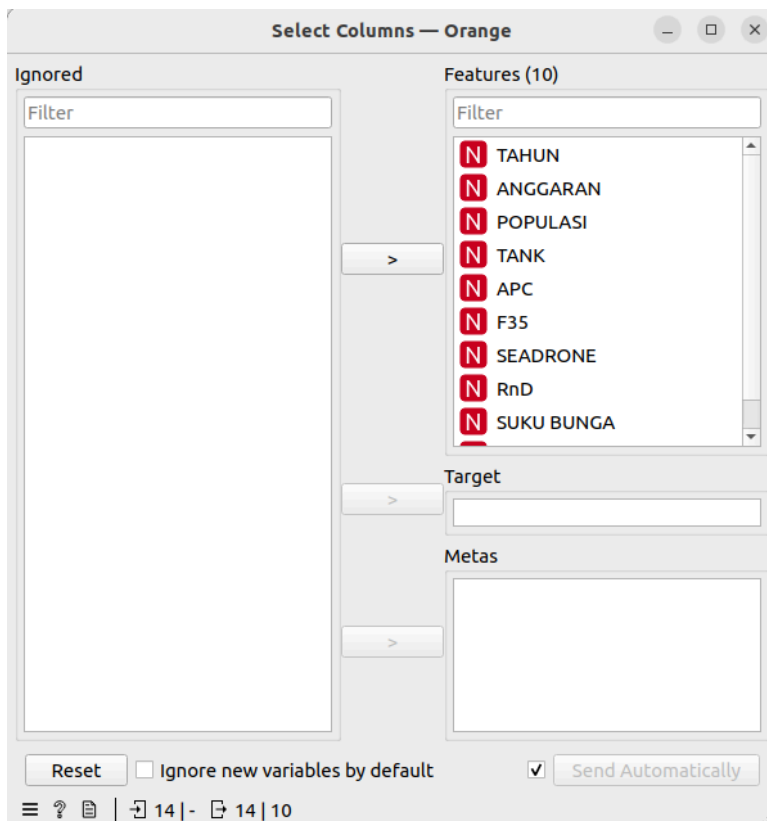
Setelah kita memilih data dari bagian "File", data tersebut secara otomatis akan ditampilkan di bagian "Select Column". Di sini kita bisa langsung melihat jenis atau peran dari setiap data (misalnya, apakah data itu berfungsi sebagai fitur atau bukan).

Dengan mengklik pada bagian "Select Column", kita bisa melihat secara detail bahwa saat ini semua data yang dipilih masuk dalam kategori "Fitur". Agar kita bisa melakukan analisis lebih lanjut, kita perlu setidaknya memiliki satu data yang masuk dalam kategori "Target". Data "Target" ini bisa diibaratkan sebagai hasil yang ingin kita prediksi (seperti nilai Y

dalam persamaan matematika), sedangkan data "Fitur" adalah faktor-faktor yang mempengaruhi hasil tersebut (seperti nilai X).

Kita bisa dengan mudah memindahkan data dari satu kategori ke kategori lainnya. Misalnya, jika kita ingin menghilangkan pengaruh suatu data dalam analisis, kita bisa memindahkannya ke kategori "Ignore". Atau, jika kita ingin menjadikan suatu data sebagai target prediksi, kita bisa memindahkannya ke kategori "Target". Selain itu, kita juga bisa mengelompokkan data ke dalam kategori "Metadata" jika diperlukan.

Saat pertama kali kita memuat data, semua data akan secara otomatis masuk ke kategori "Fitur". Artinya, kita perlu secara manual memindahkan data ke kategori yang sesuai agar analisis dapat dilakukan.

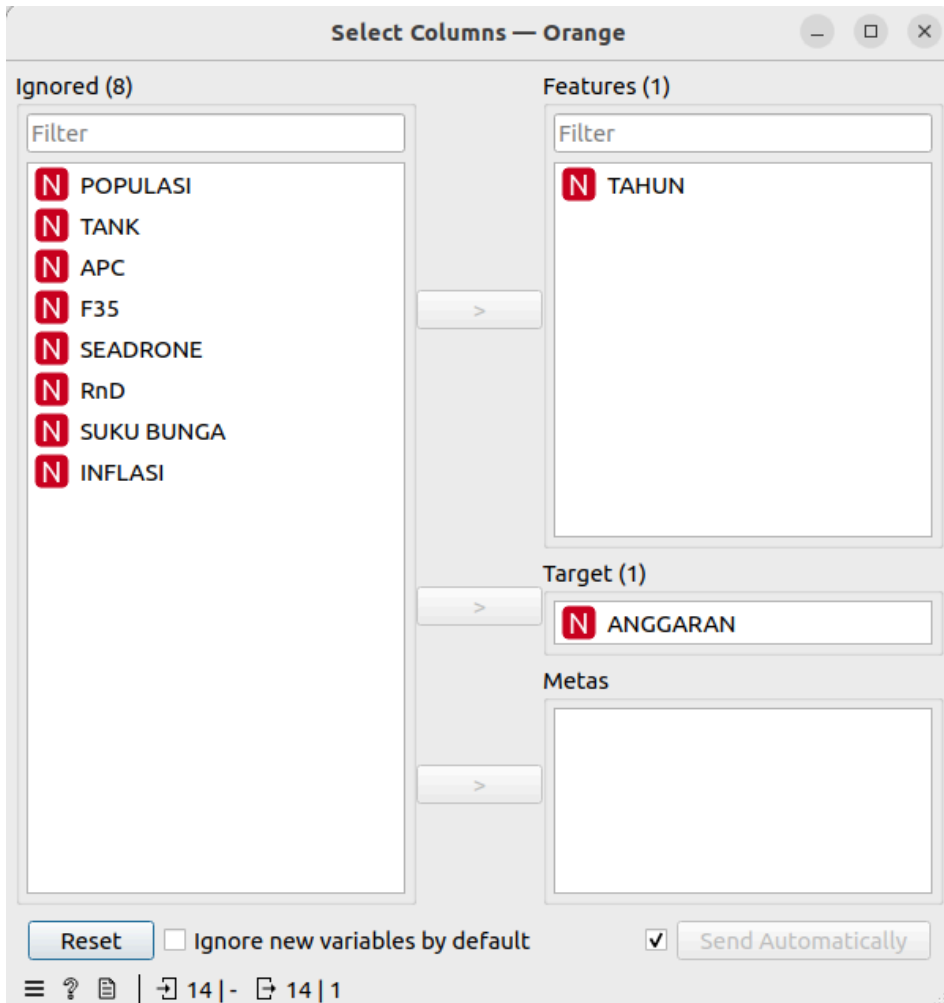


Gambar 6.15. Semua kolom menjadi fitur.

Jika kita hanya ingin menganalisa Anggaran setiap tahun, tanpa ada pengaruh dari kolom-kolom yang lain, maka kita dapat memindahkan

semua kolom ke Ignored. Sementara hanya kolom anggaran berada di Target. Kolom Tahun berada di fitur. Hasil memindahkan kolom ini terlihat pada gambar berikutnya.

Dengan memindahkan sebagian besar kolom ke Ignore, hanya menyisakan kolom Tahun dan Anggaran saja di analisa, sebetulnya kita sama saja dengan meload file data dengan dua kolom saja, yaitu, kolom Tahun dan kolom Anggaran saja.



Gambar 6.16. Memilih dua kolom untuk di analisa.

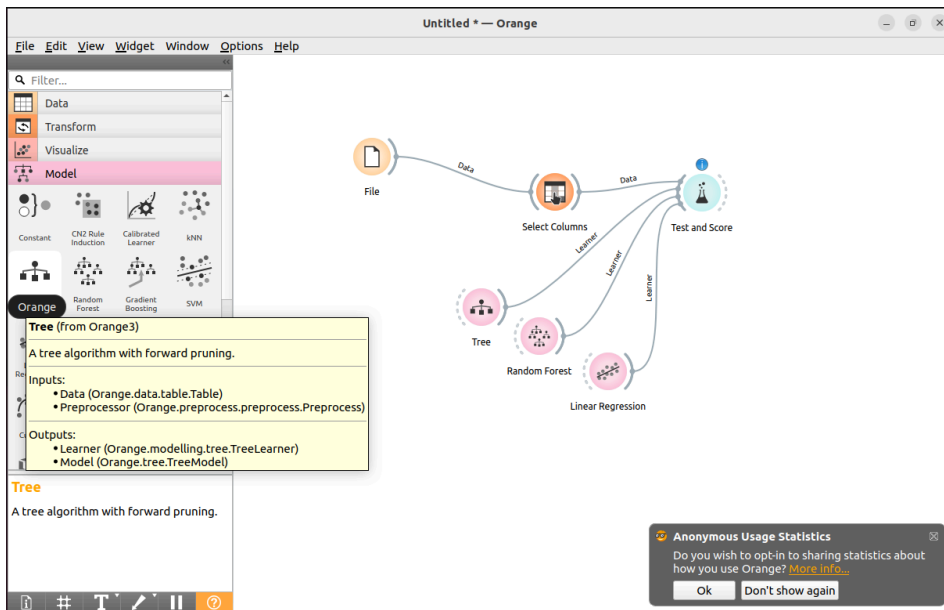
Setelah kolom yang akan digunakan dalam analisa dipilih, langkah selanjutnya adalah mengetes algoritma / model mana yang memberikan

error paling minimal dengan kondisi data yang ada. Disini sebetulnya kita perlu mengetahui jenis data yang kita analisa untuk menentukan algoritma mana yang kemungkinan akan memberikan hasil yang baik.

Contoh,

- Algoritma Linear Regression hanya baik jika target yang di analisa merupakan data linier, seperti nilai, 0, 0.1, 0.2, 0.3 dst yang banyak variasinya.
- Algoritma Logistic regression hanya baik jika target yang dianalisa hanya mempunyai beberapa nilai, seperti, baik - jelek, tua - muda dsb.
- Beberapa algoritma mempunyai kemampuan untuk menganalisa kedua jenis target data tersebut.

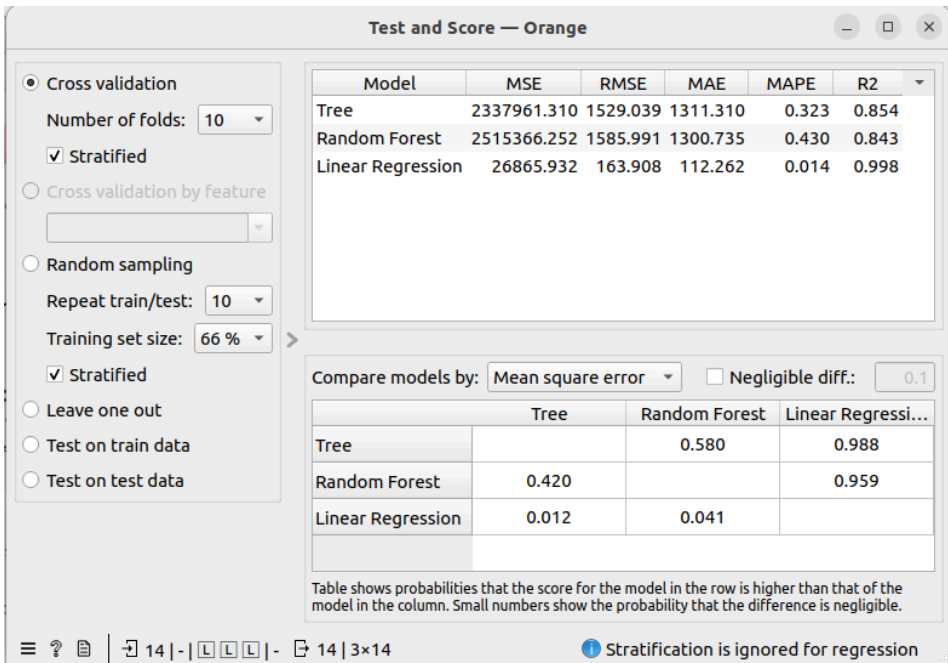
Untuk mengetes keakuratan beberapa algoritma untuk data yang kita miliki, kita dapat menggunakan widget Test & Score. Caranya data yang sudah kita Select Column dimasukan ke Widget Test & Score dan keluaran berbagai algoritma yang ingin di test di masukan ke Widget Test & Score seperti pada gambar.



Gambar 6.17. Test keakuratan Algoritma di Widget Test & Score.

Kebetulan data anggaran yang kita gunakan untuk percobaan ini merupakan data kontinu, maka algoritma yang cocok sebetulnya adalah Linear Regression. Dalam gambar di tes algoritma linear regression, tree dan random forest.

Hasil dari tes dan score di perlihatkan pada tampilan Widget Test & Score yang bisa di klik. Pada gambar berikut diperlihatkan tampilan widget Test & Score.



Gambar 6.18. Tampilan hasil Test & Score.

Pada tampilan hasil Test & Score untuk data linier biasanya dipergunakan ukuran Error, seperti, MSE, RMSE, MAR, MAPE dan R2, untuk menyatakan tingkatan kebaikan dari masing-masing algoritma, Penjelasan lebih lengkap tentang MSE, RMSE, MAR, MAPE dan R2 dijelaskan di lampiran.

Tampak pada hasil Test & Score algoritma yang akan memberikan error yang paling rendah, secara berurutan, adalah,

- Linear Regression
- Random Forest

- Tree

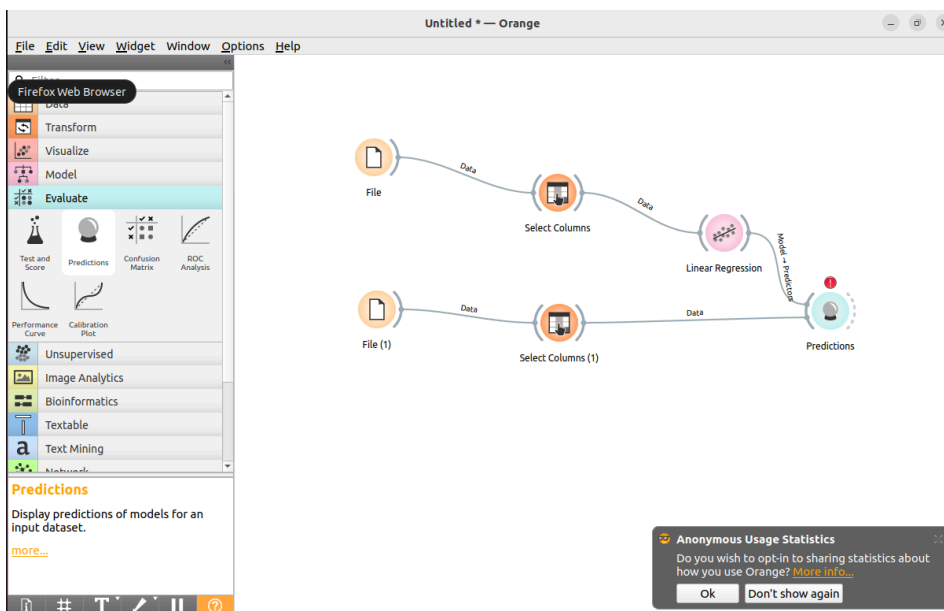
Karena algoritma linear regression memberikan error yang paling kecil, maka kita akan menggunakan widget linear regression untuk memperoleh model dari data training yang ada pada kita.

Agar lebih seru, program yang kita buat sekaligus melakukan prediksi menggunakan model yang diperoleh menggunakan widget Prediction.

Widget prediction memperoleh dua (2) masukan,

- Model dari data training yang di oleh dari keluaran widget Linear Regression.
- Data yang harus ditebak target-nya, mengambil dari widget input file yang ke dua. Dalam file ke dua ini, beberapa kolom target mempunyai nilai variabel yang kosong untuk di tebak (di prediksi).

Penampakan keseluruhan program adalah seperti gambar di bawah ini.

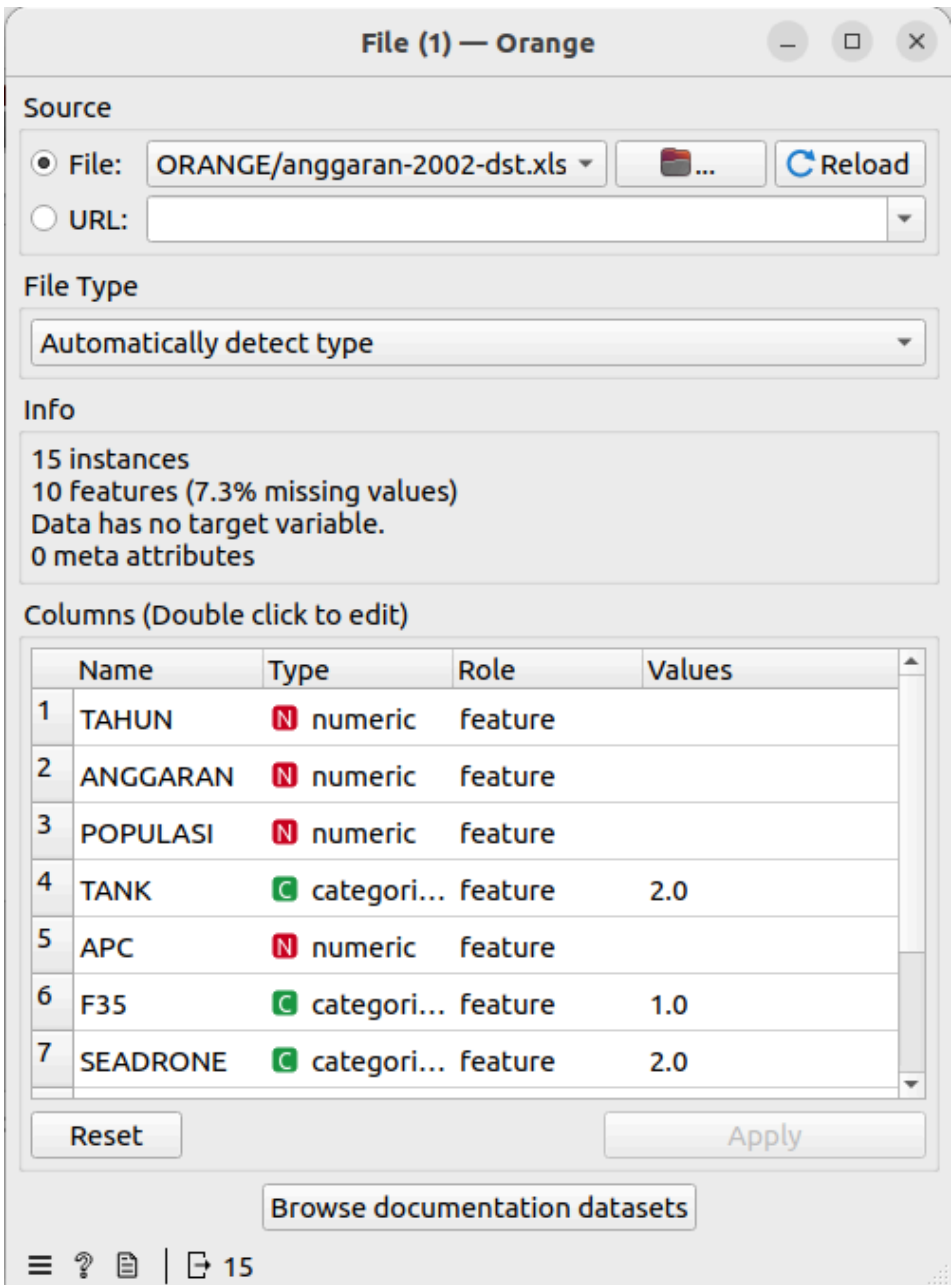


Gambar 6.19. Program untuk Prediksi

Pada dasarnya Widget Predictions akan membutuhkan dua (2) input, yaitu, model yang diperoleh dari data history atau data training; dan data yang akan akan di prediksi. Tentunya model dan data yang akan diprediksi

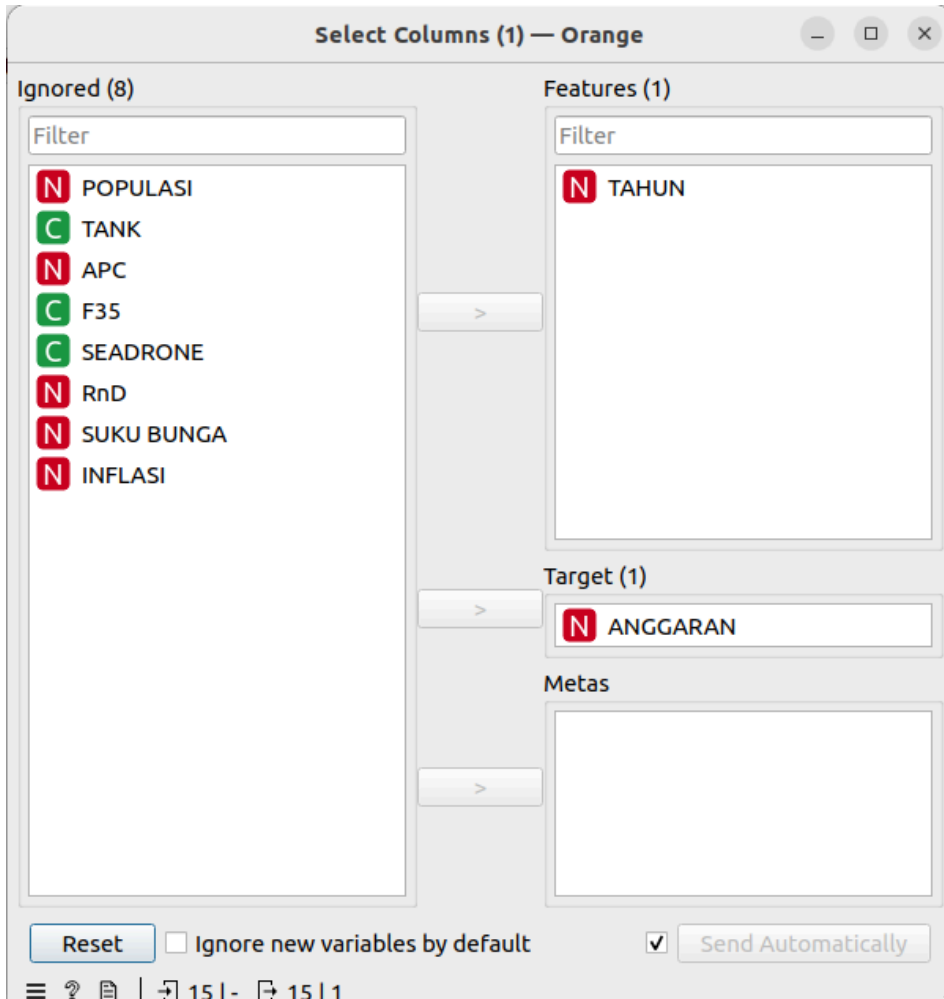
harus mempunyai kolom yang sama, dalam bahasa yang lebih keren harus mempunyai struktur data yang sama..

Contoh tampilan summary file data yang akan di prediksi diperlihatkan pada gambar selanjutnya.



Gambar 6.20. Tampilan Data yang akan diprediksi.

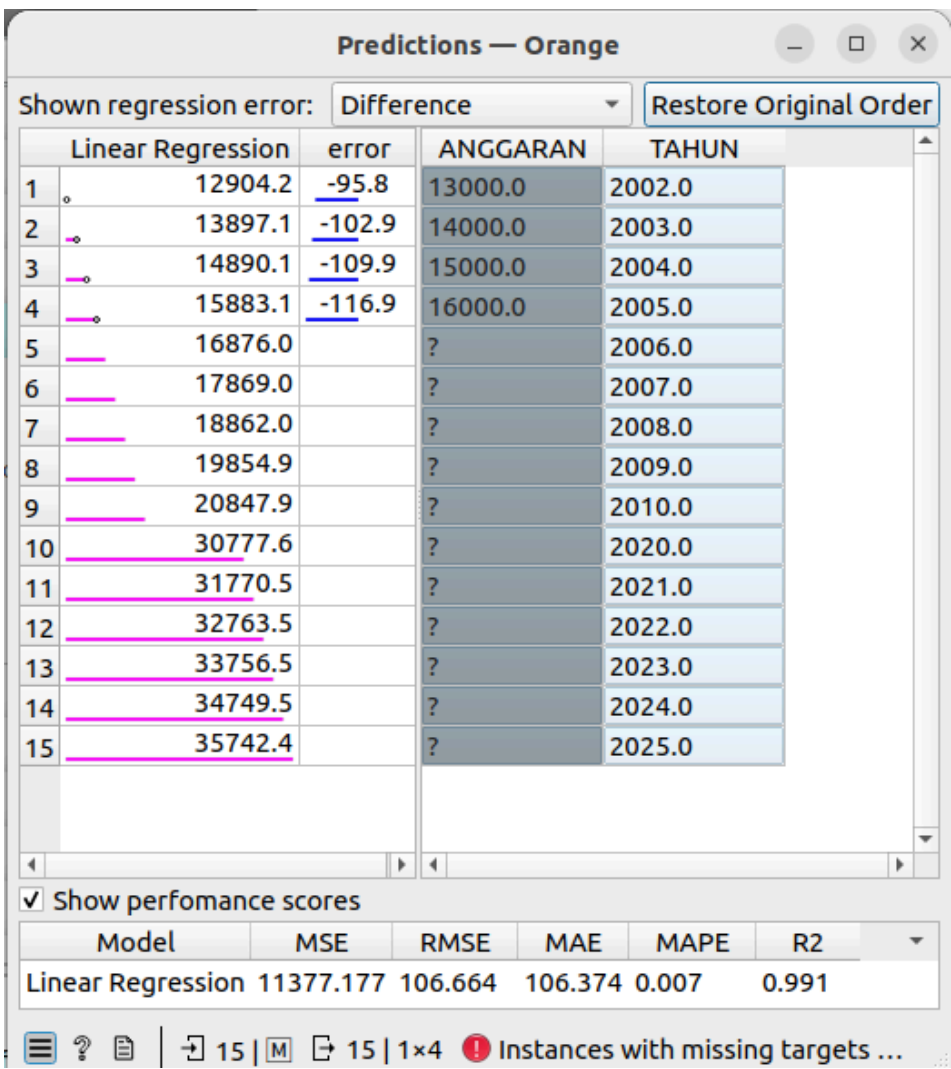
Terlihat jumlah fitur yang dimiliki sama, yaitu sepuluh (10) fitur. Jadi sama dengan data training yang digunakan. Beda utamanya adalah di missing value, di data untuk test ada 7.3% missing value.



Gambar 6.21. Widget Select Column untuk Data Test

Kita harus membuat agar data yang dimasukkan ke widget predictions sama dengan data yang digunakan untuk model. Untuk itu, kita masukan semua kolom ke bagian Ignored. Hanya data kolom Tahun dimasukkan ke

Fitur dan kolom Anggaran sebagai Target, seperti tampak pada gambar di atas.



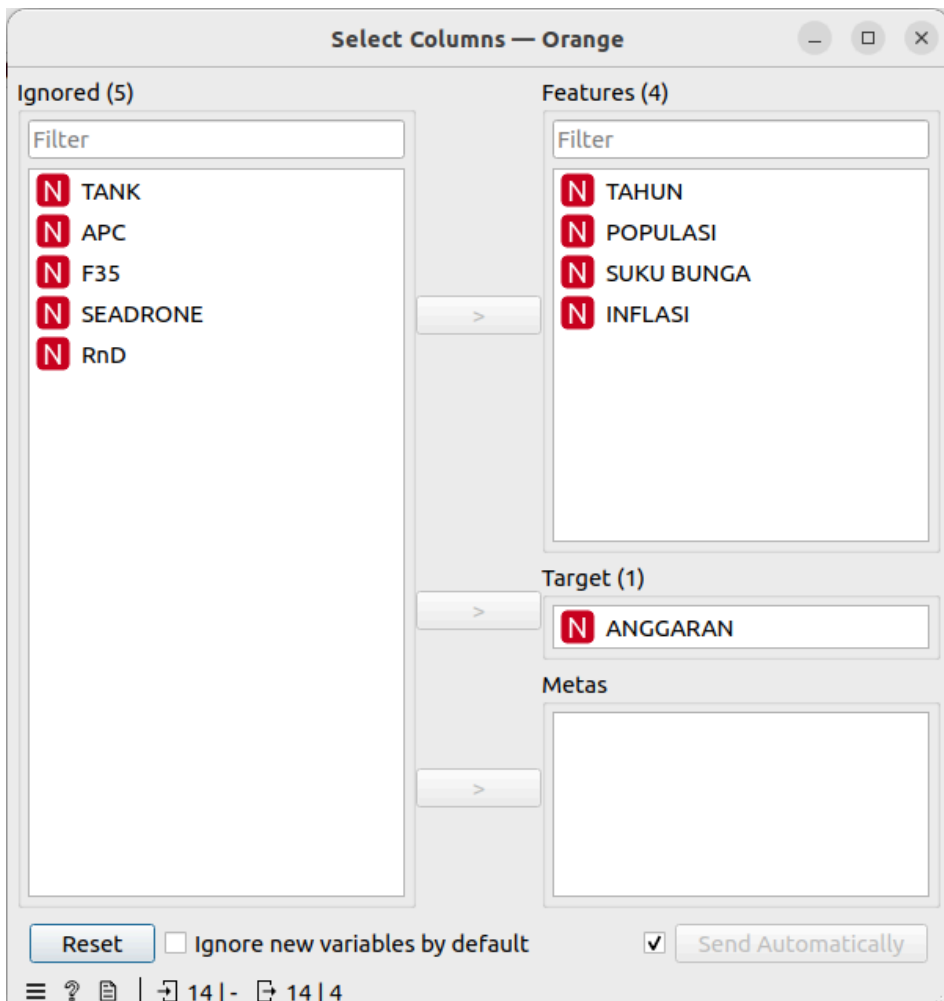
Gambar 6.22. Hasil Prediksi.

Penampakan hasil prediksi di Widget Predictions tampak pada gambar di atas. Keterangan kolom yang ditampilkan adalah sebagai berikut,

- Kolom Linear Regression - berisi hasil prediksi dari model yang diperoleh menggunakan algoritma Linear Regression.
- Kolom error - memperlihatkan error antara hasil prediksi linear regression dengan isi kolom anggaran yang kita isi manual

sebelumnya. Perhatikan bahwa error tidak ada nilainya jika di kolom anggaran kosong. Error adalah kesalahan antara hasil prediksi dengan isi anggaran yang kita isi secara manual.

- Kolom Anggaran - perkiraan anggaran yang kita masukan secara manual. Nilai Anggaran ini akan menjadi referensi untuk perhitungan kolom error.
- Kolom Tahun - merupakan salah satu fitur yang digunakan dalam perhitungan. Dalam kenyataannya, kita bisa menggunakan banyak kolom fitur.



Gambar 6.23. Pemilihan Kolom untuk Multi-Fitur.

Untuk bisa menganalisa data dengan multi-fitur maka kita perlu memasukan beberapa kolom yang akan dianalisa ke menjadi fitur. Pada gambar di atas, target anggaran adalah fungsi dari fitur tahun, populasi, suku bunga dan inflasi. Setelah kedua widget select column disesuaikan untuk multi-fitur, maka akan dihasilkan prediksi seperti pada gambar di berikut.

Predictions — Orange

Shown regression error: Difference Restore Original Order

	Linear Regression	error	ANGGARAN	TAHUN	POPULA
1	12908.3	-91.7	13000.0	2002.0	1300.0
2	13869.8	-130.2	14000.0	2003.0	1400.0
3	14831.3	-168.7	15000.0	2004.0	1500.0
4	15792.9	-207.1	16000.0	2005.0	1600.0
5	19638.7		?	2006.0	2000.0
6	17716.0		?	2007.0	1800.0
7	18677.5		?	2008.0	1900.0
8	19639.0		?	2009.0	2000.0
9	29253.5		?	2010.0	3000.0
10	58097.7		?	2020.0	6000.0
11	67712.3		?	2021.0	7000.0
12	144627.7		?	2022.0	15000.0
13	202314.3		?	2023.0	21000.0
14	211928.8		?	2024.0	22000.0
15	192700.1		?	2025.0	20000.0

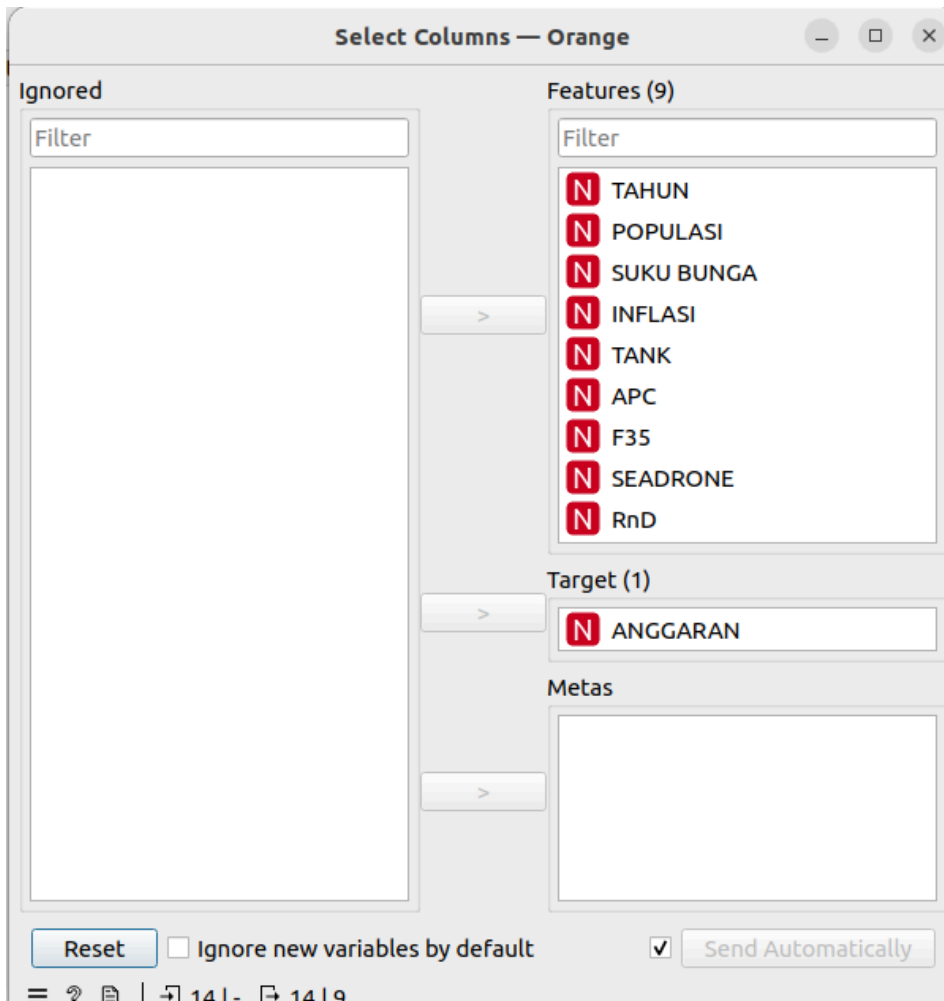
☒ Show performance scores

Model	MSE	RMSE	MAE	MAPE	R2
Linear Regression	24178.585	155.495	149.430	0.010	0.981

Instances with missing targets ...

Gambar 6.24. Hasil Prediksi dengan Data Mukti-Fitur.

Setelah ditambahkan tiga (3) kolom fitur tambahan, yaitu, populasi, suku bunga dan inflasi, maka terlihat jelas bahwa prediksi anggaran hasil dari linear regression berbeda dari anggaran hanya satu fitur saja. Perubahan ini juga akan terjadi jika semakin banyak fitur yang mempengaruhi anggaran. Pada gambar di bawah, di perlihatkan semua kolom yang ada di file kita masukan sebagai fitur.

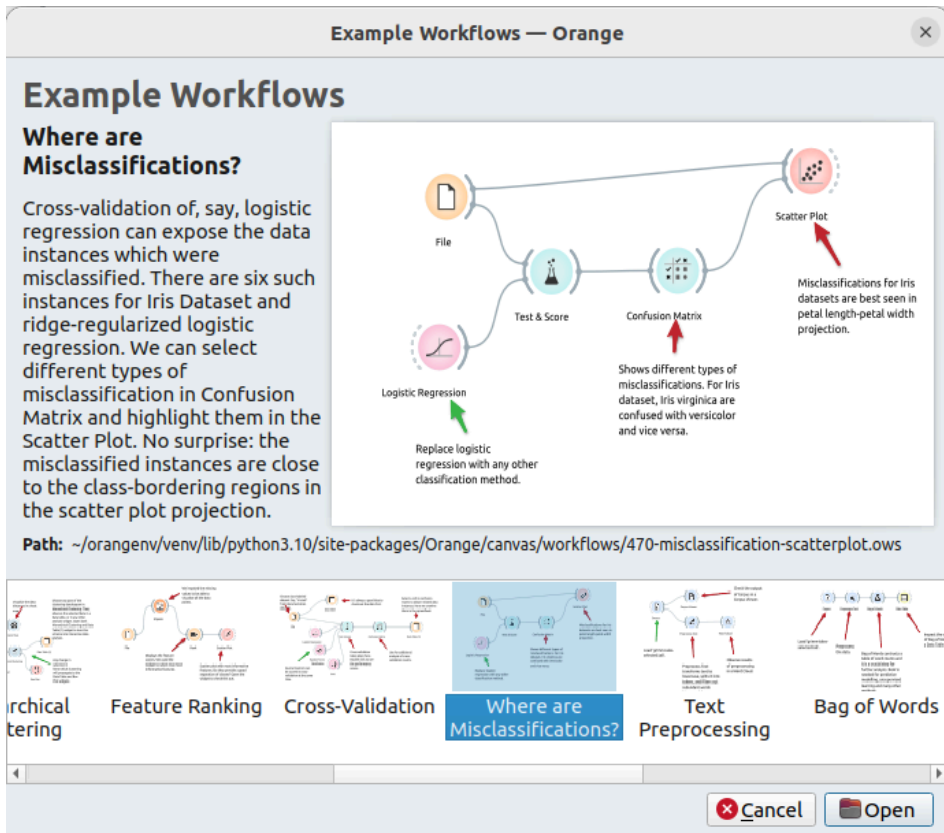


Gambar 6.25. Tampilan Widget Select Column dengan semua kolom menjadi fitur.

BAB 7: Explorasi Contoh Bawaan Orange

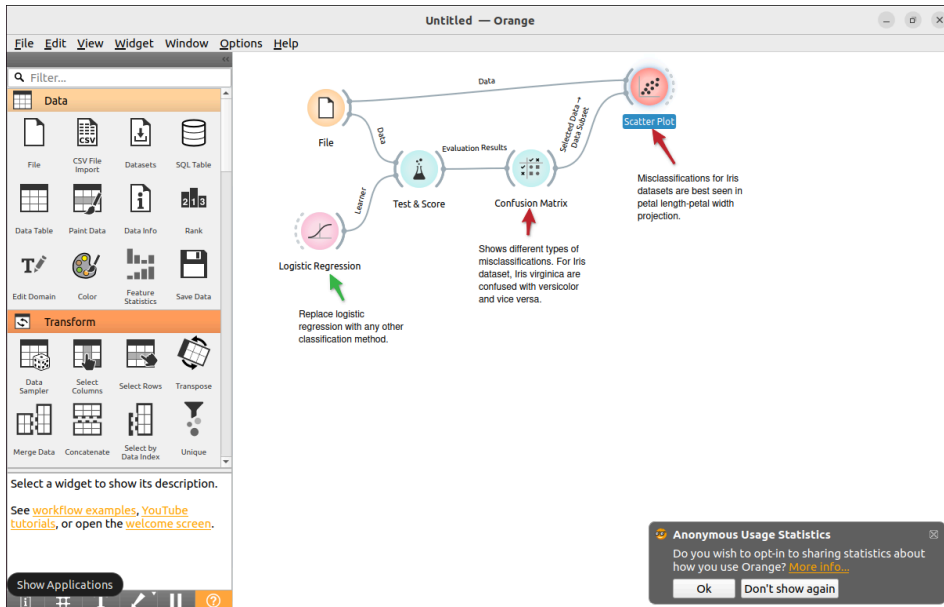
Menganalisa Model dari Data

Disini kita akan mencoba membahas salah satu contoh workflow yang juga tersedia di orange data mining yaitu “Where are Misclassifications?”. Inti dari workflow tersebut adalah untuk melihat kesalahan klasifikasi atau sering dikenal misclassification yang terjadi dalam artificial intelligence atau machine learning. Memvisualisasi misclassification tersebut. Dan yang juga tidak kalah menarik adalah memvisualisasi data yang bermasalah di Scatter Plot.



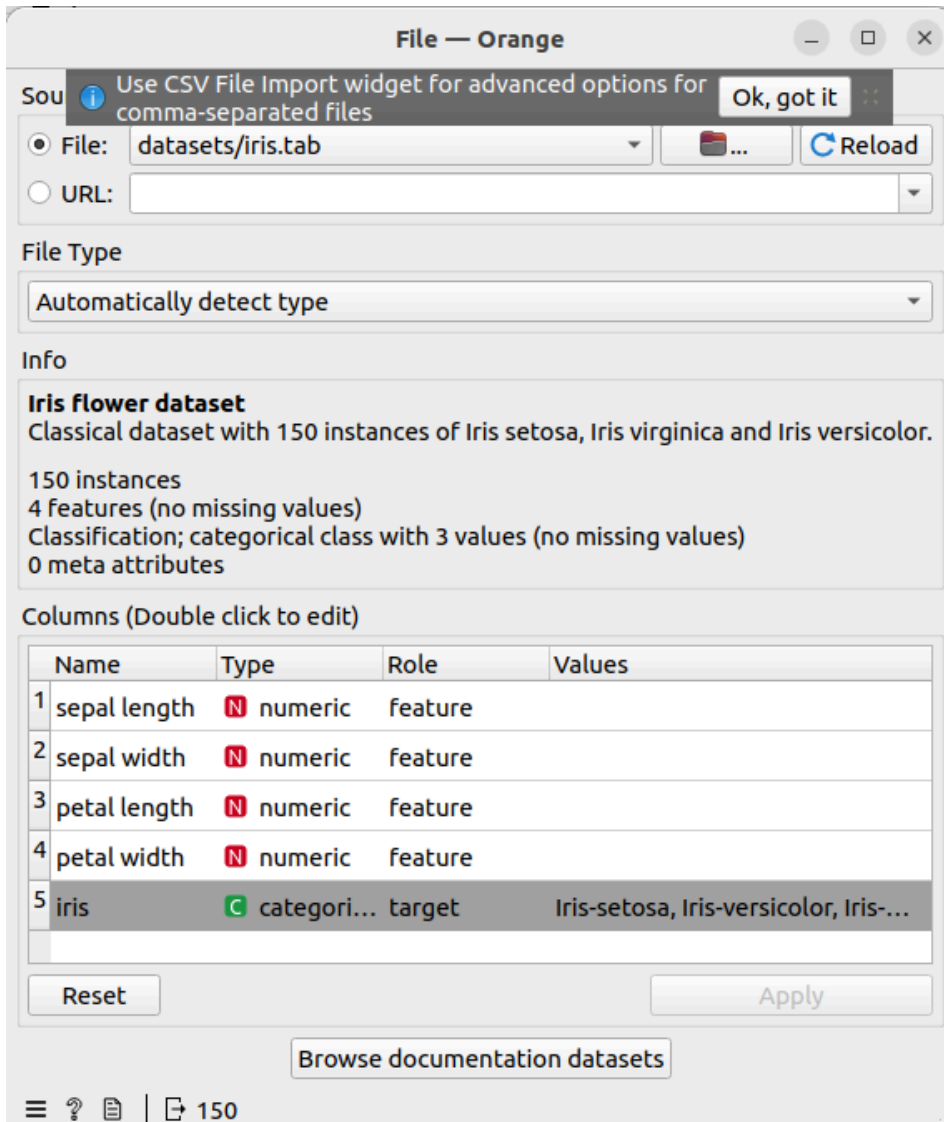
Gambar 7.1. Tampilan Kumpulan Contoh Workflow khususnya Where are Misclassifications?

Gambar 7.1 memperlihatkan kumpulan contoh workflow yang ada di orange data mining, termasuk workflow where are misclassifications?.



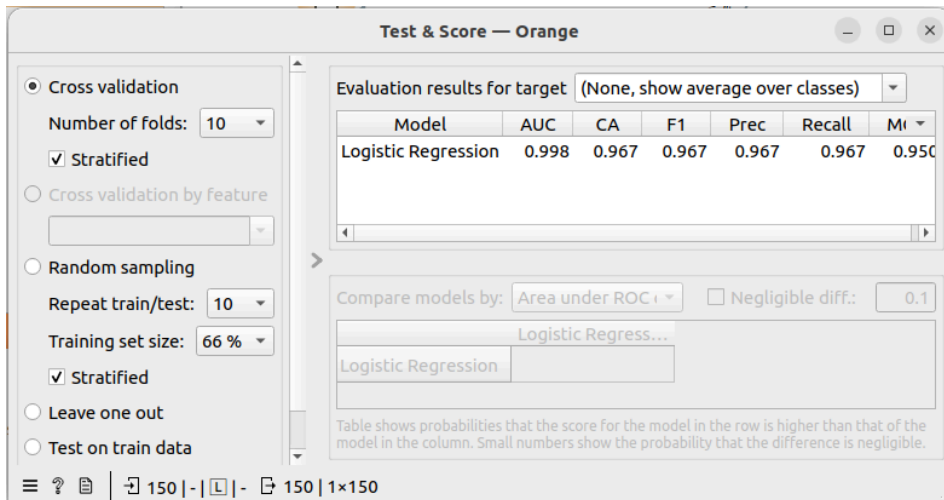
Gambar 7.2. Tampilan Workflow Where are Misclassifications di orange data mining.

Pada gambar di atas memperlihatkan saat workflow Where are Misclassifications? di load ke orange data mining. Disini ada tiga (3) widget yang belum digunakan sebelumnya, yaitu, widget Test & Score, widget Confusion Matrix, dan Widget Logistic Regression. Pada dasarnya semua Widget akan langsung beroperasi saat workflow di load. Data akan dikirim ke dua Widget, yaitu Widget Test & Score dan Widget Scatter Plot. Di Widget Test & Score, kita dapat melakukan test akan algoritma untuk pemodelan yang ingin kita gunakan terhadap input data dalam hal ini algoritma yang dites adalah Logistic Regression. Keluaran Widget Test & Score kita masukan ke Widget Confusion Matrix yang akan memperlihatkan untuk algoritma yang sedang kita test, berapa data yang secara benar di tebak, berapa yang salah tebak misclassification. Untuk analisa lebih lanjut, kita dapat memilih data di Confusion Matrix untuk ditampilkan di Scatter Plot untuk bisa memperoleh gambaran mengapa algoritma sampai melakukan misclassification.



Gambar 7.3. Tampilan Widget File.

Tampilan yang familiar bukan? Ya, betul, file yang di load sama persis dengan contoh sebelumnya yaitu file iris.tab. Data ini yang selanjutnya dikirim ke dua Widget yang berbeda, yaitu, Widget Text & Score dan Widget Scatter Plot.

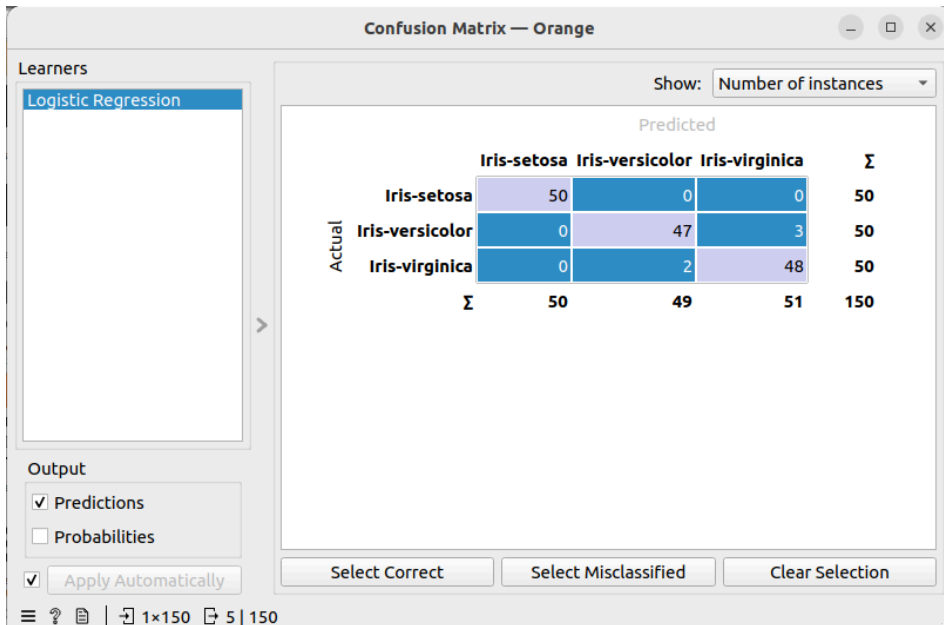


Gambar 7.4. Tampilan Widget Test & Score

Pada Gambar di atas, kita dapat melihat tampilan Widget Test & Score. Pada bagian evaluation result kita dapat melihat hasil test dan score evaluasi berbagai algoritma untuk modelling untuk memperoleh model dari data training yang dimasukan. Dalam workflow ini, algoritma modelling yang dimasukan hanya Logistic Regression saja. Jika kita masukan memasukan lebih dari satu Algoritma Modelling maka hasil yang akan diperoleh di hasil evaluasi test & score untuk beberapa model.

Yang penting kita perhatikan dari hasil test & score biasanya tingkat error atau tingkat presisi (Prec) dari model yang diperoleh. Dalam gambar di perlihatkan Model Logistic Regression yang diperoleh mampu untuk mencapai presisi (Prec) 0.967 atau 96.7% presisi. Ini merupakan angka yang sangat baik sekali, karena tingkat kesalahan yang akan diperoleh sangat kecil sekali hanya 3.3% saja.

Di sebelah kiri dari Widget Test & Score ada beberapa parameter yang dapat kita ubah2 jika kita inginkan. Walaupun perubahan parameter ini biasanya tidak diperlukan, jika kita ingin melakukan beberapa yang mungkin menarik adalah gunakan Random Sampling, dan mengubah Training set size, bisa dinaikkan dari 66%.



Gambar 7.5. Confusion Matrix.

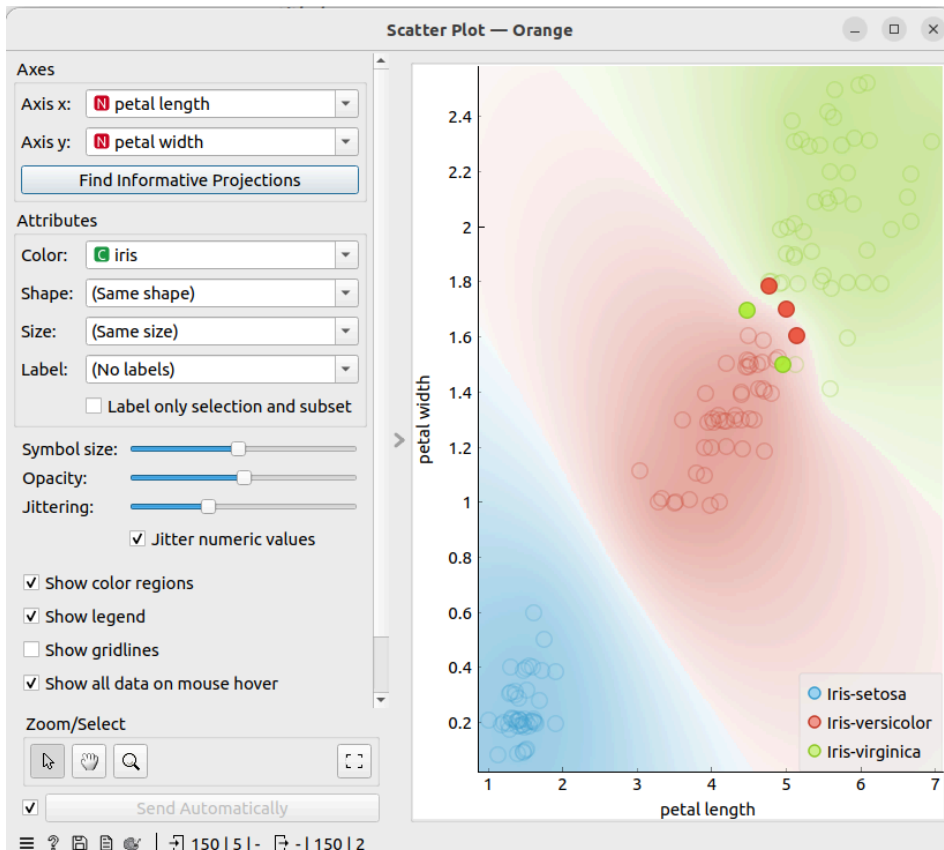
Pada Gambar di atas, diperlihatkan tampilan Widget Confusion Matrix yang menerima input langsung dari Widget Test & Score. Dalam bahasa sederhana confusion matrix adalah sebuah tabel atau matrix yang memperlihatkan kebingungan (*confusion*) dari algoritma dalam melakukan klasifikasi. Sumbu Y dari matrix adalah nilai yang sebenarnya. Sementara, sumbu X dari matrix adalah nilai yang diprediksi oleh model.

Idealnya jika model yang digunakan berhasil mendeteksi 100% tanpa kesalahan. Artinya dalam confusion matrix, semua angka hasil prediksi sama persis dengan nilai sebenarnya (actual). Artinya, semua angka berada di diagonal dari confusion matrix, diluar diagonal matrix semua angka adalah NOL.

Kenyataannya tidak semua angka diluar diagonal adalah NOL. Pada gambar di atas terlihat bahwa,

- Ada 2 Iris Virginica yang terprediksi sebagai Iris Versicolor.
- Ada 3 Iris Versicolor yang terprediksi sebagai Iris Virginica.
- Khusus untuk Iris Setosa semua berhasil terprediksi dengan baik.

Dalam bahasa yang sederhana, model tidak berhasil melakukan klasifikasi 100% untuk Iris Virginica dan Iris Versicolor. Sedang untuk Iris Setosa, model berhasil 100% melakukan klasifikasi.



Gambar 7.6. Scatter Plot

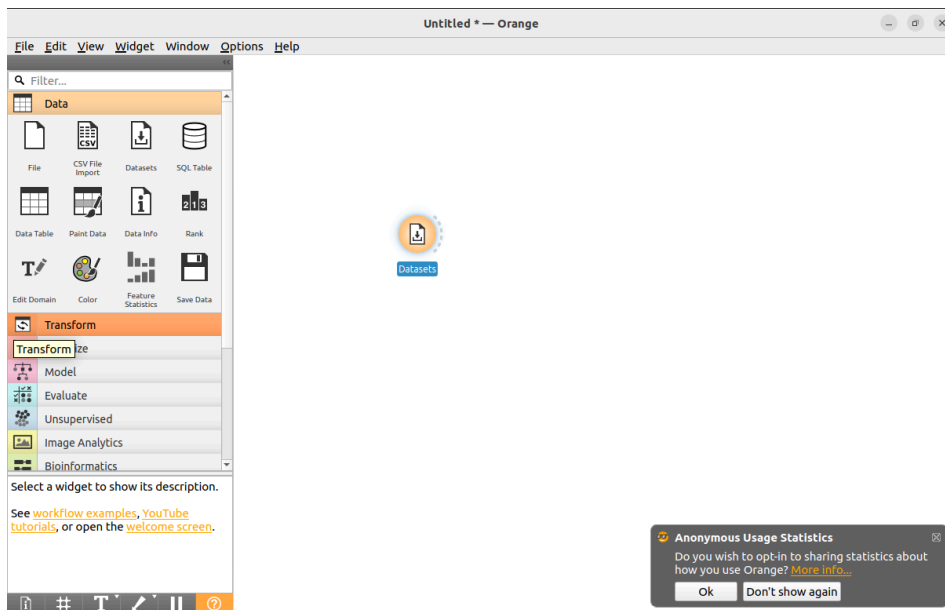
Gambar di atas, memperlihatkan hasil Scatter Plot jika data misclassifications di pilih dan di gabungan dengan seluruh data. Terlihat ada dua (2) Iris Virginica berwarna hijau berada di wilayah merah. Dan ada tiga (3) Iris Versicolor berwarna merah berada di wilayah hijau. Sementara tidak ada biru yang salah klasifikasi.

Contoh Pengambilan Keputusan Pembelian Mobil

Pada contoh berikut kita akan coba menganalisa proses pengambilan keputusan sekelompok customer yang dataset-nya telah dikumpulkan

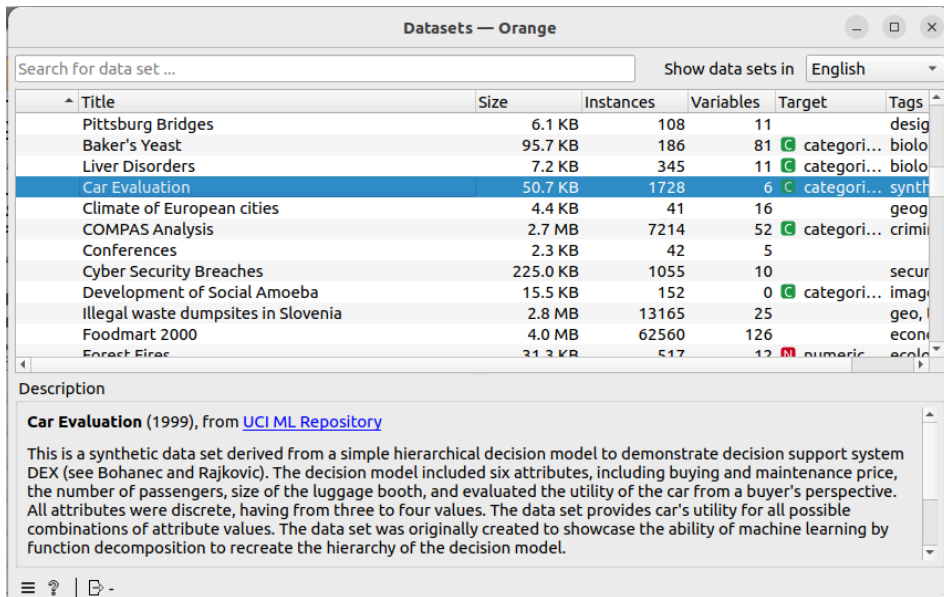
dalam orange data mining. Dataset tersebut adalah arsip transaksi keputusan customer dalam pembelian mobil. Untuk lebih jelasnya ada baiknya kita load data pembelian mobil dari kumpulan dataset yang ada di orange data mining.

Langkah pertama yang perlu dilakukan adalah mengambil widget Datasets seperti tampak pada gambar.



Gambar 7.7. Menyiapkan Widget Dataset.

Jika kita klik Widget dataset akan tampak bahwa cukup banyak dataset untuk latihan yang sudah disiapkan di orange data mining. Kita hanya perlu memilih salah satu dari dataset yang ada untuk mulai bermain-main.



Gambar 7.8. Kumpulan Dataset dalam Widget Dataset.

Setelah kumpulan dataset terpampang dalam widget dataset, langkah yang perlu kita lakukan adalah memilih dataset mana yang akan kita gunakan untuk melakukan latihan.

Kita dapat scroll ke bawah, dengan mengklik dataset yang diinginkan kita dapat membaca keterangan lebih lengkap tentang dataset tersebut. Tentunya semua tulisan dalam bahasa Inggris.

Alternatif lain adalah dengan mencari nama atau keyword dari dataset tersebut pada input text Search.

Dalam contoh gambar di atas, kita mencari dataset dengan judul

Car Evaluation

Seperti dijelaskan sebelumnya, ini berisi catatan customer saat mengevaluasi pembelian mobil.

	car	buying	maint	doors	persons	lug_boot	safety
1	unacc	vhigh	vhigh	2	2	small	low
2	unacc	vhigh	vhigh	2	2	small	med
3	unacc	vhigh	vhigh	2	2	small	high
4	unacc	vhigh	vhigh	2	2	med	low
5	unacc	vhigh	vhigh	2	2	med	med
6	unacc	vhigh	vhigh	2	2	med	high
7	unacc	vhigh	vhigh	2	2	big	low
8	unacc	vhigh	vhigh	2	2	big	med
9	unacc	vhigh	vhigh	2	2	big	high
10	unacc	vhigh	vhigh	2	4	small	low
11	unacc	vhigh	vhigh	2	4	small	med
12	unacc	vhigh	vhigh	2	4	small	high
13	unacc	vhigh	vhigh	2	4	med	low
14	unacc	vhigh	vhigh	2	4	med	med
15	unacc	vhigh	vhigh	2	4	med	high
16	unacc	vhigh	vhigh	2	4	big	low
17	unacc	vhigh	vhigh	2	4	big	med
18	unacc	vhigh	vhigh	2	4	big	high
19	unacc	vhigh	vhigh	2	more	small	low
20	unacc	vhigh	vhigh	2	more	small	med

Gambar 7.9. Data Table dari dataset Car Evaluation.

Setelah widget datasets di sambungkan dengan widget data table, maka kita dapat mengevaluasi atau melihat data yang ada dalam dataset yang kita pilih.

Dalam widget data tabel kita dapat melihat summary / ringkasan tentang dataset yang sedang kita amati, antara lain adalah,

- Ada 1728 data atau instance yang ada dalam dataset.
- Ada 6 fitur, yaitu, buying, maintenance, doors, persons, lug_boot, safety
- Ada target (car) dengan 4 nilai, yaitu, unacc, acc, vgood, good

Khusus, untuk dataset Car Evaluation, terlihat dengan jelas kolom yang tersedia dengan kemungkinan nilainya adalah sebagai berikut,

Tabel 7.1 Kolom dan Nilai Dataset Car Evaluation

Kolom	Nilai			
car	unacc	acc	vgood	good
buying	vhigh	high	med	low
maint	vhigh	high	med	low

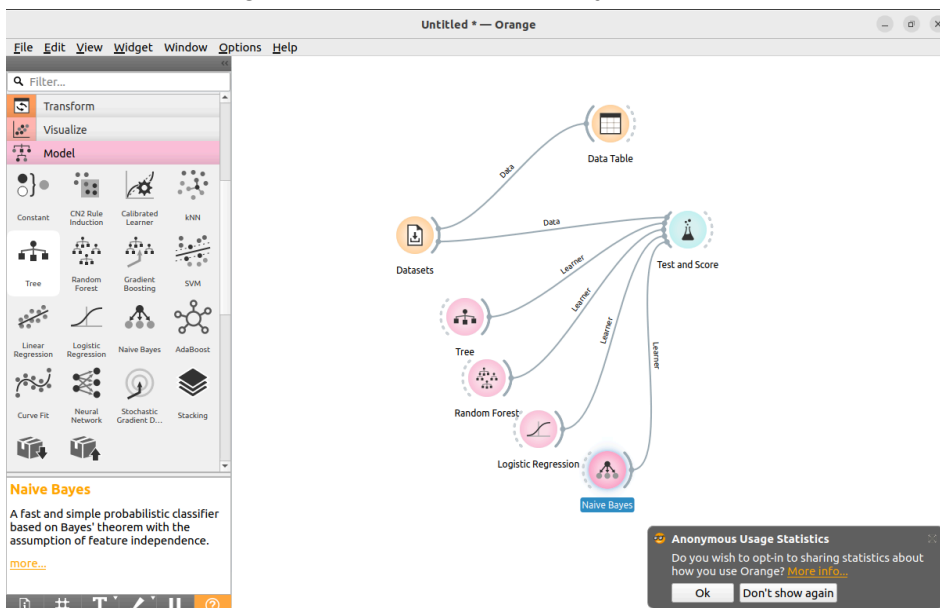
doors	2	3	4	5more
person	2	4	more	
lug	big	med	small	
safety	high	med	low	

Yang sangat menarik dari dataset yang akan kita gunakan adalah,

- Kita tidak mengolah angka
- Kita mengolah kata-kata sebagai nilai dari parameter

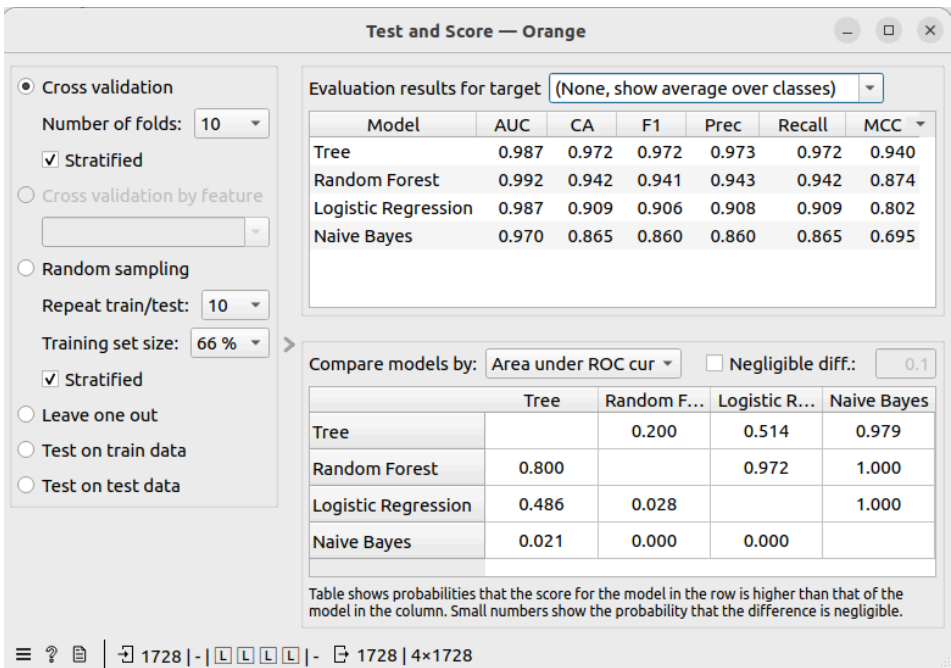
Kemampuan untuk menganalisa data yang tidak berbentuk angka menjadikan teknik-teknik machine learning dan deep learning ini menjadi sangat menarik.

Parameter yang menjadi target adalah car, dimana pembeli akan memutuskan apakah jadi atau tidak untuk membeli mobil, apakah kesan mobil tersebut sangat baik atau cukup baik saja.



Gambar 7.10 Test semua Algoritma untuk Modeling.

Sebelum proses analisa dilakukan sebaiknya kita melakukan analisa akan algoritma untuk modeling yang akan menghasilkan hasil yang paling presisi. Proses test dilakukan dengan menyambungkan widget algoritma dan data ke widget test and score.

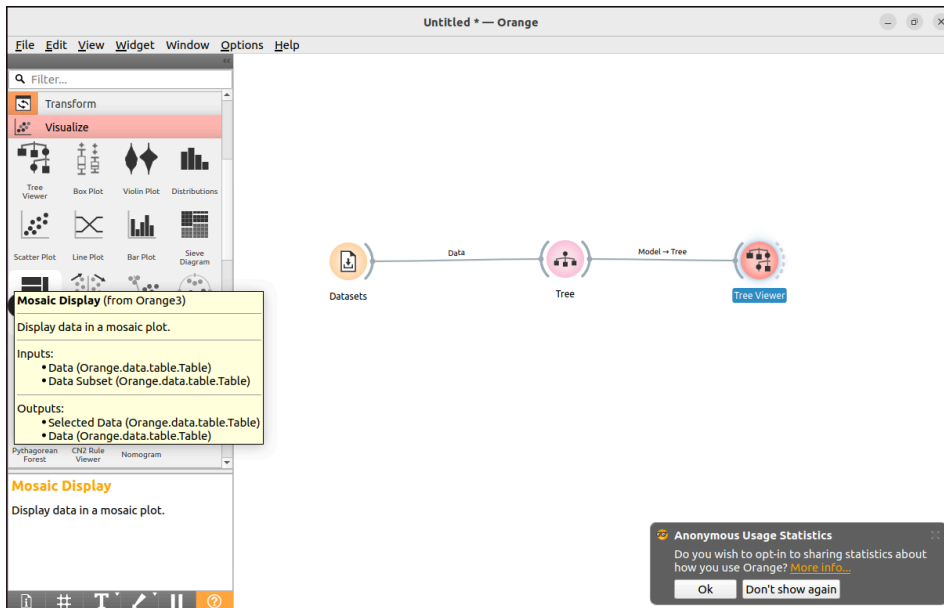


Gambar 7.11. Hasil Test and Score

Secara umum dari hasil Test and Score, kita perlu memperhatikan kolom Prec (atau Precision) atau kolom Error. Kolom Precision biasanya diperlihatkan jika kita melakukan analisa klasifikasi yang sifatnya diskrit. Sementara kolom Error biasanya diperlihatkan jika kita melakukan analisa target yang bersifat kontinyu, bukan diskrit.

Jika hasil kolom precision mendekati nilai 1 atau 100% maka algoritma yang digunakan sangat baik (sangat presisi) dalam menebak. Jika hasil kolom Error mendekati nol maka hasil analisa target sangat baik.

Dalam kasus Car Evaluation, algoritma yang menghasilkan presisi yang paling baik dalam hal ini adalah algoritma Tree. Oleh karena itu kita akan melanjutkan analisa menggunakan Tree.

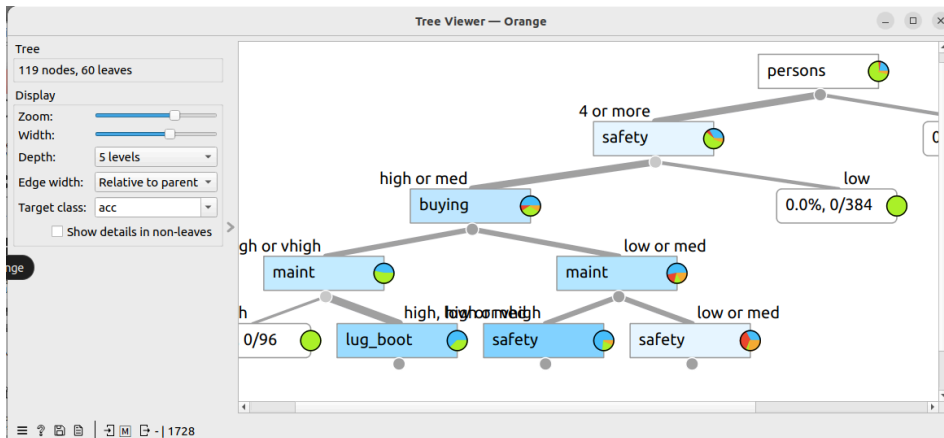


Gambar 7.12. Workflow dengan Tree dan Tree Viewer

Gambar di atas menunjukkan bagaimana kita bisa menghubungkan kumpulan data evaluasi mobil dengan algoritma pohon (Tree) dalam sebuah alur kerja (workflow). Hasil dari pemodelan menggunakan algoritma pohon ini kemudian kita visualisasikan menggunakan tool yang disebut Tree Viewer.

Cara menghubungkan ketiga komponen ini (dataset, algoritma Tree, dan Tree Viewer) adalah pendekatan yang umum digunakan dalam pembelajaran mesin. Dengan hanya tiga komponen sederhana ini, kita sudah bisa mendapatkan pemahaman awal tentang data yang kita analisis. Proses ini sangat mudah dilakukan dengan menggunakan perangkat lunak Orange. Namun, prinsip yang sama juga bisa diterapkan pada bahasa pemrograman lain seperti R, Python, TensorFlow, atau Keras.

Gambar selanjutnya menunjukkan hasil visualisasi dari algoritma pohon kita. Untuk bisa memahami dengan benar apa yang ditampilkan oleh visualisasi ini, kita perlu memahami masalah yang sedang kita analisis. Pohon yang kita lihat adalah pohon keputusan, yang merupakan model untuk menggambarkan bagaimana kita membuat keputusan.



Gambar 7.13. Tree dari Car Evaluation.

Gambar di atas memperlihatkan hasil modeling Tree dari data Car Evaluation. Inti-nya disini kita dapat melihat bagaimana orang barat mengambil keputusan dalam menentukan pembelian mobil.

Untuk rakyat +62 Indonesia kemungkinan besar saat membeli mobil yang akan diperhatikan pertama kali adalah harga (atau buying) jika murah (low) kemungkinan besar akan dibeli.

Jika kita perhatikan baik-baik, maka dari dataset Car Evaluation yang menggunakan data dari barat, maka ternyata yang urutan fitur yang akan dilihat dalam prose pengambilan keputusan adalah sebagai berikut,

- Pertama kali yang dilihat adalah person atau jumlah penumpang.
- Kedua yang akan dilihat adalah safety
- Baru harga mobil
- di lanjut dengan biaya maintenance
- terakhir baru akan dilihat besarnya bagasi

Hasil penelitian ini hanya berlaku untuk negara-negara di Eropa, karena data yang digunakan berasal dari sana. Ini penting untuk diingat, karena data yang kita gunakan sangat mempengaruhi hasil akhir dari sebuah analisis. Kita tidak bisa begitu saja mengambil kesimpulan tentang kondisi di Indonesia dengan menggunakan data dari negara lain.

Saat kita membeli perangkat lunak atau aplikasi kecerdasan buatan, data yang digunakan untuk melatih aplikasi tersebut sangat penting. Data inilah yang menentukan seberapa akurat hasil yang diberikan oleh aplikasi. Jangan langsung percaya dengan klaim vendor bahwa hasil analisis mereka pasti akurat. Jika vendor tersebut tidak memiliki data yang spesifik untuk Indonesia, maka hasil analisisnya kemungkinan besar tidak akan sesuai dengan kondisi yang sebenarnya di negara kita.

Contoh Klasifikasi Binatang

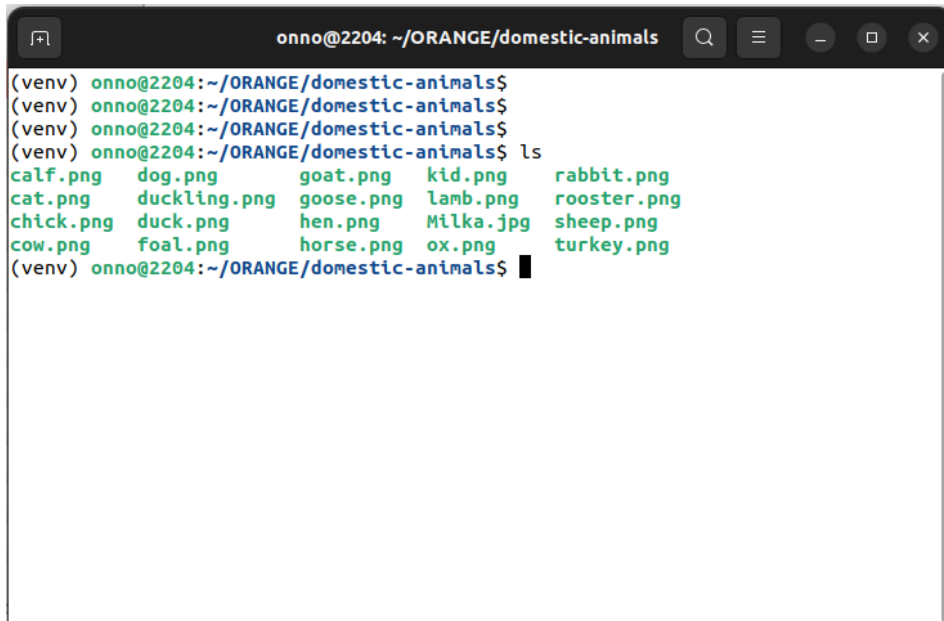
Contoh selanjutnya adalah menggunakan orange untuk mengklasifikasi gambar. Teknik ini banyak sekali digunakan untuk banyak hal, beberapa contoh yang paling sering kita gunakan misalnya,

- Mendeteksi nomor kendaraan di parkir. Jika kita perhatikan baik-baik, biasanya ada kamera untuk palang parkir untuk melihat nomor kendaraan agar mudah di cocokkan dengan tiket parkir yang kita dapat saat masuk.
- Di beberapa kantor biasanya yang memerlukan security yang tinggi, dipasang kamera untuk mendeteksi orang yang akan masuk. Ini akan di cocokkan dengan database wajah di pintu tersebut.
- KAI pernah juga memasang deteksi wajah untuk masuk penumpang ke stasiun KA. Entah apakah teknik ini diteruskan atau tidak.

Untuk bisa melakukan klasifikasi gambar binatang, kita perlu mengambil contoh gambar-gambar tersebut dari URL berikut,

<http://file.biolab.si/images/domestic-animals.zip>

Setelah file .zip di download dan dibuka maka akan tampak 20 file gambar dengan nama binatang sebagai nama file dalam format .png dalam folder domestic-animals seperti tampak pada gambar berikut.

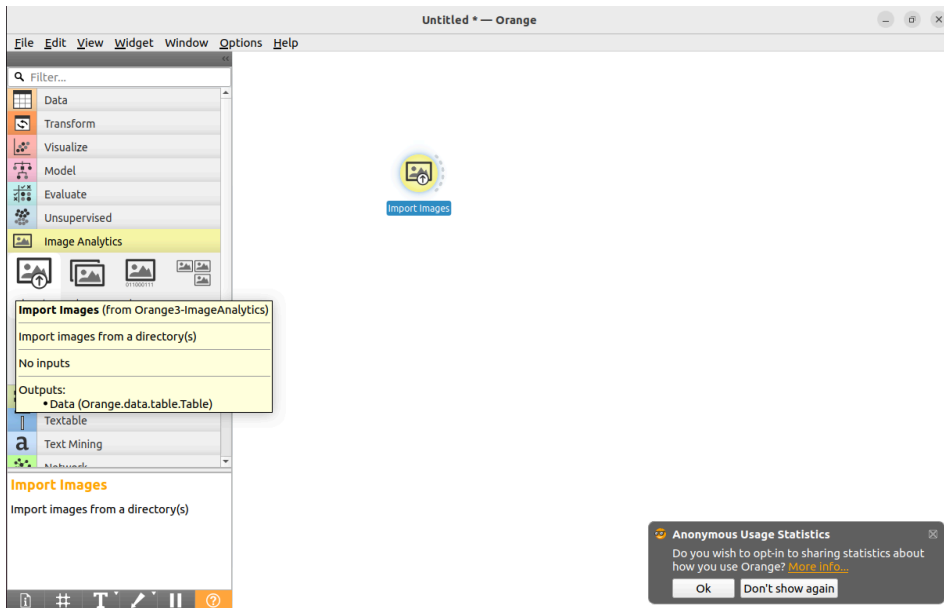
A terminal window titled 'onno@2204: ~/ORANGE/domestic-animals' with search, menu, and window control icons. The terminal shows a series of commands and their outputs. The 'ls' command lists 17 files: calf.png, dog.png, goat.png, kid.png, rabbit.png, cat.png, duckling.png, goose.png, lamb.png, rooster.png, chick.png, duck.png, hen.png, Milka.jpg, sheep.png, cow.png, foal.png, horse.png, ox.png, and turkey.png.

```
(venv) onno@2204:~/ORANGE/domestic-animals$  
(venv) onno@2204:~/ORANGE/domestic-animals$  
(venv) onno@2204:~/ORANGE/domestic-animals$  
(venv) onno@2204:~/ORANGE/domestic-animals$ ls  
calf.png  dog.png  goat.png  kid.png  rabbit.png  
cat.png   duckling.png  goose.png  lamb.png  rooster.png  
chick.png duck.png   hen.png   Milka.jpg sheep.png  
cow.png   foal.png   horse.png ox.png    turkey.png  
(venv) onno@2204:~/ORANGE/domestic-animals$
```

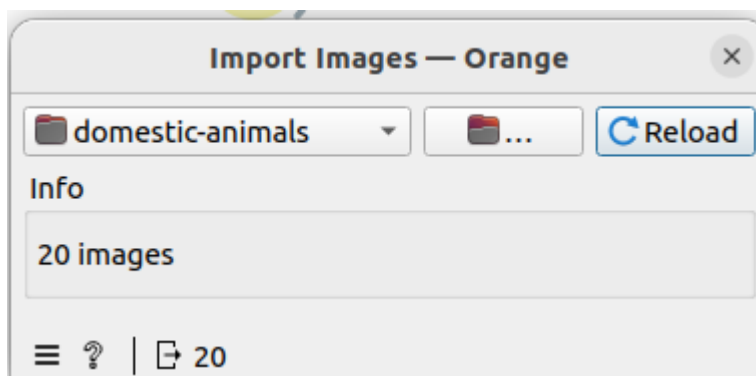
Gambar 7.14. Tampilan file .png nama binatang domestic-animals.

Untuk bisa menganalisa gambar, kita perlu memastikan bahwa orange yang kita gunakan sudah menginstalasi opsi Image Analytics. Jika belum menginstal opsi Image Analysis, kita perlu memastikan sambungan Internet berjalan dengan baik kemudian masuk ke menu options -> add ons dan pilih Image Analysis.

Dalam Image Analysis, ada Widget Import Load untuk memasukan gambar ke dalam orange. Widget Import Image dapat ditarik dari kumpulan menu sebelah kiri untuk di drag dan dimasukkan ke dalam program. Widget Import Image yang sudah di masukan ke program orange tampak seperti pada gambar.



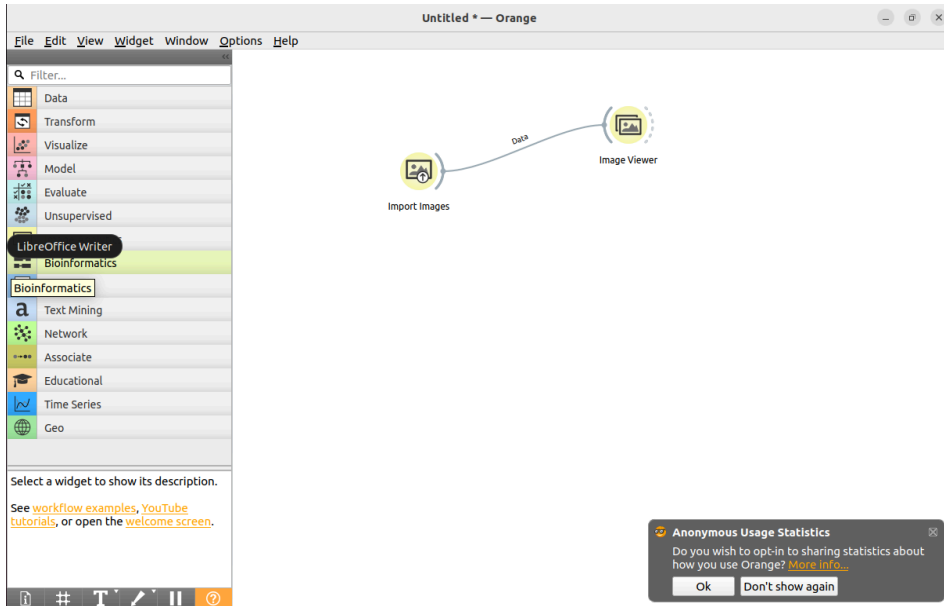
Gambar 7.15. Widget Import Image.



Gambar 7.16. Rangkuman Import Images

Melalui widget Import Images, kita dapat mengimport gambar kedalam Orange. Kebetulan gambar-gambar hewan peliharaan yang diambil dari Internet dimasukan dalam folder domestic-animals. Maka pada summary widget Import Images, kita dapat melihat ada 20 image binatang dari folder domestic-animals yang sudah berhasil dimasukan seperti tampak pada gambar.

Untuk melihat semua gambar bintang yang di import ke program, kita dapat menyambungkan Widget Image Viewer ke output dari Widget Import Images seperti tampak pada gambar,



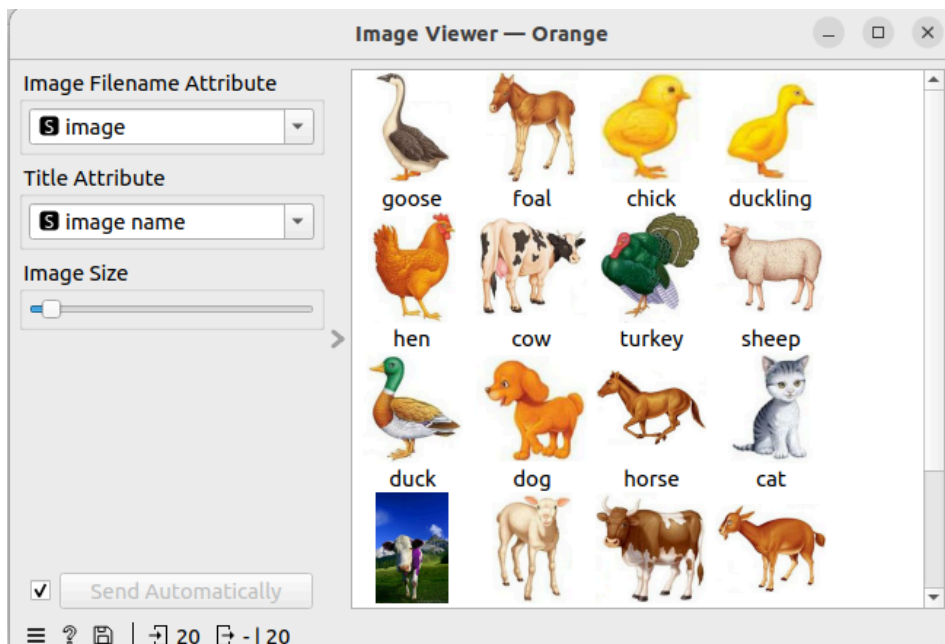
Gambar 7.17. Menyambungkan Widget Import Images ke Widget Image Viewers.

Berbagai hewan domestic animals yang berhasil di import dapat kita lihat dengan mudah dengan meng-klik widget Image Viewer. Gambar-gambar tersebut dapat kita lihat pada gambar berikutnya. Terlihat pada gambar ada,

Bebek, Itik, Kambing, Sapi, Kuda, Kalkun, Ayam, Kucing, Anjing, Anak Ayam, Anak Itik, dan masih banyak lagi.

Tugas algoritma yang kita kembangkan menggunakan orange adalah melakukan pengelompokan binatang-binatang tersebut, padahal sama sekali tidak ada informasi binatang sebut masuk ke kelompok mana, jumlah kakinya berapa, apakah ada sayapnya, apakah ada bulu-nya, apakah ada rambutnya, apakah gemuk, apakah ramping, apakah besar, apakah kecil, apakah berwarna putih, apakah berwarna kuning, apakah

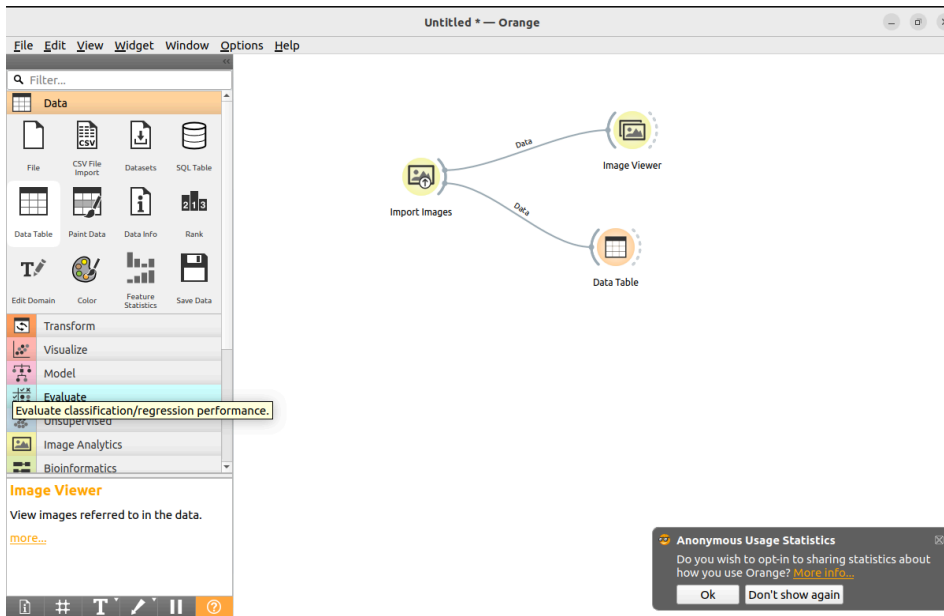
bisa berlari cepat dll. Detail gambar yang dimasukan terlihat pada gambar berikut ini,



Gambar 7.18. Kumpulan Hewan yang berhasil di Import.

Tugas pengelompokan ini akan menjadi sangat berat karena informasi yang ada hanya gambar dan sedikit informasi tentang teknis gambar tersebut. Untuk bisa melihat informasi teknis dari gambar yang di import, kita dapat menyambungkan widget Data Table ke output dari widget Import Image.

Pada gambar berikutnya, terlihat output widget Data Table kita sambungkan ke input widget Data Table, sehingga kita akan dapat memperoleh informasi terkait gambar yang berhasil di Import oleh widget Import Images ke program yang kita kembangkan. Untuk lebih jelasnya kita dapat melihat gambar berikutnya.



Gambar 7.19. sambungan output widget Import Images yang di sambungkan ke input widget Data Table.

Pada gambar di atas di perlihatkan input widget Data Table yang telah di sambungkan ke widget Import Image. Jika kita klik widget Data Table, maka kita akan melihat informasi / data tentang gambar yang di import dalam bentuk tabel.

Dari tampilan data yang diberikan dari widget Data Table, penting untuk kita sadari bahwa tabel yang tersebut hanya memuat sedikit sekali data gambar yang berhasil di import. Adapun data yang ada hanya, nama file gambar, besar gambar (byte), lebar dan tinggi (pixel). Untuk lebih jelasnya bisa dilihat pada gambar berikutnya.

Tidak ada informasi yang kita butuhkan untuk melakukan seperti, jumlah kaki, jumlah sayap, berbulu, berambut, berekor, berjengger (ayam), kecil, besar, imut2 (kucing) dsb ... oleh karena itu kita perlu mengekstrak fitur-fitur tersebut menggunakan algoritma deep learning atau neural network agar memudahkan kita dapat melakukan pengelompokan.

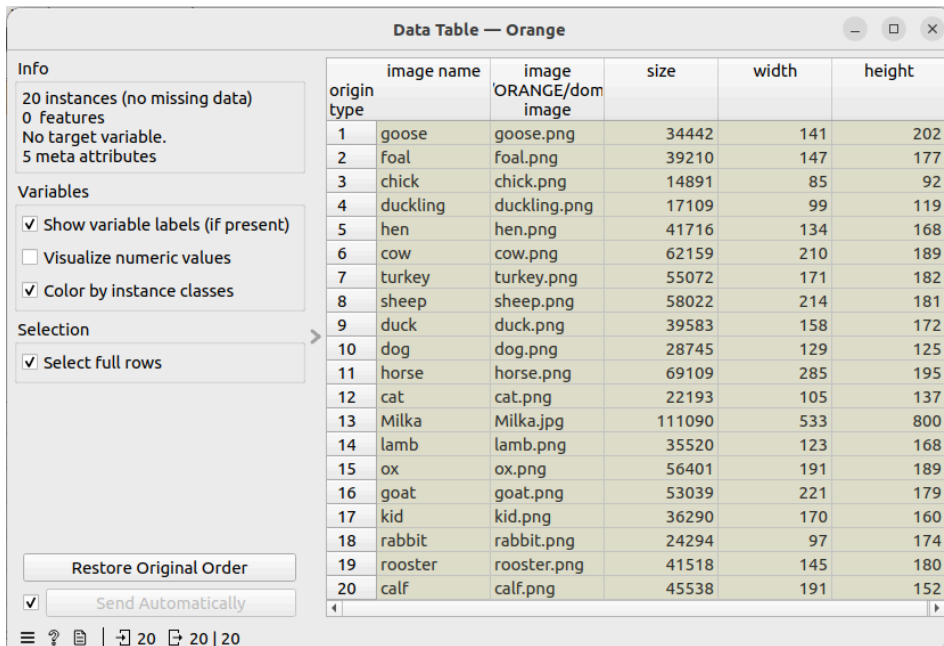


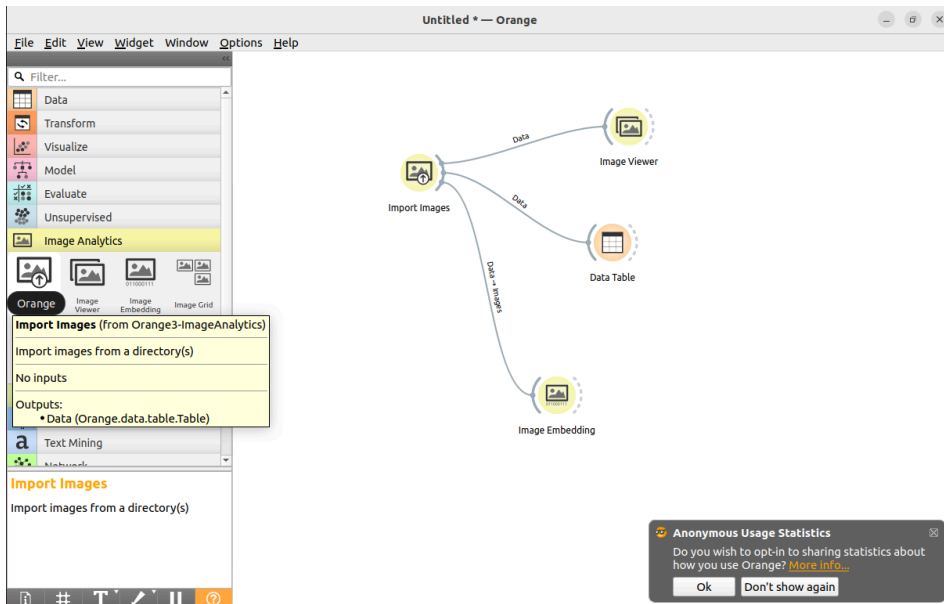
	image name	image 'ORANGE/dom image	size	width	height
1	goose	goose.png	34442	141	202
2	foal	foal.png	39210	147	177
3	chick	chick.png	14891	85	92
4	duckling	duckling.png	17109	99	119
5	hen	hen.png	41716	134	168
6	cow	cow.png	62159	210	189
7	turkey	turkey.png	55072	171	182
8	sheep	sheep.png	58022	214	181
9	duck	duck.png	39583	158	172
10	dog	dog.png	28745	129	125
11	horse	horse.png	69109	285	195
12	cat	cat.png	22193	105	137
13	Milka	Milka.jpg	111090	533	800
14	lamb	lamb.png	35520	123	168
15	ox	ox.png	56401	191	189
16	goat	goat.png	53039	221	179
17	kid	kid.png	36290	170	160
18	rabbit	rabbit.png	24294	97	174
19	rooster	rooster.png	41518	145	180
20	calf	calf.png	45538	191	152

Gambar 7.20. Tampilan Tabel data gambar yang berhasil di import.

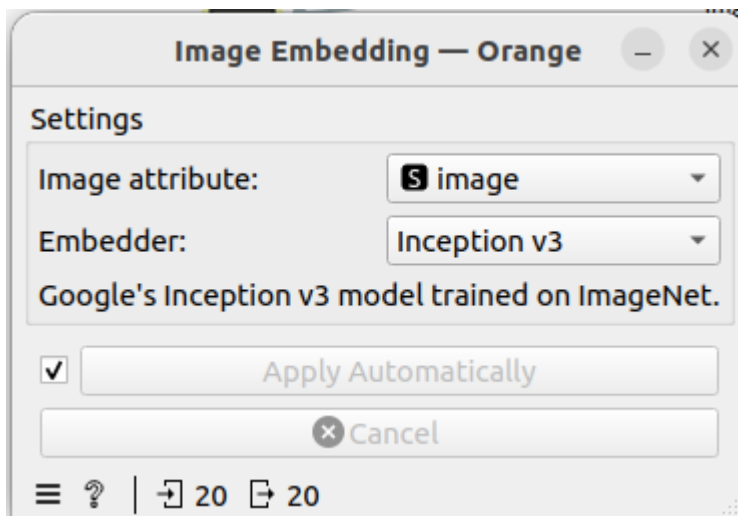
Pada gambar di atas di perhatikan tabel asli dari gambar yang di import ke dalam program yang kita buat. Terlihat dengan jelas tidak ada data fitur pada masing-masing gambar.

Untuk mengextrak fitur kita perlu mengirimkan semua gambar tersebut ke mesin neural network sehingga diperoleh fitur dari masing-masing gambar yang membedakan dengan gambar yang satu dengan yang lain. Ada beberapa pilihan mesin neural network yang dapat kita gunakan untuk melakukan ekstraksi tersebut.

Pada kesempatan ini, kita akan menggunakan mesin neural network Google Inception v3. Untuk itu kita perlu menyambung keluaran dari widget Import Images ke input widget Image Embedding seperti tampak pada Gambar. Pada widget Image Embedding kita dapat memilih beberapa mesin neural network untuk melakukan kalkulasi, pada kesempatan ini kita memilih mesin yang disediakan oleh Google yaitu Google Inception v3.



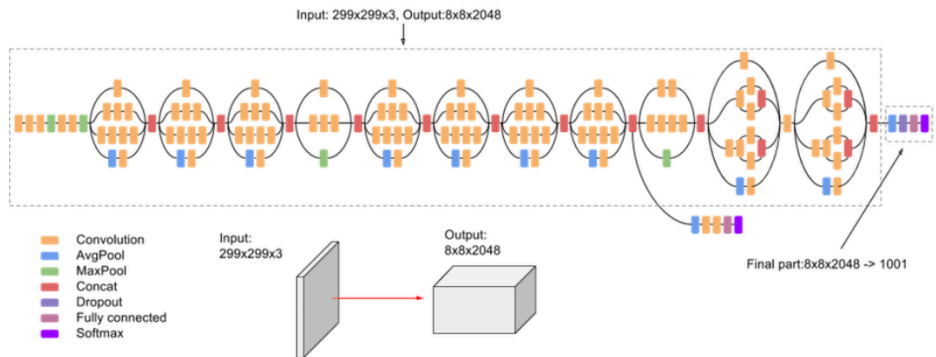
Gambar 7.21. sambukan widget Import Images ke widget Image Embedding.



Gambar 7.22. Penggunaan Opsi Google Inception v3 pada widget Image Embedding

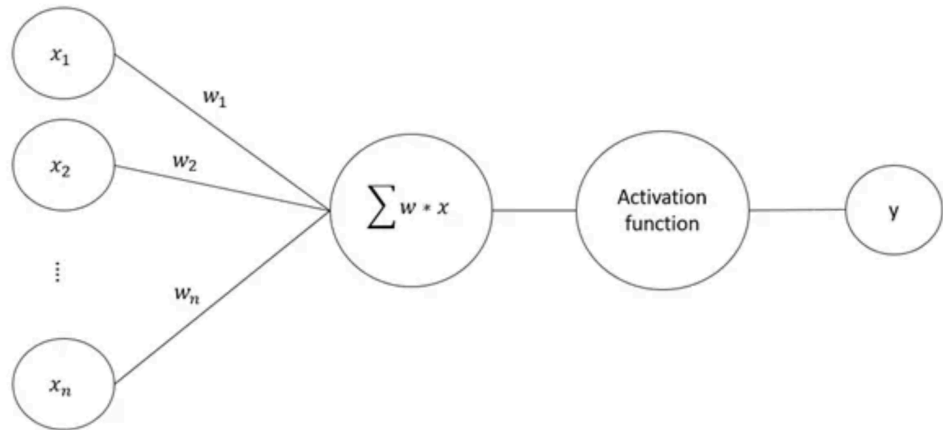
Google Inception v3 adalah model jaringan saraf konvolusional (CNN) yang membantu analisis gambar dan deteksi objek. Google Inception v3 memiliki 48 lapisan, termasuk blok penyusun simetris dan asimetris, dan

dapat mencapai akurasi lebih dari 78,1% pada kumpulan data ImageNet. Model ini diperkenalkan selama ImageNet Recognition Challenge dan didasarkan pada makalah Rethinking the Inception Architecture for Computer Vision oleh Szegedy, et. al. . Gambaran arsitektur mesin neural network tersebut terlihat pada gambar berikut ini.



Gambar 7.23. Arsitektur Jaringan Google Inception v3.

Apa hubungannya neural network berlapis-lapis ini dengan regresi linear. Untuk bisa melihat hubungan sebuah node neural network dengan regresi linear ada baiknya kita melihat gambar berikut ini.



Gambar 7.24. Regresi Linier dalam Neural Network

Dalam gambar di atas di perlihatkan hubungan antara regresi linier dengan neural network. Pada gambar di atas di perlihatkan ada banyak input X (x1, x2, ..., xn) yang dimasukkan ke fungsi. Masing-masing input X dikalikan dengan pengali w masing-masing (w1, w2, ... wn). Jumlah dari semua input yang sudah dikalikan dengan pengali W di masukan ke

activation function, yang bisa berupa sebuah nilai, misalnya B. Maka keluarannya adalah Y. Jika ditulis dalam persamaan matematika adalah sebagai berikut,

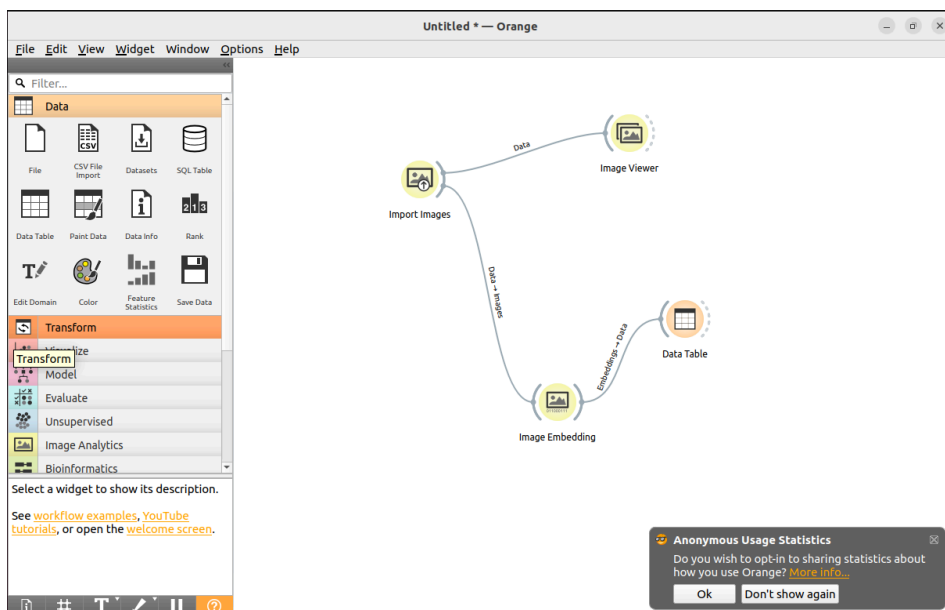
$$x_1 w_1 + x_2 w_2 + \dots x_n w_n + B = Y \quad 7.1$$

Jika kita perhatikan baik-baik, maka persamaan tersebut kira-kira hampir sama dengan persamaan linier yang kita familiar, yaitu,

$$A X + B = Y \quad 7.2$$

hanya bedanya, disini A X-nya terdiri dari banyak variabel $x_1 w_1, x_2 w_2, \dots x_n w_n$.

Melalui proses regresi linear dalam neural network ini, akan diekstraksi berbagai fitur yang ada pada gambar.



Gambar 7.25. Menyambungkan Output widget Image Embedding ke input widget Data Table.

Fitur yang diperoleh dari Google Inception v3 yang di set dalam widget Image Embedding dikirim langsung ke widget Data Table seperti tampak pada gambar.

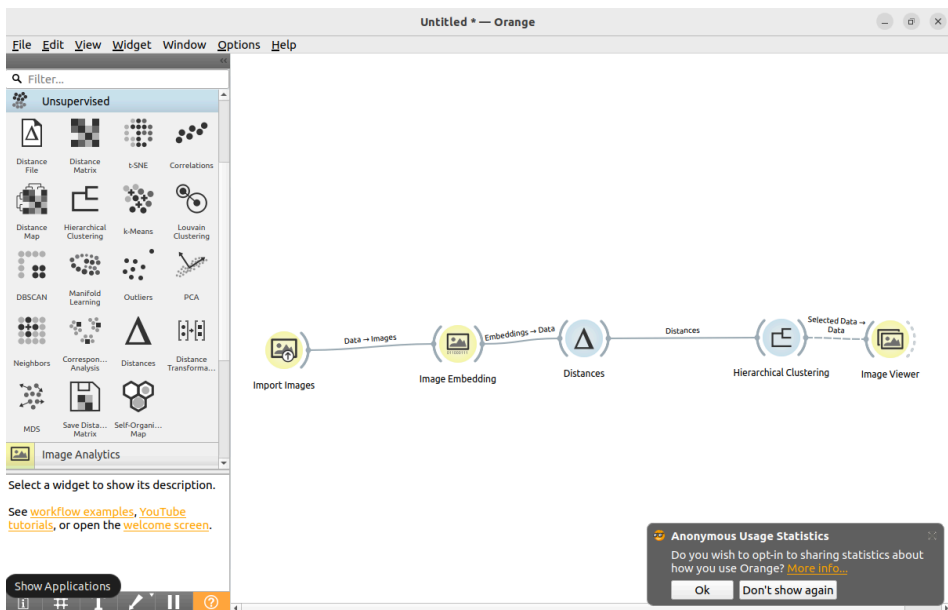
	hidden origin type	image name	image	size	width	height	n0 True	n1 True	n2 True	n3 True
1		goose	goose.png	34442	141	202	0.152611	0.1434	0.268933	0.548
2		foal	foal.png	39210	147	177	0.340434	0.360381	0.240731	0.329
3		chick	chick.png	14891	85	92	0.273149	0.143223	0.0883515	0.366
4		duckling	duckling.png	17109	99	119	0.368307	0.0462069	0.181636	0.4
5		hen	hen.png	41716	134	168	0.0623566	0.176984	0.0311257	0.738
6		cow	cow.png	62159	210	189	0.332465	0.319818	0.119457	0.289
7		turkey	turkey.png	55072	171	182	0.0719494	0.0737333	0.295801	0.543
8		sheep	sheep.png	58022	214	181	0.115838	0.226645	0.274766	0.685
9		duck	duck.png	39583	158	172	0.183795	0.0271835	0.082604	0.0570
10		dog	dog.png	28745	129	125	0.0673088	0.0701765	0.206464	0.48
11		horse	horse.png	69109	285	195	0.367071	0.0921347	0.460571	0.214
12		cat	cat.png	22193	105	137	0.0477391	0.0314866	0.170575	0.533
13		Milka	Milka.jpg	111090	533	800	0.291256	0.0888083	0.117755	0.423
14		lamb	lamb.png	35520	123	168	0.20294	0.226755	0.132112	0.238
15		ox	ox.png	56401	191	189	0.10597	0.133489	0.238158	0.568
16		goat	goat.png	53039	221	179	0.359154	0.110843	0.135006	0.176
17		kid	kid.png	36290	170	160	0.304173	0.137448	0.103495	0.490
18		rabbit	rabbit.png	24294	97	174	1.18352	0.0764723	0.0888809	0.390
19		rooster	rooster.png	41518	145	180	0.0293618	0.0419532	0.18331	0.170
20		calf	calf.png	45538	191	152	0.215285	0.183625	0.054805	0.268

Gambar 7.26. Tampilan widget Data Tabel berisi data hasil Google Inception v3

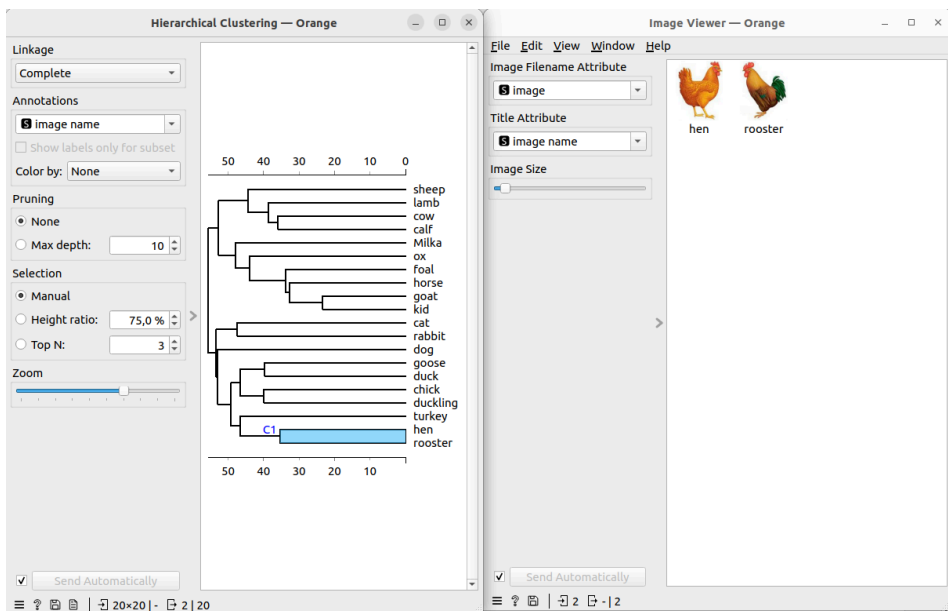
Hasil ekstraksi fitur gambar dari Google Inception v3 diperlihatkan di pada gambar di atas. Terlihat dengan jelas kolom yang berwarna putih adalah hasil ekstraksi fitur dari gambar-gambar yang kita import. Memang nama kolom-nya adalah n0, n1, n2, n3 n2013 .. bukan jumlah kaki, jumlah sayap dsb.

Dengan adanya kolom fitur ini, kita dapat dengan mudah mengelompokkan masing-masing gambar dengan menghitung fitur-fitur yang berdekatan menjadi satu kelompok yang sama. Tentunya dalam proses pengelompokan ini kita harus mengoptimalkan jarak antar nilai dari ribuan kolom.

Proses pengoptimalan ribuan kolom ini dapat dilakukan dengan menggunakan widget Distance yang merupakan bagian dari Unsupervised Learning, untuk menganalisa data yang tidak diberi label. Atau dari konsep persamaan linier $Ax+B=Y$ adalah data yang tidak ada Y-nya, jadi hanya X saja yang tersedia. Algoritma berusaha mengestimasi kelompok dari X tersebut. Nantinya kelompok ini dapat kita beri label sesuai keinginan kita.



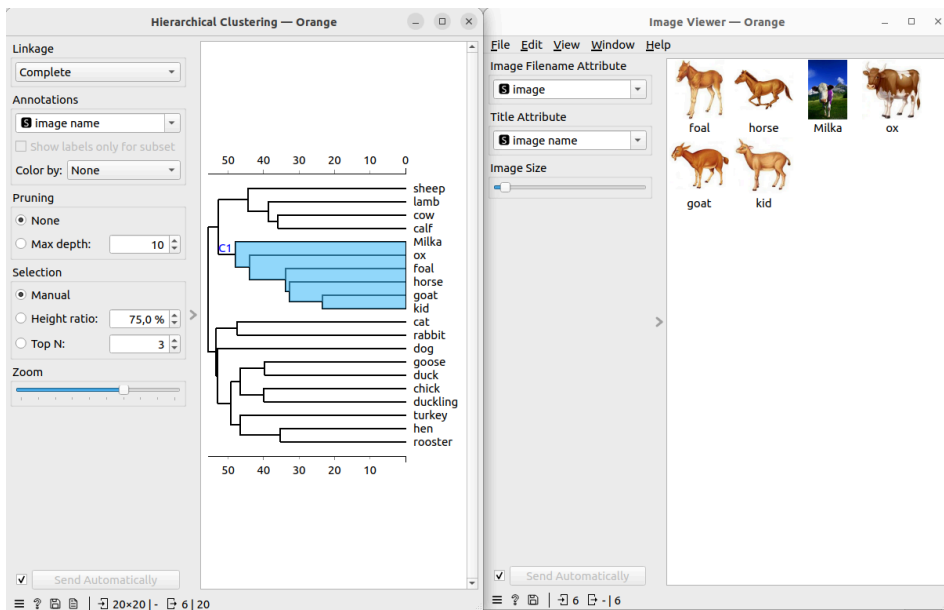
Gambar 7.27. Seluruh program lengkap dengan visualisasi oleh widget Clustering dan widget Image Viewer.



Gambar 7.28. Visualisasi Clustering dan Image Viewers

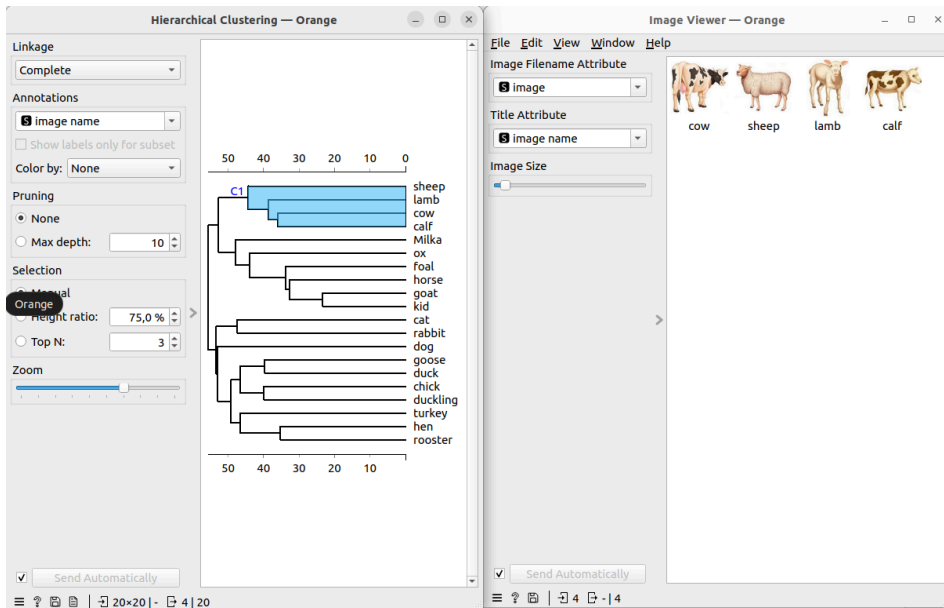
Tampilan Hirarki Clustering, pengelompokan, dari keluaran widget Distance dan gambar yang dihasilkan tampak pada gambar di atas. Yang menarik pada widget Clustering kita dapat memilih cluster yang kita inginkan dengan cara mengklik di tengah-tengah kelompok yang kita inginkan. Jika sebuah cluster di klik, maka pada Image Viewer akan ditampilkan gambar / image dari cluster tersebut.

Dalam contoh di atas kita memilih sebuah cluster dengan dengan dua anggota yang terletak paling bawah. Dengan mengklik cluster tersebut, maka akan tampak dua gambar ayam di widget Image Viewer yang memperoleh masukan dari cluster yang kita pilih.



Gambar 7.29. Pemilihan Cluster Tengah dengan 6 anggota.

Pada gambar diatas, kita memilih cluster yang terletak agak di tengah dengan anggota enam gambar. Setelah cluter tersebut dipilih, terlihat dengan jelas gambar ditampilkan pada widget Image Viewer. Disini berbeda, dengan sebelumnya ke enam gambar yang dipilih ternyata merupakan keluarga hewan berkaki empat seperti tampak pada gambar di widget Image Viewer.



Gambar 7.30. Cluster paling atas.

Pada analisis gambar kali ini, kita akan mengkhhususkan perhatian pada satu kelompok gambar tertentu. Kelompok gambar yang dipilih ini merupakan cluster yang terletak di posisi paling atas dalam tampilan hasil pengelompokan. Cluster ini terdiri dari empat gambar yang memiliki kesamaan tertentu.

Dari keempat gambar dalam cluster tersebut, kita dapat mengamati dengan jelas bahwa semua gambar menampilkan hewan berkaki empat. Ini menunjukkan bahwa algoritma pengelompokan berhasil mengidentifikasi kesamaan ciri fisik pada keempat gambar tersebut. Namun, jika kita bandingkan dengan kelompok gambar sebelumnya, terdapat perbedaan yang menarik. Pada kelompok gambar sebelumnya, mayoritas hewan memiliki warna coklat. Sedangkan pada cluster ini, warna putih mendominasi, seperti pada gambar sapi, biri-biri, dan kambing.

Hasil pengelompokan yang dihasilkan oleh algoritma distance umumnya cukup masuk akal dan sesuai dengan ekspektasi kita. Namun, kita perlu menyadari bahwa algoritma ini tidak selalu sempurna. Terkadang, algoritma dapat melakukan kesalahan dalam mengidentifikasi gambar. Sebagai contoh, gambar domba kecil yang diperbesar secara signifikan

dapat terlihat mirip dengan gambar sapi yang diambil dari jarak jauh. Hal ini menunjukkan bahwa algoritma masih memiliki keterbatasan dalam mengenali detail-detail kecil pada gambar.

Teknik pengelompokan gambar dan pola pikir yang digunakan dalam analisis ini memiliki potensi yang sangat besar untuk dikembangkan lebih lanjut. Kemampuan untuk mengenali pola, wajah, dan gambar secara otomatis memiliki banyak manfaat dalam berbagai bidang yang berkaitan dengan pengolahan gambar. Misalnya, teknologi ini dapat digunakan untuk mengorganisir album foto, melakukan pencarian gambar yang lebih efisien, atau bahkan untuk mengembangkan sistem keamanan berbasis pengenalan wajah.

BAB 8: Prediksi Data & Situasi Kompleks

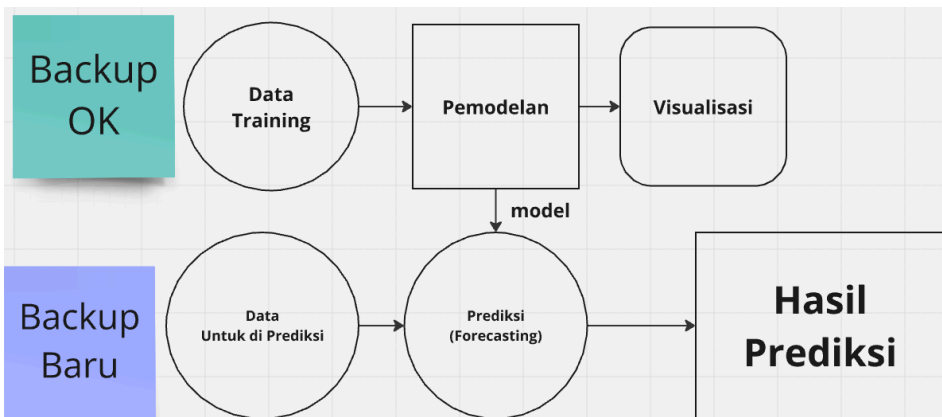
Penerapan teknologi mutakhir seperti kecerdasan buatan (AI) dan big data di Indonesia telah mengubah cara pemerintah melayani masyarakat. Melalui analisis data yang mendalam, pemerintah dapat merancang kebijakan yang lebih tepat sasaran, seperti program penanggulangan kemiskinan dan mitigasi bencana. AI juga meningkatkan efisiensi pelayanan publik, seperti pembuatan e-KTP dan deteksi penipuan. Selain itu, teknologi ini membantu dalam perencanaan pembangunan yang lebih baik, meningkatkan transparansi pemerintahan, dan membuka peluang untuk layanan publik yang lebih personal. Dengan demikian, AI dan big data berperan penting dalam membangun Indonesia yang lebih maju dan sejahtera.

Pola pikir prediksi

Untuk melakukan sebuah prediksi (dukun), sebetulnya pola pikir yang perlu diadopsi tidak terlalu sulit. Logika tersebut adalah sebagai berikut,

- Kita **membuat model** dari data history yang ada sebelumnya. Dalam bahasa data science, model diperoleh dari training data.
- Dari model yang diperoleh, kita dapat **melakukan prediksi** dengan cara menggunakan model dengan data yang tidak ada di history / training data. Hasilnya adalah prediksi dari data yang kita masukan terhadap model yang kita peroleh dari training data.

Logika berfikir ini dapat digambarkan seperti gambar di bawah ini.



Gambar 8.1. Logika Berfikir untuk Prediksi

Untuk mengukur dampak suatu kebijakan, kita bisa melakukan simulasi sederhana. Caranya adalah dengan menerapkan kebijakan tersebut pada data historis. Misalnya, kita punya data tentang kondisi masyarakat di masa lalu. Nah, data ini kita jadikan sebagai 'bahan latihan' untuk membuat sebuah model. Setelah model ini jadi, kita coba terapkan pada data yang belum kita gunakan sebelumnya. Dengan membandingkan hasil prediksi model dengan data aslinya, kita bisa melihat kira-kira apa saja yang akan terjadi jika kebijakan itu benar-benar diterapkan.

Misalnya, kita ingin tahu apa yang terjadi kalau kita memperpanjang masa pendidikan seseorang selama satu tahun. Kita bisa mencoba hal ini dengan cara mengubah data historis kita. Kita 'tambahkan' satu tahun masa pendidikan untuk setiap orang dalam data tersebut, lalu kita masukkan data yang sudah kita ubah ini ke dalam model yang sudah kita buat tadi. Hasil prediksinya akan menunjukkan perkiraan perubahan kondisi masyarakat setelah kebijakan perpanjangan masa pendidikan diterapkan. Kita bisa melihat, misalnya, berapa banyak orang yang tadinya sangat miskin menjadi tidak terlalu miskin setelah masa pendidikannya diperpanjang.

Konsekuensi menaikkan wajib belajar selama 1 tahun adalah alokasi anggaran yang perlu dilakukan dikalikan dengan jumlah rakyat yang harus disubsidi sekolahnya. Dalam bahasa birokrasi misalnya,

Dalam program pemerintah daerah ini, kita Insya Allah dapat mengubah $x\%$ rakyat dari status sangat miskin menjadi miskin, dan $y\%$ dari miskin menjadi sejahtera dengan cara menambahkan waktu wajib belajar selama 1 tahun, dengan beban anggaran Rp. y juta, atau sekitar $z\%$ dari total anggaran belanja daerah.

Contoh Sederhana Klasifikasi Kemiskinan

Dalam contoh analisis data ini, kita akan mencoba mengelompokkan data kemiskinan. Tujuan utamanya adalah untuk melihat apakah hasil pengelompokan yang dilakukan oleh komputer sama dengan pengelompokan yang biasanya dilakukan secara manual. Dengan kata lain, kita akan menguji kemampuan komputer dalam mengidentifikasi pola-pola kemiskinan.

Analisis data kemiskinan ini memiliki implikasi yang luas di dunia nyata. Seringkali, data tentang kemiskinan dimanipulasi untuk kepentingan tertentu, misalnya agar orang yang tidak berhak mendapatkan bantuan sosial justru mendapatkannya. Dengan menggunakan teknik klasifikasi data, kita dapat mendeteksi adanya manipulasi data tersebut. Hal ini sangat penting untuk memastikan bahwa bantuan sosial benar-benar tepat sasaran dan diterima oleh mereka yang benar-benar membutuhkan.

Data kemiskinan yang digunakan merupakan rekayasa penulis, dengan kolom sebagai berikut,

Penghasilan, Pendidikan, Pendidikan (Lama), Rumah, Motor, Mobil, Anak, Utang, Nama, Klasifikasi, Stunting.

Disini fitur yang diambil adalah penghasilan, pendidikan, lama pendidikan, rumah, motor, mobil bahkan utang. Ada dua (2) kolom yang dapat digunakan sebagai target untuk proses selanjutnya yaitu kolom

klasifikasi
stunting

Perkiraan data training di perlihatkan pada gambar di bawah ini.

	PENDIDIKAN	PENDIDIKAN (LAMA)	RUMAH	MOTOR	MOBIL	ANAK	UTANG	NAMA	KLASIFIKASI	STUNTING
1	SD	6	0	0	0	4	0	Amir	miskin	ya
2	SD	6	0	0	0	5	0	Badu	miskin	ya
3	SD	6	0	0	0	3	0	Candra	miskin	ta
4	SD	7	0	0	0	5	0	Dedi	miskin	ya
5	SD	7	0	0	0	3	0	Eddy	miskin	ya
6	SD	6	0	0	0	3	0	Fahri	miskin	tidak
7	SD	7	0	0	0	4	0	Gagah	miskin	tidak
8	SMP	9	0	0	0	2	0	Hadi	miskin	tidak
9	SMP	9	0	0	0	2	0	Ira	miskin	tidak
10	SMP	10	0	1	0	4	0	Jaja	miskin	tidak
11	SMP	10	1	1	0	3	0	Kaka	miskin	tidak
12	SMP	9	1	1	0	3	0	Lala	miskin	tidak
13	SMP	9	1	0	0	4	0	Mami	miskin	tidak
14	SMA	11	1	1	1	0	100000	Nani	miskin	ya
15	SMA	11	1	1	1	1	100000	Opa	miskin	ya
16	SMA	11	1	1	1	2	100000	Qra	miskin	ya
17	SMA	12	1	1	1	2	200000	Rani	miskin	ya
18	SMA	12	1	1	1	3	0	Sasa	sejahtera	tidak
19	SMA	12	1	1	1	3	0	Tati	sejahtera	tidak
20	SMA	13	1	1	1	2	0	Udi	sejahtera	tidak
21	UNIV	14	1	1	1	2	0	Vava	sejahtera	tidak
22	UNIV	14	1	1	1	2	0	Wawa	sejahtera	tidak
23	UNIV	18	1	1	1	1	0	Xaxa	sejahtera	tidak
24	UNIV	19	1	1	1	2	0	Yaya	sejahtera	tidak
25	NO	0	0	0	0	4	0	Mamir	sangatmiskin	tidak
26	NO	1	0	0	0	6	0	Mbanir	sangatmiskin	tidak

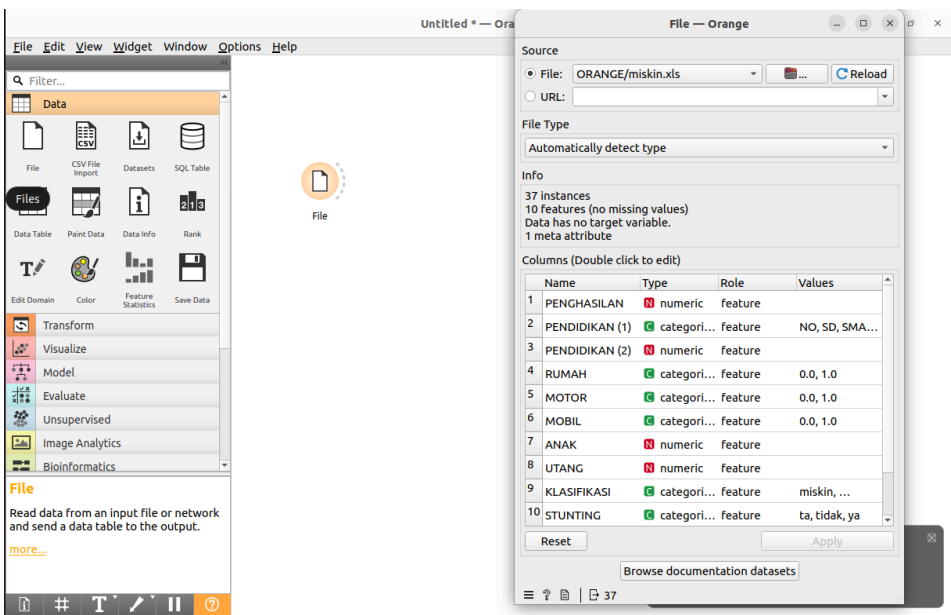
Gambar 8.2. Data Training Kemiskinan.

Data yang kita gunakan untuk melatih model kita sebenarnya belum cukup detail. Misalnya, data tentang rumah seharusnya mencakup informasi yang lebih spesifik, seperti jumlah kamar, jenis material bangunan, dan ukuran listrik. Namun, untuk tahap awal percobaan ini, kita akan menyederhanakan data dengan hanya menghitung jumlah rumah.

Tujuan yang ingin kita capai dalam analisis ini sebenarnya ada dua: mengklasifikasikan rumah tangga dan memprediksi stunting. Sayangnya, kebanyakan metode analisis kesulitan memprediksi dua hal sekaligus dengan akurat. Oleh karena itu, dalam model kita, kita akan fokus pada satu tujuan saja terlebih dahulu.

Pada tahap ini, kita ingin melihat apakah hasil klasifikasi yang kita dapatkan dari komputer (misalnya, sangat miskin, miskin, atau sejahtera) sesuai dengan data yang sudah kita siapkan sebelumnya. Untuk melakukan ini, kita akan menggunakan suatu tool yang disebut "widget distance" yang termasuk dalam kelompok metode analisis data *unsupervised learning*.

Pertama-tama kita load file miskin.xls yang berisi data yang kita buat menggunakan widget Load Data seperti tampak pada gambar.



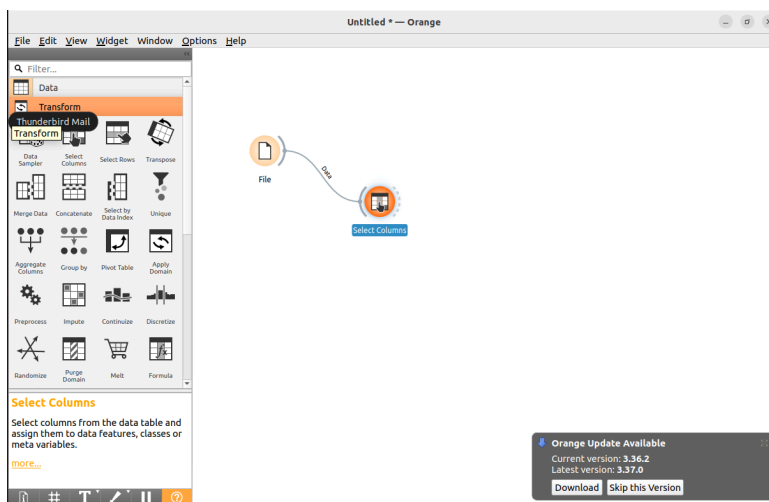
Gambar 8.3. Penampakan Data miskin.xls

Pada gambar kita dapat melihat rangkuman dari data miskin.xls yang di load menggunakan widget File. Secara umum terdeteksi ada 37 line data (instances), 10 fitur (tidak ada nilai yang kosong), nama kolom-kolom seperti penghasilan, pendidikan, rumah, mobil dll. Yang menarik untuk diamati bahwa semua kolom terdeteksi sebagai fitur, tidak ada kolom target. Kebetulan dalam data yang kita upload sebetulnya ada dua (2) kolom yang sebetulnya target, yaitu,

- Klasifikasi (variabelnya adalah sejahtera, miskin dan sangat miskin)
- Stunting (variabelnya adalah ya dan tidak).

Analisis data yang kita lakukan saat ini memerlukan penyesuaian jenis kolom agar lebih akurat. Algoritma yang kita gunakan memiliki keterbatasan dalam melakukan regresi untuk dua target secara bersamaan. Oleh karena itu, kita akan melakukan pemodelan secara bertahap, dimulai dengan klasifikasi terlebih dahulu.

Untuk mengubah peran suatu kolom, misalnya dari fitur menjadi target atau bahkan diabaikan, kita dapat memanfaatkan widget Select Columns. Caranya adalah dengan menghubungkan output dari widget File ke input dari widget Select Columns. Dengan demikian, kita dapat mengontrol jenis kolom mana yang akan digunakan dalam analisis.

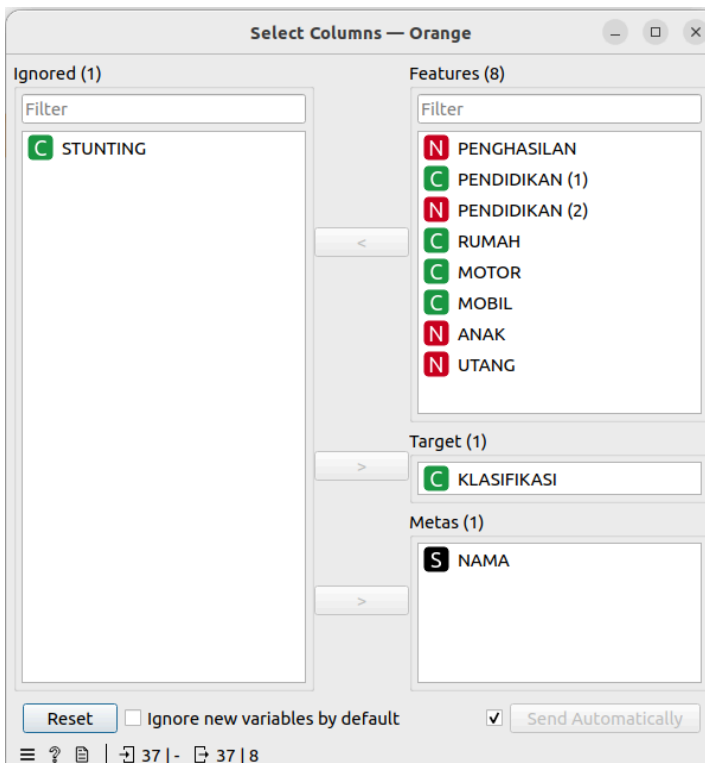


Gambar 8.4. Keluaran Widget File di sambung ke masukan Widget Select Columns.

Objektif analisa pengelompokan kemiskinan untuk kesempatan pertama ini adalah untuk memastikan bahwa klasifikasi yang dibuat di data training dengan klasifikasi yang dilakukan oleh mesin sama, paling tidak error-nya tidak banyak. Untuk itu pada Widget Select Columns, kita akan membuat supaya,

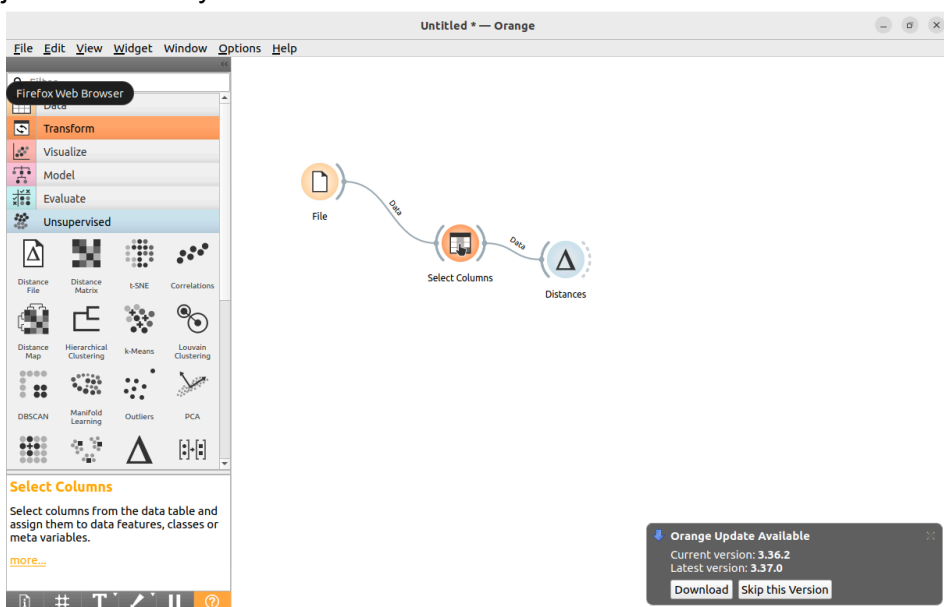
- Kolom Klasifikasi menjadi Target, untuk di cocokan dengan target hasil analisa komputer.
- Kolom nama menjadi metadata saja, karena tidak berfungsi dalam perhitungan.
- Kolom Stunting sementara dimasukan ke Ignored, karena algoritma yang ada tidak memungkinkan untuk menganalisa dua (2) target sekaligus.

Penampakan widget Select Columns, setelah kolom yang ada di sesuaikan dengan kebutuhan analisis kita adalah seperti pada gambar.



Gambar 8.5. Konfigurasi di widget Select Columns.

Setelah memastikan bahwa data pada setiap kolom memiliki fungsi yang tepat sebagai fitur atau target, langkah selanjutnya adalah mengelompokkan data berdasarkan kemiripannya. Proses pengelompokkan ini dikenal sebagai clustering. Dengan menggunakan tool yang sesuai, seperti widget Distance pada Orange, kita dapat membandingkan hasil pengelompokkan yang dihasilkan oleh sistem dengan kelompok yang telah ditentukan sebelumnya oleh peneliti. Proses ini merupakan bagian dari *unsupervised learning*, di mana sistem belajar menemukan pola dalam data tanpa diberikan label atau kategori yang jelas sebelumnya.

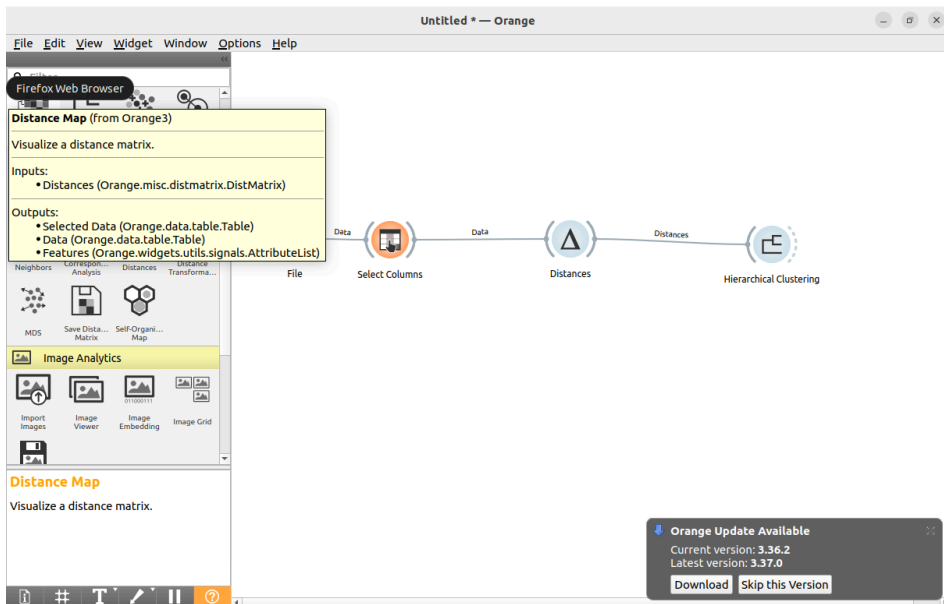


Gambar 8.6. Menyambungkan Keluaran widget Select Columns ke masukan widget Distance.

Untuk memvisualisasi hasil perhitungan kita dapat menyambungkan keluaran dari widget Distance untuk menjadi masukan dari widget Hierarchical Clustering agar kita dapat melihat dengan jelas struktur kelompok yang dibuat. Agar lebih seru lagi, kita dapat menyambungkan keluaran dari widget Hierarchical Clustering.

Penampakan akhir dari program untuk melakukan pengelompokan kategori kemiskinan terlihat dalam gambar berikut. Secara umum ada 3 bagian besar, yaitu,

- Input data - data sering kali harus diperoleh melalui kegiatan pengambilan sensus, atau bahkan mata-mata.
- Pemodelan - disini kita menggunakan widget Distance karena pemodelan menggunakan metode unsupervised learning.
- Visualisasi - menggambarkan hasil akhir agar mudah dimengerti oleh manusia.

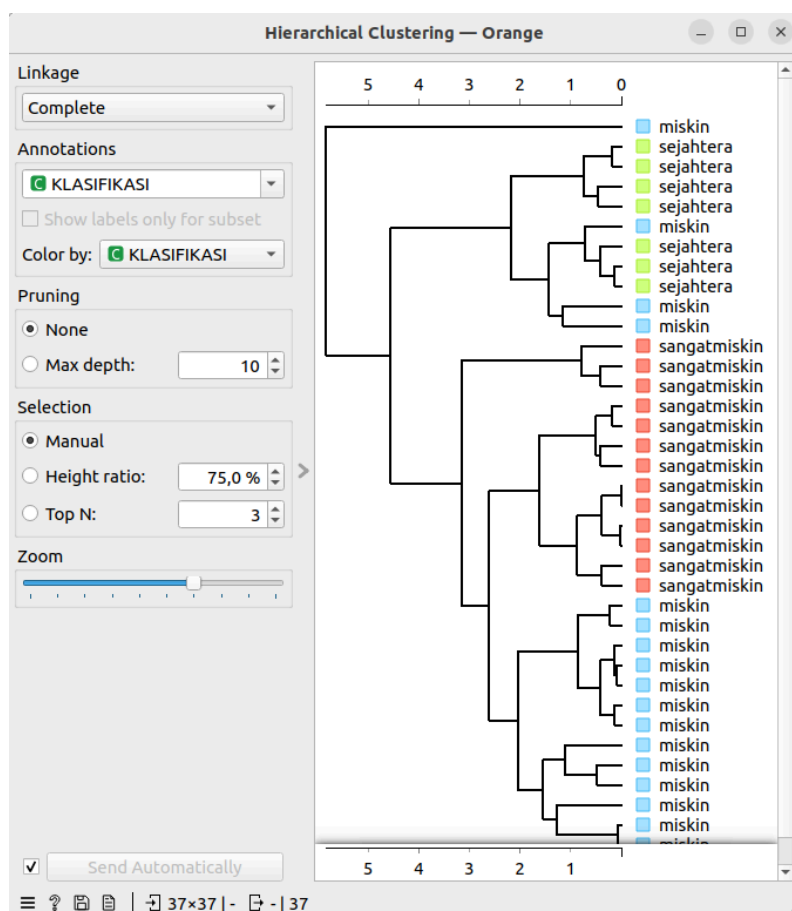


Gambar 8.7. Program untuk pengelompokan (*clustering*) data.

Dalam gambar berikutnya di perlihatkan hasil visualisasi melalui widget Hierarchical Clustering. Beberapa hal yang penting pada saat kita melakukan visualisasi ini adalah,

- Annotations - pada gambar terlampir di pilih kolom klasifikasi untuk menjadi nilai / nama yang akan digunakan dalam anggota cluster. Pilihan lain yang dapat kita gunakan untuk menjadi nama dari anggota cluster adalah nama dari kolom nama, agar kita dapat tahu nama orang anggota cluster tersebut.
- Color By - gunakan nilai kolom klasifikasi untuk membedakan warna setiap kelompok. Karena dalam kolom klasifikasi ada tiga (3) nilai, maka kita akan mempunyai tiga (3) pilihan warna.

Penampilan gambar kelompok (cluster) berhasil ditangkap seperti pada gambar berikut.



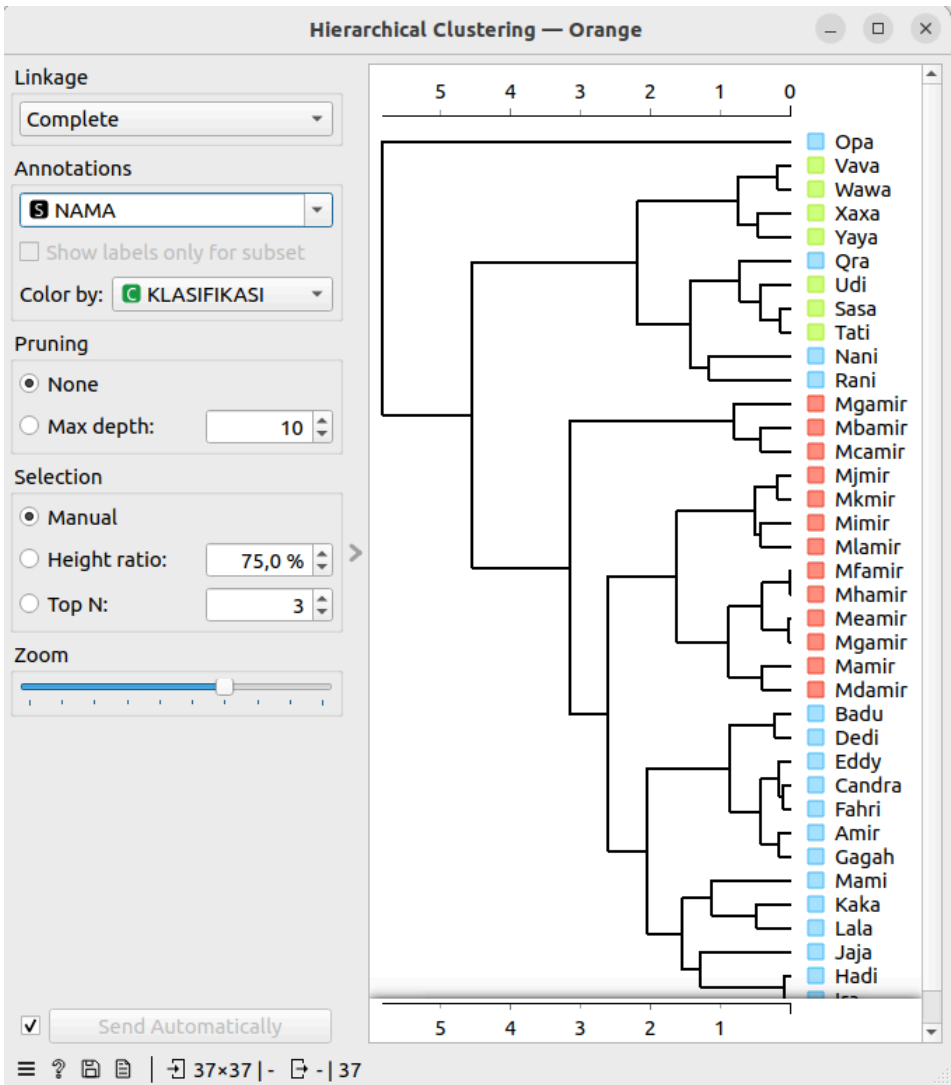
Gambar 8.8. Tampilan Cluster / Kelompok.

Dalam gambar *cluster* terlihat dengan jelas bahwa klasifikasi miskin & sangatmiskin hasil antara program dan data penulis sama persis. Tidak ada kesalahan klasifikasi antara penulis dan program.

Yang jadi masalah adalah dalam klasifikasi sejahtera, terlihat ada beberapa yang diklasifikasi sebagai miskin oleh penulis, ternyata terklasifikasi sebagai sejahtera oleh program. Ada paling tidak empat kasus salah klasifikasi diklasifikasi sejahtera. Di dunia nyata, kesalahan klasifikasi seperti ini juga sering terjadi, ada orang-orang yang harusnya

masih sejahtera ternyata oleh operator dimasukan kedalam kategori miskin untuk satu dan lain hal keperluan.

Untuk mengecek siapa saja yang salah, kita dapat mengganti Annotations menjadi nama bukan klasifikasi. Untuk memudahkan parameter color by tetap menggunakan Klasifikasi untuk menentukan warna agar nanti kita dapat memilahnya berdasarkan warna. Hasilnya kita dapat lihat pada gambar berikut ini.



Gambar 8.9. Penggunaan Anotasi Nama di Cluster.

Dalam gambar di atas, terlihat empat (4) nama yang bermasalah, yaitu, Opa, Qra, Nani dan Rani. Keempat nama ini dalam data dimasukkan dalam kategori miskin padahal oleh komputer keempat orang ini harusnya masuk kategori sejahtera.

Langkah selanjutnya yang dapat kita lakukan adalah mencari kembali data excel untuk memastikan bahwa ke empat orang ini apakah seharusnya miskin atau sejahtera. Kesalahan seperti ini sangat sering sekali terjadi di republik ini, sehingga banyak orang yang seharusnya tidak memperoleh bantuan BLT malah dapat bantuan, sedang orang yang seharusnya tidak memperoleh bantuan BLT malah dapat karena datanya diubah.

Mudah-mudahan cara sederhana ini dapat menjelaskan bahwa kita sebetulnya cukup mudah untuk mendeteksi kesalahan klasifikasi.

Contoh Sederhana Klasifikasi Stunting

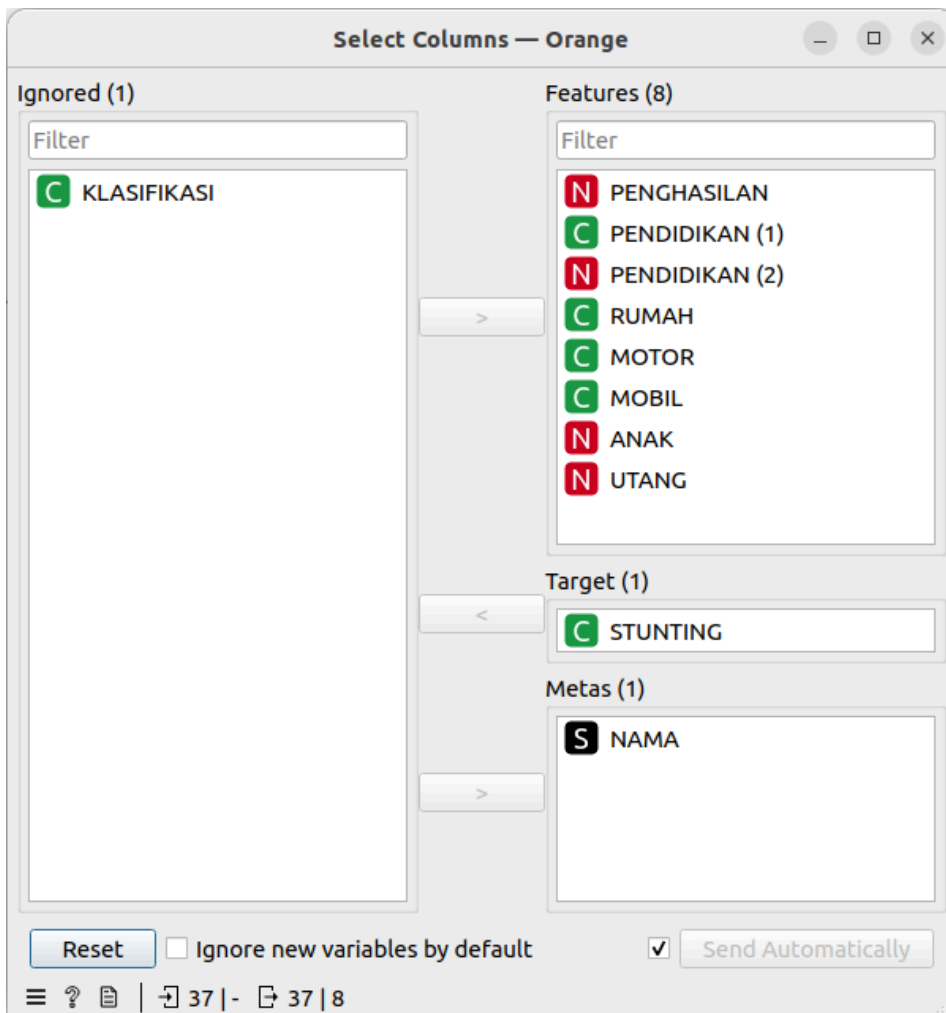
Kebetulan data kemiskinan yang kita gunakan ada kolom Stunting yang dapat kita jadikan target. Pada kesempatan ini kita akan mencoba data kemiskinan yang sama, akan tetapi kita akan menganalisa klasifikasi Stunting.

Pertama-tama yang perlu diubah widget Select Columns untuk menset agar,

- Klasifikasi masuk ke Ignored
- Stunting masuk ke target.

Penampakan widget Select Columns setelah kita ubah terlihat pada gambar.

Hasil analisa yang diperoleh dapat kita lihat pada gambar selanjutnya. Jujur apa adanya, data yang ada tampaknya sulit untuk digunakan mengklasifikasi Stunting.



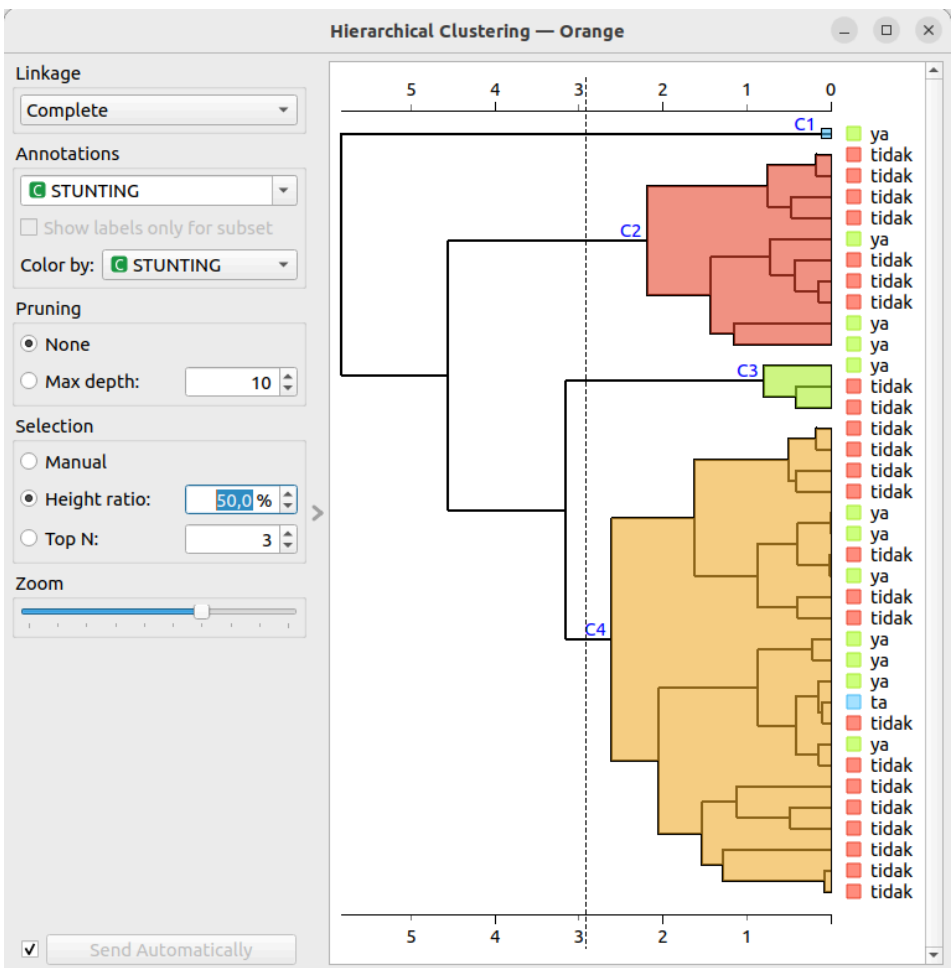
Gambar 8.10. widget Select Columns dengan Stunting sebagai target.

Bahkan untuk visualisasi Cluster kita perlu mengubah parameter visualisasi, yaitu,

- Annotations - Stunting.
- Color By - Stunting.
- Selection - Height Ratio di set misalnya ke 50%

Terlihat gambar yang dihasilkan, ada empat (4) kelompok, akan tetapi masing-masing kelompok tidak konsisten dalam membagi Stunting karena

di masing-masing kelompok ada yang Stunting ada yang tidak kena Stunting.



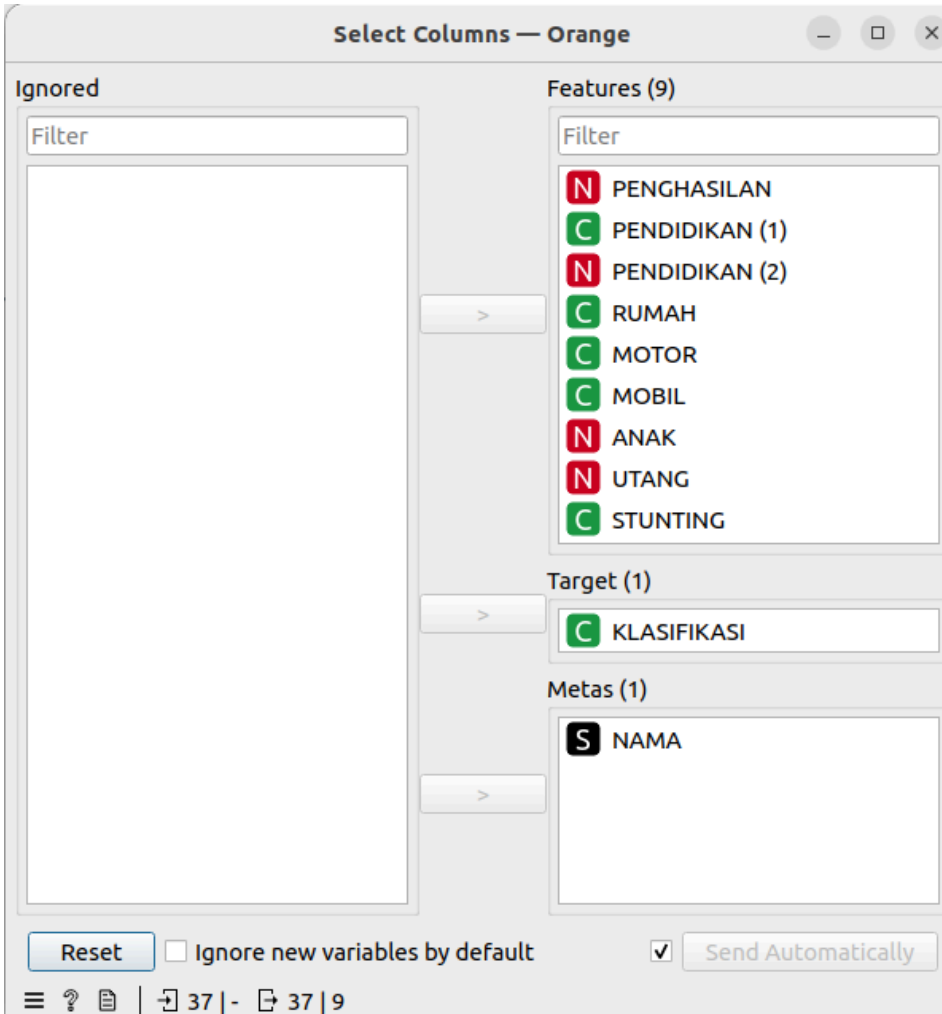
Gambar 8.11. Hasil Pengelompokan Stunting.

Pada dasarnya untuk data yang kita miliki kita agak sulit untuk melakukan pengelompokan Stunting. Kita mungkin perlu usaha lain untuk melihat perilaku pengelompokan Stunting yang lebih masuk akal.

Kondisi-kondisi seperti seringkali akan kita hadapi pada saat melakukan pengolahan data, dimana data tidak dapat dengan mudah kita kelompokkan sesuai dengan keinginan kita.

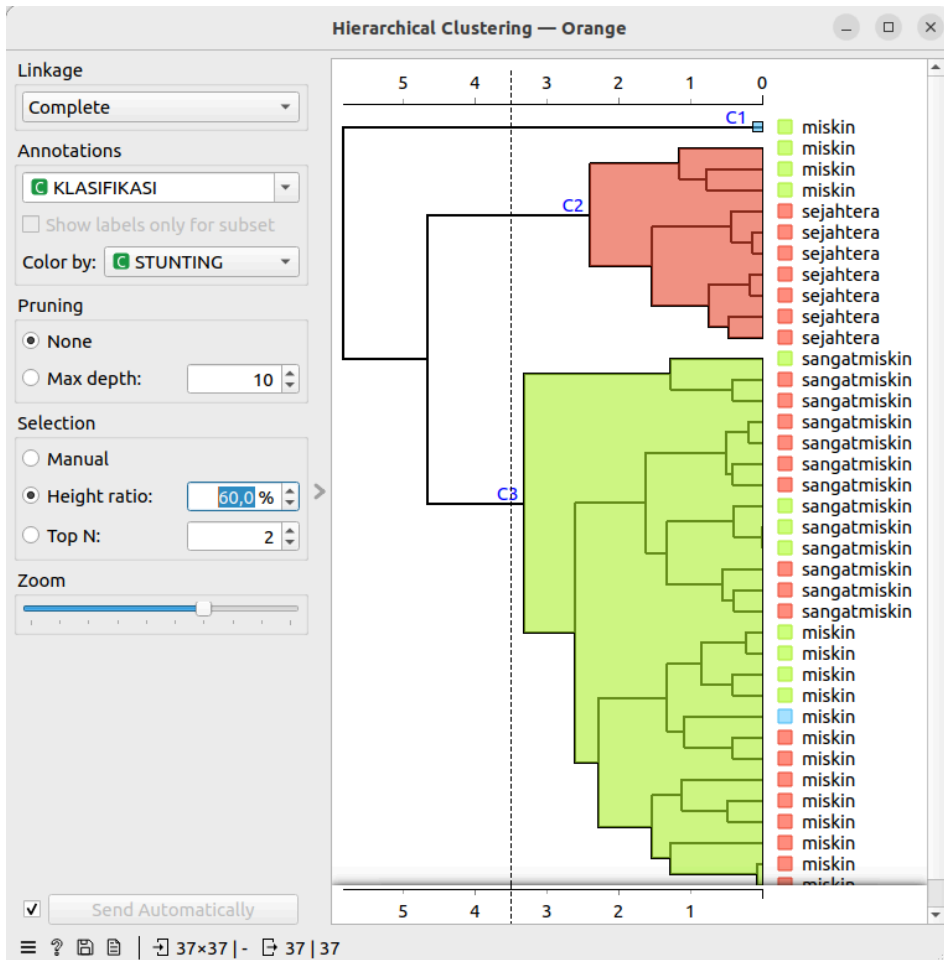
Contoh Klasifikasi Stunting dengan Korelasi Kemiskinan

Selanjutnya kita akan melihat korelasi Stunting dengan data kemiskinan yang kita miliki. Untuk itu kita perlu berexperiment mengubah sedikit isi widget Select Columns sebagai berikut,



Gambar 8.12. setup widget Select Columns dengan menset Stunting sebagai salah satu fitur.

Pada widget Select Columns di atas, kita mencoba memasukan Stunting menjadi salah satu fitur. Dengan mengubah struktur kolom kita bisa melihat hubungan antara Stunting dengan kemiskinan. Visualisasi hasil clustering-nya diperlihatkan pada gambar berikut.



Gambar 8.13. Visualisasi Clustering dengan Color By Stunting.

Gambar diatas diperoleh dengan men-set,

- Annotation menggunakan kolom Klasifikasi.
- Color by menggunakan kolom Stunting.
- Selection Height Ratio bisa di experimen, disini kita menggunakan 60% untuk memperoleh cluster yang sedikit.

Hasilnya terlihat dengan jelas bahwa ada dua (2) kelompok besar yang terkait Stunting,

- Masyarakat sejahtera, tidak ditemukan kasus Stunting.

- Stunting biasanya ada pada kelompok masyarakat miskin dan sangat miskin.
- Tapi memang tidak semua masyarakat miskin dan sangat miskin yang ada Stunting, sebagian tidak ada Stunting. Untuk lebih presisi mengklasifikasi situasi Stunting ini kita membutuhkan data yang lebih dalam dan detail lagi. Data apa adanya, tampaknya sulit untuk memperoleh gambaran klasifikasi penyebab Stunting.

Contoh Prediksi Program Pengentasan Kemiskinan

Selanjutnya berdasarkan pengalaman analisa klasifikasi kemiskinan di sebelumnya, mungkinkah kita memperoleh model dari data kemiskinan yang kita miliki? Jawaban singkatnya, ya bisa.

Dengan memiliki model, kita dapat dengan mudah melakukan prediksi. Contoh skenario,

- Jika pendidikan kita subsidi agar tingkat pendidikan rakyat misalnya satu tingkat di atas sebelumnya, berapa persen kemiskinan rakyat dapat kita turunkan?
- Jika kita dapat mendorong lapangan kerja, sehingga penghasilan dapat kita naikan 25%, berapa persen kemiskinan dapat kita turunkan?
- Parameter apa yang dapat menurunkan Stunting?

Evaluasi fitur yang dominan mengubah target

Sebelum kita terlalu jauh melakukan operasi perdukunan / forecasting, ada baiknya kita mengevaluasi dari sekian banyak fitur yang ada di data training, mana yang paling dominan untuk melakukan perubahan pada target kolom yang kita inginkan.

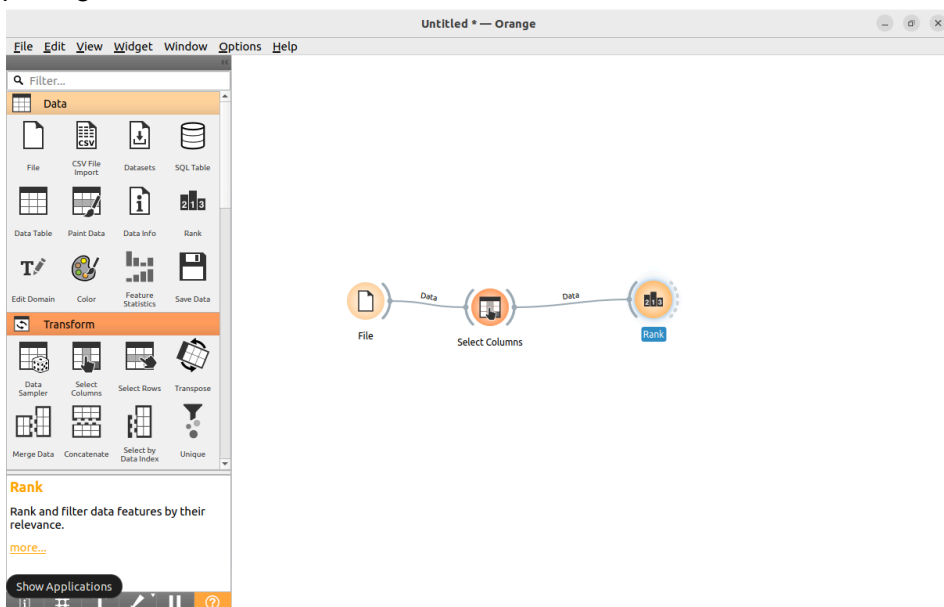
Data training kemiskinan yang kita miliki mempunyai banyak kolom fitur, mulai dari Penghasilan, Pendidikan, Lama Pendidikan, Rumah, Motor, Mobil, Anak, Utang, Nama, Klasifikasi / Status dan Stunting. Masing-masing kolom dapat kita jadikan target untuk analisa, dan tergantung pada fitur lainnya.

Beberapa target kolom yang akan menarik untuk dianalisa ketergantungannya terhadap kolom lain adalah,

- Klasifikasi / Status
- Stunting
- Penghasilan

Tentunya kolom lainnya juga terbuka untuk dijadikan target. Mari kita mulai dengan tiga (3) target di atas, dan mengidentifikasi dua (2) atau tiga (3) kolom mana yang paling berpengaruh pada target kolom tersebut.

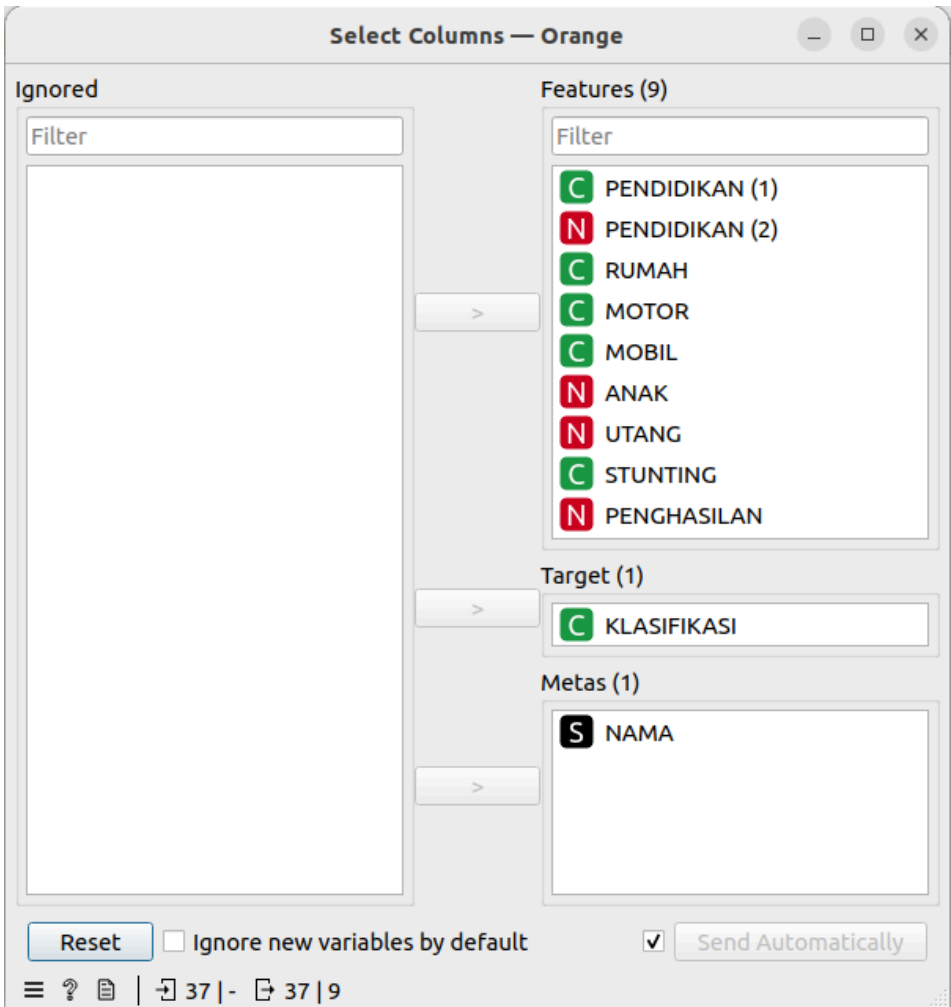
Misalnya kita akan menganalisa kolom klasifikasi / status sebagai target analisa, dan mencari tiga (3) kolom fitur yang paling berpengaruh terhadap klasifikasi. Kita dapat mengembangkan program seperti tampak pada gambar.



‘Gambar 8.14. Analisa efek fitur terhadap target.

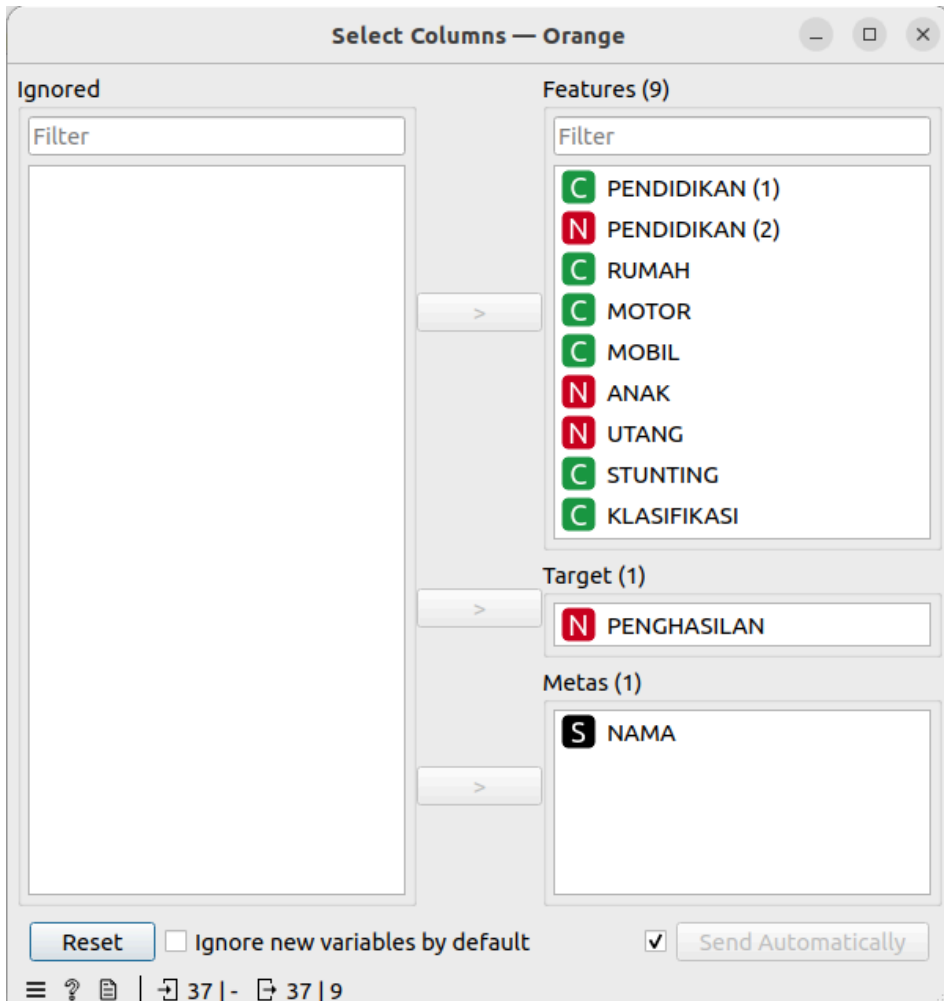
Widget kunci yang akan memberikan informasi tentang fitur mana yang paling berpengaruh pada target adalah widget Rank. Ada beberapa metoda pilihan metoda untuk menganalisa ranking fitur yang paling berpengaruh pada target, kita dapat menganalisa dengan semua metoda yang ada dan memperoleh gambaran tentang fitur yang paling berpengaruh.

Untuk melakukan analisa kita perlu menset terlebih dulu, kolom mana yang menjadi target sisanya bisa menjadi fitur. Pada gambar berikut, kita menggunakan kolom klasifikasi sebagai target.



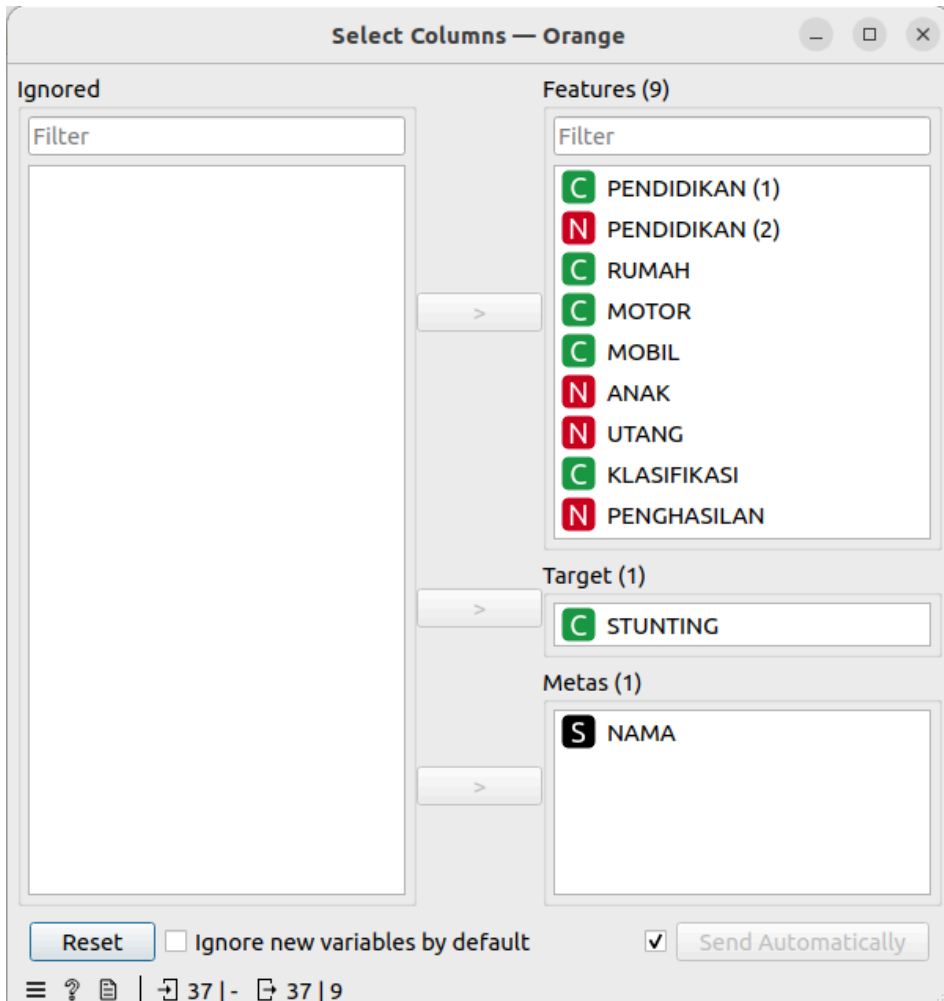
Gambar 8.15. Kolom Klasifikasi sebagai Target.

Untuk analisa lainnya, kita dapat menggunakan kolom penghasilan sebagai target. Dengan demikian kita dapat menganalisa fitur apa saja yang akan berpengaruh pada penghasilan. Penampakan widget Select Columns saat memilih kolom penghasilan sebagai target tampak pada gambar.



Gambar 8.16. Kolom Penghasilan sebagai target.

Untuk analisa lainnya, kita dapat menggunakan kolom Stunting sebagai target. Dengan demikian kita dapat menganalisa fitur apa saja yang akan berpengaruh pada Stunting. Penampakan widget Select Columns saat memilih kolom Stunting sebagai target tampak pada gambar.



Gambar 8.17. Kolom Stunting sebagai target.

Untuk melakukan rank efek berbagai fitur terhadap target ada beberapa metoda. Berikut adalah perbedaan Rank fitur berdasarkan hasil analisa Information Gain, Information Gain Ratio, Gini Decrease, ANOVA, X2, ReliefF, FCBF

- Information Gain (IG) mengukur pengurangan entropi yang dihasilkan oleh pembagian data berdasarkan atribut tertentu. Atribut dengan nilai IG tinggi menunjukkan kemampuannya untuk memisahkan data ke dalam kelas yang berbeda.

- Information Gain Ratio (IGR) merupakan modifikasi dari IG yang mempertimbangkan bias terhadap atribut multi-nilai. IGR membagi IG dengan intrinsic information dari atribut tersebut. Atribut dengan nilai IGR tinggi menunjukkan kemampuannya untuk memisahkan data dengan lebih efisien.
- Gini Decrease (GD) mengukur penurunan Gini Index setelah pembagian data berdasarkan atribut tertentu. Gini Index merupakan metrik untuk mengukur ketidakmurnian data. Atribut dengan nilai GD tinggi menunjukkan kemampuannya untuk meningkatkan kemurnian data.
- ANOVA (Analysis of Variance) merupakan metode statistik untuk menguji apakah terdapat perbedaan yang signifikan antara rata-rata kelompok data yang berbeda. ANOVA digunakan untuk mengukur kekuatan hubungan antara atribut numerik dan target variabel.
- X2 (Chi-Square) merupakan statistik uji yang digunakan untuk menguji apakah terdapat hubungan antara dua variabel kategorikal. X2 tinggi menunjukkan kemungkinan besar terdapat hubungan antara kedua variabel tersebut.
- ReliefF merupakan algoritma yang mengukur relevansi atribut dengan menghitung perbedaan jarak terdekat antara instance dengan kelas yang sama dan berbeda. Atribut dengan nilai ReliefF tinggi menunjukkan kemampuannya untuk memisahkan instance dengan kelas yang berbeda.
- FCBF (Fast Correlation Based Feature Selection) merupakan algoritma yang memilih atribut berdasarkan korelasi mutlak dengan target variabel. FCBF memilih atribut yang memiliki korelasi tinggi dengan target variabel dan korelasi rendah antara atribut yang dipilih.

Untuk melihat ringkasan perbedaan antara berbagai metoda, tabel berikut akan membantu melihat hal tersebut.

Tabel 8.1 Perbandingan Metoda untuk Men-Ranking

Metode	Deskripsi	Keunggulan	Kekurangan
Information Gain	Mengukur pengurangan entropi	Sederhana, mudah dipahami	Bias terhadap atribut multi-nilai
Information Gain Ratio	Mengurangi bias terhadap atribut multi-nilai	Lebih adil untuk atribut multi-nilai	Lebih kompleks daripada IG
Gini Decrease	Mengukur penurunan Gini Index	Efektif untuk data yang tidak seimbang	Sensitif terhadap outliers
ANOVA	Menguji perbedaan rata-rata kelompok data	Kuat untuk data numerik	Lemah untuk data kategorikal
X ²	Menguji hubungan antara dua variabel kategorikal	Sederhana, mudah dipahami	Tidak dapat mengukur kekuatan hubungan
ReliefF	Mengukur relevansi atribut dengan perbedaan jarak terdekat	Efektif untuk data dengan banyak atribut	Sensitif terhadap outliers
FCBF	Memilih atribut berdasarkan korelasi mutlak	Sederhana, mudah dipahami	Tidak dapat mengukur kekuatan hubungan

Metode yang terbaik untuk memilih fitur tergantung pada jenis data, tujuan analisis, dan preferensi pengguna.

Catatan:

- Penting untuk diingat bahwa rank fitur hanyalah indikator dan tidak boleh menjadi satu-satunya faktor yang dipertimbangkan dalam memilih fitur.
- Kombinasi dari beberapa metode rank fitur dapat digunakan untuk mendapatkan hasil yang lebih optimal.

Berikut adalah tiga besar kolom yang berpengaruh pada target Klasifikasi. Gambar di bawah diperoleh dari widget Rank yang diurutkan untuk Information Gain (IG).

	#	In...In	Gain ratio	Gini	ANOVA	χ^2	ReliefF	FCBF
1 C PENDIDIKAN (1)	5	1.347	0.609	0.568	NA	32.772	0.613	2.475
2 N PENGHASILAN		1.330	0.665	0.540	NA	26.748	0.358	2.923
3 N PENDIDIKAN (2)		1.102	0.553	0.455	NA	24.572	0.280	0.000
4 C MOBIL	2	0.740	0.843	0.262	NA	22.895	0.534	1.530
5 C MOTOR	2	0.644	0.673	0.210	NA	15.681	0.367	0.000
6 C RUMAH	2	0.644	0.673	0.210	NA	15.681	0.367	0.000
7 N ANAK		0.338	0.171	0.109	NA	5.449	0.070	0.000
8 N UTANG		0.206	0.314	0.087	NA	7.841	0.019	0.000
9 C STUNTING	2	0.004	0.004	0.001	NA	0.117	-0.064	0.000

Gambar 8.18. Rank - Information Gain

Terlihat bahwa urutannya adalah,

- Tingkat Pendidikan
- Penghasilan
- Lama Pendidikan

Untuk rank menggunakan metoda Gini Decrease hasilnya ditampilkan pada gambar selanjutnya. Rank diperoleh juga sama dengan Information Gain (IG), yaitu,

- Tingkat Pendidikan
- Penghasilan
- Lama Pendidikan

Rank — Orange

Scoring Methods

☒ Information Gain

☒ Information Gain Ratio

☒ Gini Decrease

☒ ANOVA

Select Attributes

☐ None

☐ All

☐ Manual

☒ Best ranked: 3

☒ Send Automatically

	#	Info. gain	Gain ratio	Gini	ANOVA	χ^2	ReliefF	FCBF	
1	PENDIDIKAN (1)	5	1.347	0.609	0.568	NA	32.772	0.613	2.475
2	PENGHASILAN		1.330	0.665	0.540	NA	26.748	0.358	2.923
3	PENDIDIKAN (2)		1.102	0.553	0.455	NA	24.572	0.280	0.000
4	MOBIL	2	0.740	0.843	0.262	NA	22.895	0.534	1.530
5	RUMAH	2	0.644	0.673	0.210	NA	15.681	0.367	0.000
6	MOTOR	2	0.644	0.673	0.210	NA	15.681	0.367	0.000
7	ANAK		0.338	0.171	0.109	NA	5.449	0.070	0.000
8	UTANG		0.206	0.314	0.087	NA	7.841	0.019	0.000
9	STUNTING	2	0.004	0.004	0.001	NA	0.117	-0.064	0.000

Gambar 8.19. Rank Gini Decrease

Selanjutnya hasil analisis rank menggunakan ReliefF, juga menghasilkan hasil yang sama dengan sebelumnya yaitu,

- Tingkat Pendidikan
- Penghasilan
- Lama Pendidikan

Rank — Orange

Scoring Methods

☒ Information Gain

☒ Information Gain Ratio

☒ Gini Decrease

☒ ANOVA

Select Attributes

☐ None

☐ All

☐ Manual

☒ Best ranked: 3

☒ Send Automatically

	#	Info. gain	Gain ratio	Gini	ANOVA	χ^2	ReliefF	FCBF	
1	PENDIDIKAN (1)	5	1.347	0.609	0.568	NA	32.772	0.613	2.475
2	PENGHASILAN		1.330	0.665	0.540	NA	26.748	0.358	2.923
3	PENDIDIKAN (2)		1.102	0.553	0.455	NA	24.572	0.280	0.000
4	MOBIL	2	0.740	0.843	0.262	NA	22.895	0.534	1.530
5	RUMAH	2	0.644	0.673	0.210	NA	15.681	0.367	0.000
6	MOTOR	2	0.644	0.673	0.210	NA	15.681	0.367	0.000
7	UTANG		0.206	0.314	0.087	NA	7.841	0.019	0.000
8	ANAK		0.338	0.171	0.109	NA	5.449	0.070	0.000
9	STUNTING	2	0.004	0.004	0.001	NA	0.117	-0.064	0.000

Gambar 8.20. Rank ReliefF

Menarik untuk dicatat bahwa hasil rank untuk semua metoda tidak semua sama. Hasil rank FCBF menunjukan urutan yang berbeda, yaitu,

- Penghasilan
- Tingkat Pendidikan
- Mobil

	#	Info. gain	Gain ratio	Gini	ANOVA	x²	ReliefF	FCBF
1 N PENGHASILAN		1.330	0.665	0.540	NA	26.748	0.358	2.923
2 C PENDIDIKAN (1)	5	1.347	0.609	0.568	NA	32.772	0.613	2.475
3 C MOBIL	2	0.740	0.843	0.262	NA	22.895	0.534	1.530
4 N PENDIDIKAN (2)		1.102	0.553	0.455	NA	24.572	0.280	0.000
5 C MOTOR	2	0.644	0.673	0.210	NA	15.681	0.367	0.000
6 C RUMAH	2	0.644	0.673	0.210	NA	15.681	0.367	0.000
7 N ANAK		0.338	0.171	0.109	NA	5.449	0.070	0.000
8 N UTANG		0.206	0.314	0.087	NA	7.841	0.019	0.000
9 C STUNTING	2	0.004	0.004	0.001	NA	0.117	-0.064	0.000

Gambar 8.21. Rank FCBF

Disini menjadi menarik untuk dicatat bahwa fitur-fitur yang perlu diperhitungkan untuk menganalisa kemiskinan minimal adalah

- Tingkat Pendidikan
- Penghasilan
- Lama Pendidikan
- Mobil

Fitur lainnya dapat kita ujicoba untuk disertakan dalam perhitungan atau bisa kita ignored. Tentunya ini semua harus diuji coba untuk melihat mana yang paling menghasilkan perkiraan target yang paling baik.

Dengan cara yang sama, kita dapat menganalisis rank fitur yang paling berpengaruh untuk target Stunting. Rangkuman hasil rank adalah sebagai berikut,

Rank hasil metode Information Gain (IG)

- Tingkat Pendidikan
- Utang
- Lama Pendidikan

Rank hasil metoda Gini Decrease

- Tingkat Pendidikan
- Utang
- Lama Pendidikan

Rank hasil Metoda ReliefF

- Tingkat Pendidikan
- Klasifikasi
- Rumah

Rank hasil metoda FCBF

- Utang
- Tingkat Pendidikan
- Lama Pendidikan

Jika kita kumpulkan semua fitur yang akan mempengaruhi Stunting maka, fitur tersebut adalah,

- Tingkat Pendidikan
- Utang
- Lama Pendidikan
- Klasifikasi
- Rumah

Dengan cara yang sama, kita dapat menganalisis rank fitur yang paling berpengaruh untuk target Penghasilan. Karakter target Penghasilan berbeda dengan lainnya, Penghasilan merupakan variabel kontinu. Metoda untuk melakukan rank ada dua, yaitu,

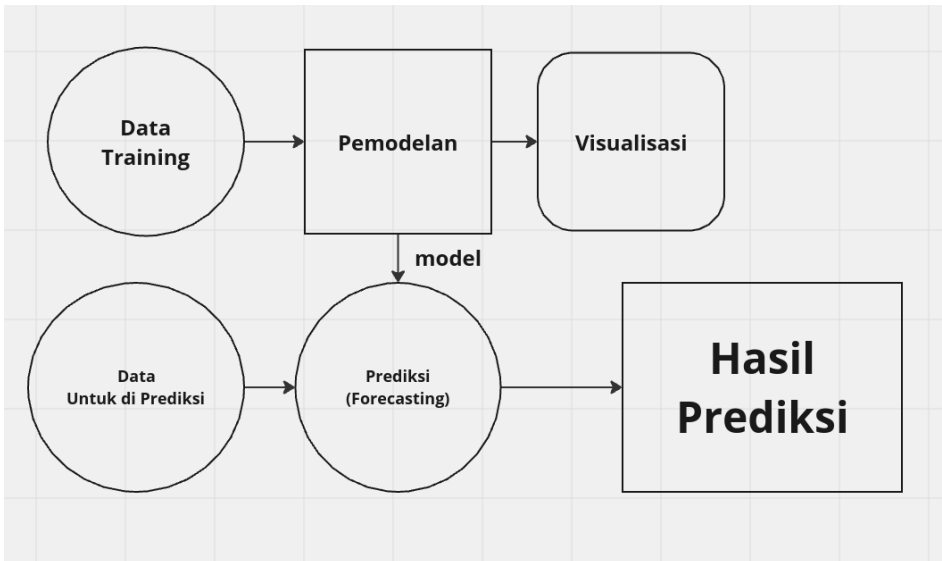
Univariate Regression
RReliefF

Untuk data yang kita miliki, tampaknya hanya bisa menggunakan RReliefF. Rangkuman hasil rank untuk 5 fitur utama yang bisa mempengaruhi target Penghasilan adalah sebagai berikut,

- Tingkat Pendidikan
- Klasifikasi
- Rumah
- Mobil
- Motor

Logika berfikir untuk prediksi

Untuk mengingatkan kita semua, logika berfikir untuk melakukan perdukunan / prediksi (forecasting) apa yang akan dilakukan ke depan dapat dijelaskan dalam gambar berikut.

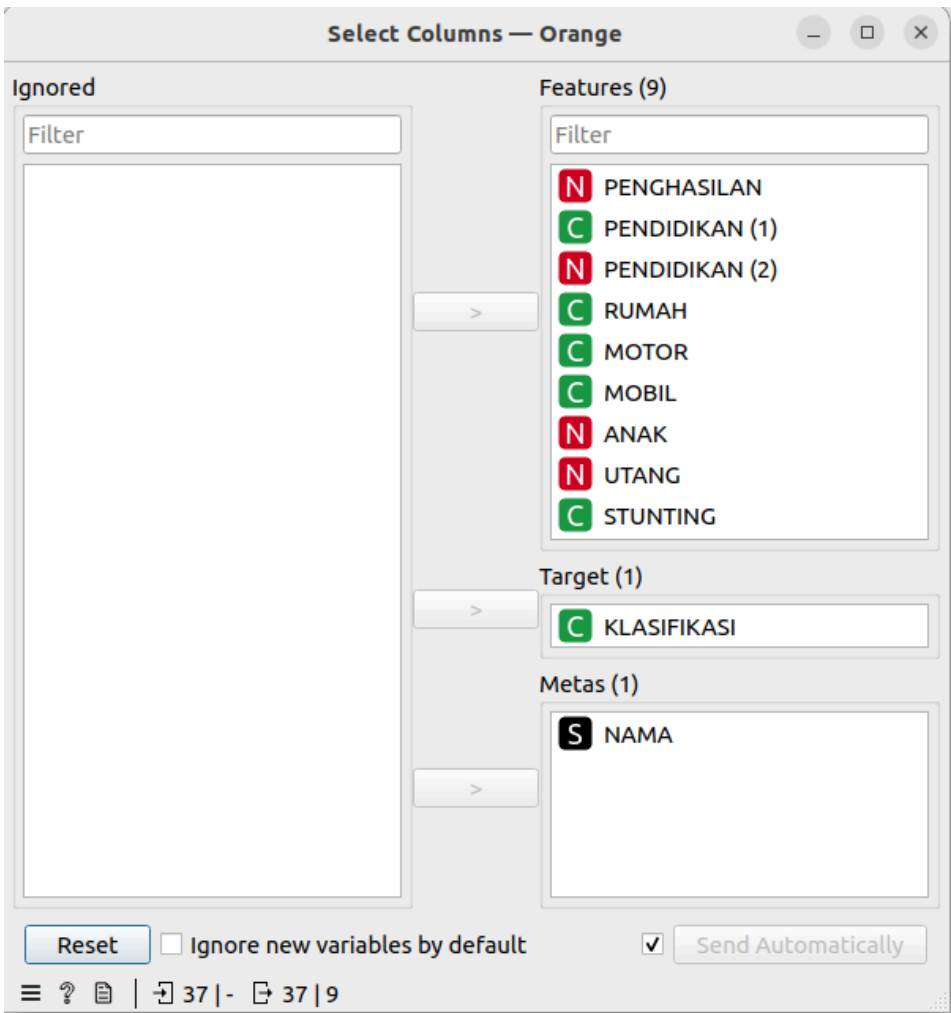


Gambar 8.22. Logika Berfikir Forecasting (Prediksi)

Logika berfikir untuk bisa melakukan prediksi sebetulnya sederhana saja.

- Kita harus memiliki model dari data. Model tersebut diperoleh melalui proses pemodelan dari training data yang kita miliki sebelumnya. Semakin banyak data training yang kita miliki biasanya semakin presisi model yang akan kita miliki.
- Prediksi dilakukan dengan memasukan data untuk diprediksi ke model yang kita miliki. Ada baiknya mengingat penjelasan tentang persamaan linear $Ax+B=Y$, disini pada dasarnya model adalah nilai A & B , x adalah data yang akan di prediksi, sementara Y adalah hasil prediksi. Untuk memudahkan proses prediksi, biasanya data untuk prediksi merupakan data history yang diubah sedikit di salah satu kolom X -nya untuk memprediksi efek perubahan data di satu kolom X terhadap Y .
- Visualisasi biasanya dilakukan setelah dilakukan pemodelan untuk membantu memberikan gambaran bagi kita manusia akan model yang dihasilkan.

Untuk bisa mengimplementasikan ide proses prediksi di atas mari kita membuat model dari data training miskin.xls yang kita gunakan sebelumnya. Langkah pertama kita harus menggunakan widget Select Columns untuk memilih fitur dan target dari training data seperti tampak pada gambar.

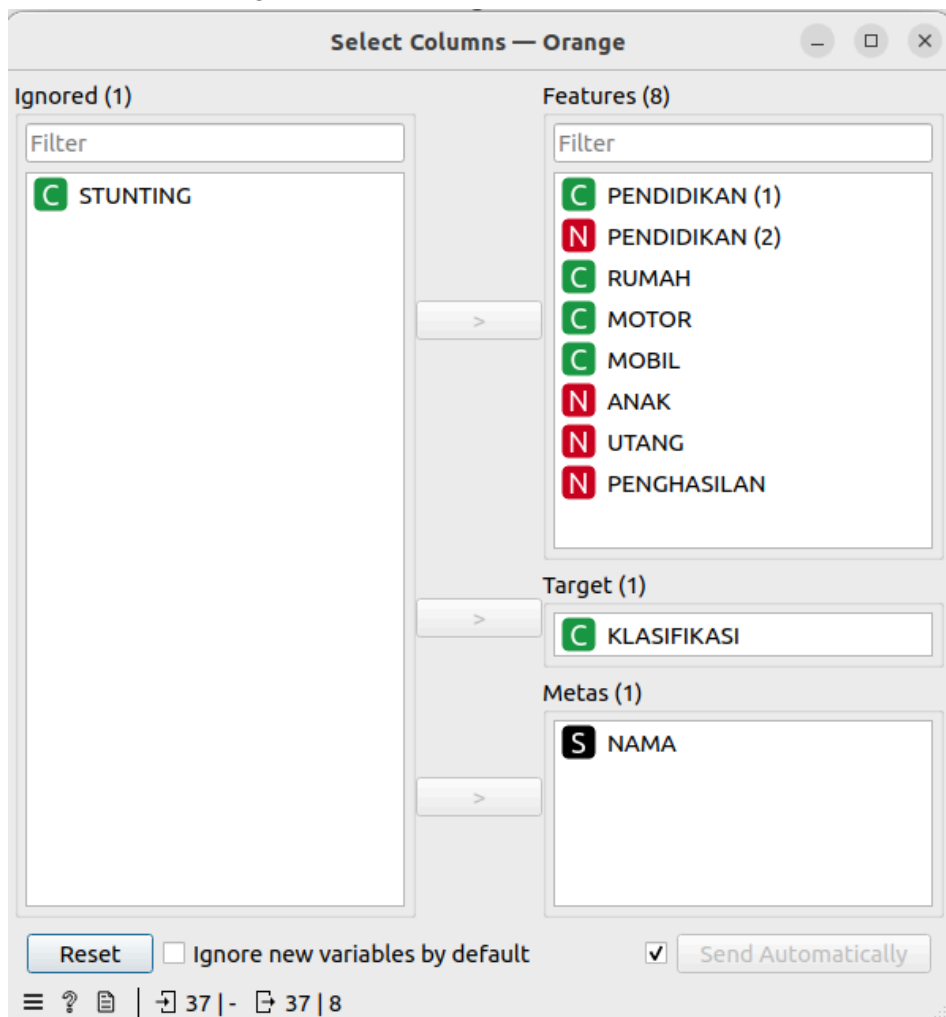


Gambar 8.23. kondisi widget Select Columns

Untuk kebutuhan melakukan prediksi program KemenkopUKM untuk pengentasan kemiskinan, kita akan menggunakan sebagian besar kolom sebagai fitur dan hanya kolom klasifikasi sebagai target seperti tampak pada gambar widget Select Columns. Jika kita ingin lebih fokus dapat

menggunakan hanya fitur-fitur yang mempunyai pengaruh besar pada target variable seperti yang dideteksi oleh widget Rank.

Contoh paling sederhana dalam pemilihan kolom yang di analisa ditampilkan pada gambar berikut,



Gambar 8.24. Meng-Ignore Stunting.

Salah satu yang paling sederhana adalah mengeluarkan Stunting sebagai parameter dari bagian data yang akan dianalisa. Salah satu hal yang paling membuat pusing saat proses analisa adalah menentukan kolom yang bisa menjadi fitur, kolom mana yang menjadi target. Beberapa tip sederhana yang dapat diambil adalah,

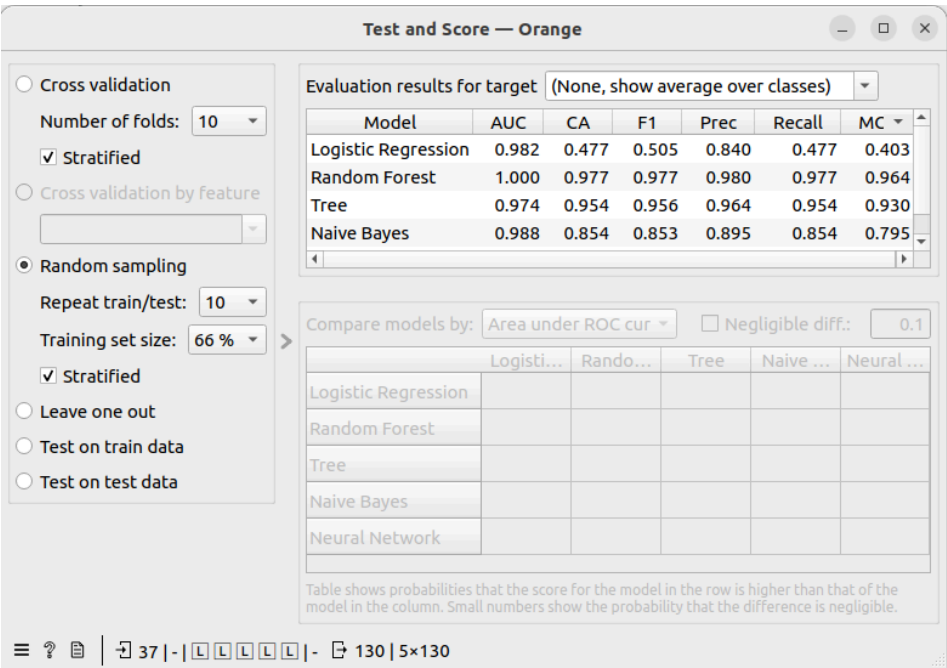
- Untuk meyakinkan bahwa kita akan menggunakan algoritma untuk memperoleh model dari data dengan presisi tertinggi, ada baiknya kita menggunakan widget Test and Score dan menyambungkan keluaran dari widget Select Columns dan masuk ke input widget Test and Score. Kemudian masukan keluaran beberapa algoritma modeling untuk widget Test and Score. Seperti tampak pada gambar.



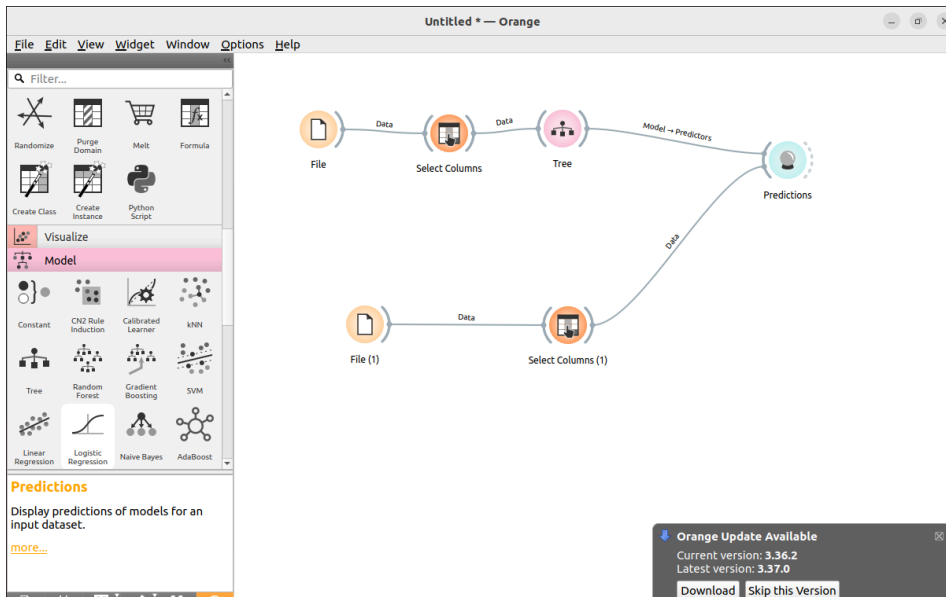
- Random Forest
- Tree
- Naive Bayes
- Logistic Regression

Gambaran detail hasil pengukuran widget Test and Score diperlihatkan pada gambar berikutnya.

Menarik untuk dicatat bahwa untuk memperoleh hasil yang lebih reliable ada baiknya kita menggunakan Random Sampling, dan Repeat train/test sebanyak 10 kali, dengan training size 66%. Tentunya disini kita dapat berexperimen untuk dapat menentukan algoritma mana yang secara konsisten akan memberikan hasil yang presisi dengan berbagai kombinasi yang ada. Untuk data yang penulis buat, algoritma yang konsisten memberikan hasil yang presisi adalah Random Forest. Untuk ujicoba kita akan mencoba algoritma Tree supaya hasilnya lebih seru.



Gambar 8.25. Hasil Test and Score.



Gambar 8.26. Program untuk melakukan prediksi.

Gambar di atas adalah program untuk melakukan prediksi. widget Prediksi memperoleh dua input, input yang pertama adalah model yang dibuat dari training data. Input ke dua adalah data test atau data untuk diprediksi target variabel-nya. Input ke dua diperoleh File excel yang di load melalui widget File. Screenshot file excel dengan nama miskin prediksi adalah seperti tampak pada gambar.

	A	B	C	D	E	F	G	H	I	J	K
1	PENGHASILAN	PENDIDIKAN	PENDIDIKAN	RUMAH	MOTOR	MOBIL	ANAK	NAMA	UTANG	KLASIFIKASI	STUNTING
2	2000000	SMA	12	0	0	0	0	4 Amir	0	miskin	ya
3	1500000	SMA	12	1	0	0	1	5 Badu	0	miskin	tidak
4	200000	SD	1	1	1	1	1	4 cad	0	miskin	tidak
5	150000000	SD	6	0	0	0	0	10 ded	1000000	sejahtera	ya
6	2000000	SMA	12	1	1	1	1	2 Rani	0	sejahtera	tidak
7	5000000	SMA	12	1	1	1	1	3 Sasa	0	sejahtera	tidak
8	4600000	SMA	12	1	1	1	1	3 Tati	0	sejahtera	tidak
9											

Gambar 8.27. Input Excel dari data untuk diprediksi.

Sebetulnya isinya agak sembarang karena semanya penulis saja mengisi kolom-kolom yang ada, setelah di sambungkan ke widget Predictions maka akan tampak seperti pada gambar.

Predictions — Orange												
Show probabilities for		Classes in data	☒ Show classification errors									Restore Original Order
Tree	error	KLASIFIKASI	NAMA	PENGHASILAN	ENDIDIKAN (1)	ENDIDIKAN (2)	RUMAH	MOTOR	MOBIL	ANAK		
1 0.00 : 1.00 → sejahtera	1.000	miskin	Amir	10000000.0	SMA	12.0	0.0	0.0	0.0	4.0	0.0	
2 1.00 : 0.00 → miskin	0.000	miskin	Badu	1500000.0	SMA	12.0	2.0	0.0	2.0	5.0	0.0	
3 0.00 : 0.00 → sangatmiskin	1.000	miskin	cad	200000.0	SD	1.0	2.0	1.0	2.0	4.0	0.0	
4 0.00 : 1.00 → sejahtera	0.000	sejahtera	ded	15000000.0	SD	6.0	0.0	0.0	0.0	10.0	10.0	
5 1.00 : 0.00 → miskin	1.000	sejahtera	Rani	2000000.0	SMA	12.0	1.0	1.0	1.0	2.0	0.0	
6 0.00 : 1.00 → sejahtera	0.000	sejahtera	Sasa	5000000.0	SMA	12.0	1.0	1.0	1.0	3.0	0.0	
0.00 : 1.00 → sejahtera	0.000	sejahtera	Tati	4600000.0	SMA	12.0	1.0	1.0	1.0	3.0	0.0	

Gambar 8.28. Hasil Prediksi Random Forest.

Pada gambar di terlihat hasil prediksi target oleh model yang dibuat oleh Tree. Di kolom paling kiri adalah hasil prediksi menggunakan Tree, kolom selanjutnya adalah error / kesalahan antara prediksi Tree dibandingkan dengan perkiraan target yang kita buat, dan kolom sebelahnya yang berwarna gelap adalah nilai target yang kita inginkan.

Analisis prediksi dapat ditingkatkan dengan menerapkan skenario yang lebih rumit. Misalnya, kita bisa diminta untuk merancang program terbaik guna mengurangi angka kemiskinan atau stunting di suatu daerah. Pertanyaan semacam ini seringkali muncul dari para pemimpin daerah yang ingin melihat solusi konkret untuk masalah sosial yang kompleks.

Tujuan para pemimpin umumnya cukup jelas. Mereka ingin mengetahui langkah-langkah paling efektif untuk mencapai tujuan tertentu. Dalam konteks pengentasan kemiskinan, misalnya, kita memiliki banyak faktor yang bisa dimanipulasi. Pertanyaan mendasarnya adalah: variabel mana yang paling berpengaruh dalam mengurangi kemiskinan? Salah satu contoh intervensi yang mungkin adalah memberikan tambahan dua tahun pendidikan bagi seluruh masyarakat. Analisis prediksi dapat membantu kita mengukur dampak potensial dari intervensi seperti ini terhadap tingkat kemiskinan.

Jika kita memilih sebuah fitur atau variable maka berapa perkiraan biaya atau anggaran yang perlu dikeluarkan? berapa lama kira-kira intervensi tersebut perlu dilakukan? Dengan kondisi data yang ada maka kita punya beberapa pilihan program untuk kebutuhan contoh pada kesempatan ini, kita akan mencoba beberapa hal seperti,

- Menaikan gaji / pendapatan bulanan setiap orang sebanyak 10%. Apa efeknya terhadap tingkat kemiskinan masyarakat.

- Menaikan gaji / pendapatan bulanan setiap orang sebanyak 10 kali lipat untuk setiap orang. Apa efek terhadap tingkat kemiskinan masyarakat.
- Menambah waktu pendidikan setiap orang dengan dua tahun. Apa efek terhadap tingkat kemiskinan masyarakat.

Percobaan Menaikan Pendapatan 10%

Untuk percobaan pertama kita akan mencoba jika penghasilan / pendapatan setiap orang berhasil kita naikan sebanyak 10% berapa banyak orang yang akan naik klasifikasinya, misalnya, dari sangatmiskin menjadi miskin, atau dari miskin ke sejahtera.

Predictions — Orange											
Show probabilities for		Classes in data	✓ Show classification errors		Restore Original Order						
Tree		error	KLASIFIKASI	NAMA	PENGHASILAN	ENDIDIKAN (1	ENDIDIKAN (2	RUMAH	MOTOR	MOBIL	ANAK
1	1.00 : 0.00 : 0.00 → miskin	0.000	miskin	Amir	2200000.0	SD	6.0	0.0	0.0	0.0	4.0
2	1.00 : 0.00 : 0.00 → miskin	0.000	miskin	Badu	1650000.0	SD	6.0	0.0	0.0	0.0	5.0
3	1.00 : 0.00 : 0.00 → miskin	0.000	miskin	Candra	1870000.0	SD	6.0	0.0	0.0	0.0	3.0
4	1.00 : 0.00 : 0.00 → miskin	1.000	sangatmiskin	Dedi	1100000.0	SD	7.0	0.0	0.0	0.0	5.0
5	1.00 : 0.00 : 0.00 → miskin	0.000	miskin	Eddy	2090000.0	SD	7.0	0.0	0.0	0.0	3.0
6	1.00 : 0.00 : 0.00 → miskin	0.000	miskin	Fahri	2178000.0	SD	6.0	0.0	0.0	0.0	3.0
7	1.00 : 0.00 : 0.00 → miskin	0.000	miskin	Gagah	1980000.0	SD	7.0	0.0	0.0	0.0	4.0
8	1.00 : 0.00 : 0.00 → miskin	0.000	miskin	Hadi	2090000.0	SMP	9.0	0.0	0.0	0.0	2.0
9	1.00 : 0.00 : 0.00 → miskin	0.000	miskin	Ira	1870000.0	SMP	9.0	0.0	0.0	0.0	2.0
10	1.00 : 0.00 : 0.00 → miskin	0.000	miskin	Jaja	2090000.0	SMP	10.0	0.0	1.0	0.0	4.0
11	1.00 : 0.00 : 0.00 → miskin	0.000	miskin	Kaka	1870000.0	SMP	10.0	1.0	1.0	0.0	3.0
12	0.00 : 0.00 : 1.00 → sejahtera	1.000	miskin	Lala	3300000.0	SMP	9.0	1.0	1.0	0.0	3.0
13	1.00 : 0.00 : 0.00 → miskin	0.000	miskin	Mami	2750000.0	SMP	9.0	1.0	0.0	0.0	4.0
14	0.00 : 0.00 : 1.00 → sejahtera	0.000	sejahtera	Nani	4400000.0	SMA	11.0	1.0	1.0	1.0	0.0
15	0.00 : 0.00 : 1.00 → sejahtera	0.000	sejahtera	Opa	4620000.0	SMA	11.0	1.0	1.0	1.0	1.0
16	0.00 : 0.00 : 1.00 → sejahtera	0.000	sejahtera	Qra	4279000.0	SMA	11.0	1.0	1.0	1.0	2.0
17	0.00 : 0.00 : 1.00 → sejahtera	0.000	miskin	Rani	2200000.0	SMA	12.0	1.0	1.0	1.0	2.0
18	0.00 : 0.00 : 1.00 → sejahtera	0.000	sejahtera	Sasa	5500000.0	SMA	12.0	1.0	1.0	1.0	3.0
19	0.00 : 0.00 : 1.00 → sejahtera	0.000	sejahtera	Tati	5060000.0	SMA	12.0	1.0	1.0	1.0	3.0
20	0.00 : 0.00 : 1.00 → sejahtera	0.000	sejahtera	Udi	5170000.0	SMA	13.0	1.0	1.0	1.0	2.0
21	0.00 : 0.00 : 1.00 → sejahtera	0.000	sejahtera	Vava	6600000.0	UNIV	14.0	1.0	1.0	1.0	2.0
22	0.00 : 0.00 : 1.00 → sejahtera	0.000	sejahtera	Wawa	7150000.0	UNIV	14.0	1.0	1.0	1.0	2.0
23	0.00 : 0.00 : 1.00 → sejahtera	0.000	sejahtera	Xaxa	6050000.0	UNIV	18.0	1.0	1.0	1.0	1.0
24	0.00 : 0.00 : 1.00 → sejahtera	0.000	sejahtera	Yaya	6710000.0	UNIV	19.0	1.0	1.0	1.0	2.0
25	0.00 : 1.00 : 0.00 → sangatmiskin	0.000	sangatmiskin	Mamir	220000.0	NO	0.0	0.0	0.0	0.0	4.0
26	0.00 : 1.00 : 0.00 → sangatmiskin	0.000	sangatmiskin	Mbamir	330000.0	NO	1.0	0.0	0.0	0.0	6.0
27	0.00 : 1.00 : 0.00 → sangatmiskin	0.000	sangatmiskin	Mcamir	440000.0	NO	2.0	0.0	0.0	0.0	7.0
28	0.00 : 1.00 : 0.00 → sangatmiskin	0.000	sangatmiskin	Midamir	275000.0	NO	3.0	0.0	0.0	0.0	4.0

Gambar 8.29. Hasil Prediksi dengan penghasilan naik 10%

Untuk melakukan prediksi ini, kita perlu mengcopykan file training data dan nama lain, misalnya miskin-program1.xls dimana isi kolom penghasilan kita naikan nilainya 10% dari nilai yang ada di file miskin.xls. File miskin-program1.xls kita masukan ke prediksi dan kita lihat hasil prediksi Tree. Yang perlu kita lihat di hasil prediksi random forest adalah berapa klasifikasi yang berubah status, persentase klasifikasi yang berubah ini yang perlu kita laporkan ke pimpinan, atau walikota, atau gubernur, atau menteri.

Contoh hasil prediksi untuk miskin-program1.xls yang dibuat penulis diperlihatkan pada gambar. Gambar di atas adalah screenshot hasil prediksi saat penghasilan kita naikan 10%. Dengan mengamati klasifikasi yang terjadi, ternyata ada beberapa yang klasifikasi yang berubah. Tepatnya,

- Satu perubahan dari sangat miskin menjadi miskin
- Satu perubahan dari miskin menjadi sejahtera.

Kebetulan, ada total 37 data, maka dengan menaikkan penghasilan 10% saja ternyata menyebabkan ada 2 perubahan artinya ada 2/37 atau 5% perubahan.

Percobaan Menaikan Pendapatan 10 kali lipat.

Untuk percobaan selanjutnya kita akan mencoba jika penghasilan / pendapatan setiap orang berhasil kita naikan sebanyak 10 kali lipat dari penghasilan sebelumnya. Disini kita coba melihat berapa banyak orang yang akan naik klasifikasinya, misalnya, dari sangatmiskin menjadi miskin, atau dari miskin ke sejahtera.

Untuk melakukan prediksi ini, kita perlu mengcopykan file training data dan nama lain, misalnya miskin-program2.xls dimana isi kolom penghasilan kita naikan nilainya 10 kali dari nilai yang ada di file miskin.xls. File miskin-program2.xls kita masukan ke prediksi dan kita lihat hasil prediksi random-forest-nya. Yang perlu kita lihat di hasil prediksi random forest adalah berapa klasifikasi yang berubah status, persentase klasifikasi yang berubah ini yang perlu kita laporkan ke pimpinan, atau walikota, atau gubernur, atau menteri.

Contoh hasil prediksi untuk miskin-program2.xls yang dibuat penulis diperlihatkan pada gambar berikut.

Predictions — Orange

Show probabilities for: ☐ Classes in data ☒ Show classification errors

	Tree	error	KLASIFIKASI	NAMA	PENGHASILAN	ENDIDIKAN (1)	ENDIDIKAN (2)	RUMAH	MOTOR	MOBIL	ANAK
1	0.00 : 0.00 : 1.00 → sejahtera	1.000	miskin	Amir	20000000.0	SD	6.0	0.0	0.0	0.0	4.0
2	0.00 : 0.00 : 1.00 → sejahtera	1.000	miskin	Badu	15000000.0	SD	6.0	0.0	0.0	0.0	5.0
3	0.00 : 0.00 : 1.00 → sejahtera	1.000	miskin	Candra	17000000.0	SD	6.0	0.0	0.0	0.0	3.0
4	0.00 : 0.00 : 1.00 → sejahtera	1.000	sangatmiskin	Dedi	10000000.0	SD	7.0	0.0	0.0	0.0	5.0
5	0.00 : 0.00 : 1.00 → sejahtera	1.000	miskin	Eddy	19000000.0	SD	7.0	0.0	0.0	0.0	3.0
6	0.00 : 0.00 : 1.00 → sejahtera	1.000	miskin	Fahri	19800000.0	SD	6.0	0.0	0.0	0.0	3.0
7	0.00 : 0.00 : 1.00 → sejahtera	1.000	miskin	Gagah	18000000.0	SD	7.0	0.0	0.0	0.0	4.0
8	0.00 : 0.00 : 1.00 → sejahtera	1.000	miskin	Hadi	19000000.0	SMP	9.0	0.0	0.0	0.0	2.0
9	0.00 : 0.00 : 1.00 → sejahtera	1.000	miskin	Ira	17000000.0	SMP	9.0	0.0	0.0	0.0	2.0
10	0.00 : 0.00 : 1.00 → sejahtera	1.000	miskin	Jaja	19000000.0	SMP	10.0	0.0	1.0	0.0	4.0
11	0.00 : 0.00 : 1.00 → sejahtera	1.000	miskin	Kaka	17000000.0	SMP	10.0	1.0	1.0	0.0	3.0
12	0.00 : 0.00 : 1.00 → sejahtera	1.000	miskin	Lala	30000000.0	SMP	9.0	1.0	1.0	0.0	3.0
13	0.00 : 0.00 : 1.00 → sejahtera	1.000	miskin	Mami	25000000.0	SMP	9.0	1.0	0.0	0.0	4.0
14	0.00 : 0.00 : 1.00 → sejahtera	0.000	sejahtera	Nani	40000000.0	SMA	11.0	1.0	1.0	1.0	0.0
15	0.00 : 0.00 : 1.00 → sejahtera	0.000	sejahtera	Opa	42000000.0	SMA	11.0	1.0	1.0	1.0	1.0
16	0.00 : 0.00 : 1.00 → sejahtera	0.000	sejahtera	Qra	38900000.0	SMA	11.0	1.0	1.0	1.0	2.0
17	0.00 : 0.00 : 1.00 → sejahtera	1.000	miskin	Rani	20000000.0	SMA	12.0	1.0	1.0	1.0	2.0
18	0.00 : 0.00 : 1.00 → sejahtera	0.000	sejahtera	Sasa	50000000.0	SMA	12.0	1.0	1.0	1.0	3.0
19	0.00 : 0.00 : 1.00 → sejahtera	0.000	sejahtera	Tati	46000000.0	SMA	12.0	1.0	1.0	1.0	3.0
20	0.00 : 0.00 : 1.00 → sejahtera	0.000	sejahtera	Udi	47000000.0	SMA	13.0	1.0	1.0	1.0	2.0
21	0.00 : 0.00 : 1.00 → sejahtera	0.000	sejahtera	Vava	60000000.0	UNIV	14.0	1.0	1.0	1.0	2.0
22	0.00 : 0.00 : 1.00 → sejahtera	0.000	sejahtera	Wawa	65000000.0	UNIV	14.0	1.0	1.0	1.0	2.0
23	0.00 : 0.00 : 1.00 → sejahtera	0.000	sejahtera	Xaxa	55000000.0	UNIV	18.0	1.0	1.0	1.0	1.0
24	0.00 : 0.00 : 1.00 → sejahtera	0.000	sejahtera	Yaya	61000000.0	UNIV	19.0	1.0	1.0	1.0	2.0
25	1.00 : 0.00 : 0.00 → miskin	1.000	sangatmiskin	Mamir	2000000.0	NO	0.0	0.0	0.0	0.0	4.0
26	1.00 : 0.00 : 0.00 → miskin	1.000	sangatmiskin	Mbamir	3000000.0	NO	1.0	0.0	0.0	0.0	6.0
27	0.00 : 0.00 : 1.00 → sejahtera	1.000	sangatmiskin	Mc Amir	4000000.0	NO	2.0	0.0	0.0	0.0	7.0
28	1.00 : 0.00 : 0.00 → miskin	1.000	sangatmiskin	Mdamir	2500000.0	NO	3.0	0.0	0.0	0.0	4.0

Show performance scores

37 | 37 | 1x37

Gambar 8.30. Hasil prediksi dengan penghasilan di naikan 10x lipat.

Gambar di atas adalah screenshot hasil prediksi saat penghasilan kita naikan 10%. Dengan mengamati klasifikasi yang terjadi, ternyata ada banyak sekali yang klasifikasi yang berubah, yaitu,

- Dari sangat miskin ke miskin 8
- Dari sangat miskin ke sejahtera 4
- Dari miskin ke sejahtera 13
- Tidak berubah 12

Terlihat dengan jelas, terjadi perubahan besar-besaran terutama dari kategori sangat miskin dan miskin menjadi kategori yang lebih baik. Kira-kira $\frac{2}{3}$ dari populasi berubah menjadi lebih baik.

Memang ada $\frac{1}{3}$ dari populasi yang tidak berubah, kategori sejahtera merupakan kategori yang paling banyak tidak berubah. Masuk akal karena memang kebetulan kategori sejahtera adalah yang tertinggi, tidak ada kategori super crazy rich. Yang jelas, dengan menaikkan penghasilan kita akan melihat perubahan dari kategori masyarakat.

Menaikan penghasilan bukan hal yang mudah, hal ini akan melibatkan banyak hal terutama memudahkan dunia usaha untuk berusaha dan

memutarkan uangnya. Ketersediaan infrastruktur, seperti, listrik, air, internet, jalan dll, yang akan sangat memudahkan dunia usaha. Jadi menaikkan penghasilan tidak semudah mengganti angka saja.

Percobaan menambahkan lama pendidikan 6 tahun.

Untuk percobaan selanjutnya kita akan mencoba jika lama pendidikan setiap orang kita tambah 6 tahun dari lama pendidikan sebelumnya. Jadi semua yang tadinya hanya lulus SD menjadi lulus SMA. Disini kita coba melihat berapa banyak orang yang akan naik klasifikasinya, misalnya, dari sangatmiskin menjadi miskin, atau dari miskin ke sejahtera.

Untuk melakukan prediksi ini, kita perlu mengcopykan file training data dan nama lain, misalnya miskin-program3.xls dimana isi kolom penghasilan kita tambahkan ke lama pendidikan 6 tahun, juga tingkat pendidikan dari SD menjadi SMA, dari SMP menjadi UNIV. File miskin-program3.xls kita masukan ke prediksi dan kita lihat hasil prediksi Tree. Yang perlu kita lihat di hasil prediksi Tree adalah berapa klasifikasi yang berubah status, persentase klasifikasi yang berubah ini yang perlu kita laporkan ke pimpinan, atau walikota, atau gubernur, atau menteri.

Contoh hasil prediksi untuk miskin-program3.xls yang dibuat penulis diperlihatkan pada gambar berikut. Dengan sangat mengejutkan tidak ada efek sama sekali terhadap perubahan kategori. Dengan kata lain, untuk data yang dibuat penulis menaikkan tingkat pendidikan mengubah status seseorang.

Predictions — Orange

Show probabilities for: ☐ Classes in data ☒ Show classification errors

	Tree	error	KLASIFIKASI	NAMA	PENGHASILAN	ENDIDIKAN (1)	ENDIDIKAN (2)	RUMAH	MOTOR	MOBIL	ANAK
1	1.00 : 0.00 : 0.00 → miskin	0.000	miskin	Amir	2000000.0	SMA	12.0	0.0	0.0	0.0	4.0
2	1.00 : 0.00 : 0.00 → miskin	0.000	miskin	Badu	1500000.0	SMA	12.0	0.0	0.0	0.0	5.0
3	1.00 : 0.00 : 0.00 → miskin	0.000	miskin	Candra	1700000.0	SMA	12.0	0.0	0.0	0.0	3.0
4	0.00 : 1.00 : 0.00 → sangatmiskin	0.000	sangatmiskin	Dedi	1000000.0	UNIV	13.0	0.0	0.0	0.0	5.0
5	1.00 : 0.00 : 0.00 → miskin	0.000	miskin	Eddy	1900000.0	UNIV	13.0	0.0	0.0	0.0	3.0
6	1.00 : 0.00 : 0.00 → miskin	0.000	miskin	Fahri	1980000.0	SMA	12.0	0.0	0.0	0.0	3.0
7	1.00 : 0.00 : 0.00 → miskin	0.000	miskin	Gagah	1800000.0	UNIV	13.0	0.0	0.0	0.0	4.0
8	1.00 : 0.00 : 0.00 → miskin	0.000	miskin	Hadi	1900000.0	UNIV	15.0	0.0	0.0	0.0	2.0
9	1.00 : 0.00 : 0.00 → miskin	0.000	miskin	Ira	1700000.0	UNIV	15.0	0.0	0.0	0.0	2.0
10	1.00 : 0.00 : 0.00 → miskin	0.000	miskin	Jaja	1900000.0	UNIV	16.0	0.0	1.0	0.0	4.0
11	1.00 : 0.00 : 0.00 → miskin	0.000	miskin	Kaka	1700000.0	UNIV	16.0	1.0	1.0	0.0	3.0
12	1.00 : 0.00 : 0.00 → miskin	0.000	miskin	Lala	3000000.0	UNIV	15.0	1.0	1.0	0.0	3.0
13	1.00 : 0.00 : 0.00 → miskin	0.000	miskin	Mami	2500000.0	UNIV	15.0	1.0	0.0	0.0	4.0
14	0.00 : 0.00 : 1.00 → sejahtera	0.000	sejahtera	Nani	4000000.0	UNIV	17.0	1.0	1.0	1.0	0.0
15	0.00 : 0.00 : 1.00 → sejahtera	0.000	sejahtera	Opa	4200000.0	UNIV	17.0	1.0	1.0	1.0	1.0
16	0.00 : 0.00 : 1.00 → sejahtera	0.000	sejahtera	Qra	3890000.0	UNIV	17.0	1.0	1.0	1.0	2.0
17	1.00 : 0.00 : 0.00 → miskin	0.000	miskin	Rani	2000000.0	UNIV	18.0	1.0	1.0	1.0	2.0
18	0.00 : 0.00 : 1.00 → sejahtera	0.000	sejahtera	Sasa	5000000.0	UNIV	18.0	1.0	1.0	1.0	3.0
19	0.00 : 0.00 : 1.00 → sejahtera	0.000	sejahtera	Tati	4600000.0	UNIV	18.0	1.0	1.0	1.0	3.0
20	0.00 : 0.00 : 1.00 → sejahtera	0.000	sejahtera	Udi	4700000.0	UNIV	19.0	1.0	1.0	1.0	2.0
21	0.00 : 0.00 : 1.00 → sejahtera	0.000	sejahtera	Vava	6000000.0	UNIV	20.0	1.0	1.0	1.0	2.0
22	0.00 : 0.00 : 1.00 → sejahtera	0.000	sejahtera	Wawa	6500000.0	UNIV	20.0	1.0	1.0	1.0	2.0
23	0.00 : 0.00 : 1.00 → sejahtera	0.000	sejahtera	Xaxa	5500000.0	UNIV	24.0	1.0	1.0	1.0	1.0
24	0.00 : 0.00 : 1.00 → sejahtera	0.000	sejahtera	Yaya	6100000.0	UNIV	25.0	1.0	1.0	1.0	2.0
25	0.00 : 1.00 : 0.00 → sangatmiskin	0.000	sangatmiskin	Mamir	200000.0	SD	6.0	0.0	0.0	0.0	4.0
26	0.00 : 1.00 : 0.00 → sangatmiskin	0.000	sangatmiskin	Mbamir	300000.0	SMP	7.0	0.0	0.0	0.0	6.0
27	0.00 : 1.00 : 0.00 → sangatmiskin	0.000	sangatmiskin	Mc Amir	400000.0	SMP	8.0	0.0	0.0	0.0	7.0
28	0.00 : 1.00 : 0.00 → sangatmiskin	0.000	sangatmiskin	Mdamir	250000.0	SMP	9.0	0.0	0.0	0.0	4.0

☐ Show performance scores

37 | 37 | 1x37

Gambar 8.31. Prediksi Tree dengan menaikkan 6 tahun lama pendidikan.

Kita tentunya tahu, meningkatkan pendidikan kemungkinan besar akan meningkatkan status seseorang di kemudian hari, yang terdeteksi dari berbagai aset yang dimiliki oleh orang tersebut. Apa adanya, perubahan lama pendidikan tanpa perubahan aset akan menyebabkan algoritma salah mendeteksi status seseorang.

Percobaan menambahkan 1 rumah.

Untuk percobaan selanjutnya kita akan mencoba jika jumlah rumah yang dimiliki seseorang di tambah satu (1) rumah dari jumlah rumah sebelumnya. Disini kita coba melihat berapa banyak orang yang akan naik klasifikasinya, misalnya, dari sangatmiskin menjadi miskin, atau dari miskin ke sejahtera.

Untuk melakukan prediksi ini, kita perlu mengcopykan file training data dan nama lain, misalnya miskin-program4.xls dimana isi kolom rumah kita tambahkan satu (1) rumah. File miskin-program4.xls kita masukan ke prediksi dan kita lihat hasil prediksi Tree. Yang perlu kita lihat di hasil prediksi Tree adalah berapa klasifikasi yang berubah status, persentase klasifikasi yang berubah ini yang perlu kita laporkan ke pimpinan, atau walikota, atau gubernur, atau menteri.

Contoh hasil prediksi untuk miskin-program4.xls yang dibuat penulis diperlihatkan pada gambar berikut.

	Tree	error	KLASIFIKASI	NAMA	PENGHASILAN	ENDIDIKAN (1)	ENDIDIKAN (2)	RUMAH	MOTOR	MOBIL	ANAK
1	1.00 : 0.00 : 0.00 → miskin	0.000	miskin	Amir	2000000.0	SD	6.0	1.0	0.0	0.0	4.0
2	1.00 : 0.00 : 0.00 → miskin	0.000	miskin	Badu	1500000.0	SD	6.0	1.0	0.0	0.0	5.0
3	1.00 : 0.00 : 0.00 → miskin	0.000	miskin	Candra	1700000.0	SD	6.0	1.0	0.0	0.0	3.0
4	0.00 : 1.00 : 0.00 → sangatmiskin	0.000	sangatmiskin	Dedi	1000000.0	SD	7.0	1.0	0.0	0.0	5.0
5	1.00 : 0.00 : 0.00 → miskin	0.000	miskin	Eddy	1900000.0	SD	7.0	1.0	0.0	0.0	3.0
6	1.00 : 0.00 : 0.00 → miskin	0.000	miskin	Fahri	1980000.0	SD	6.0	1.0	0.0	0.0	3.0
7	1.00 : 0.00 : 0.00 → miskin	0.000	miskin	Gagah	1800000.0	SD	7.0	1.0	0.0	0.0	4.0
8	1.00 : 0.00 : 0.00 → miskin	0.000	miskin	Hadli	1900000.0	SMP	9.0	1.0	0.0	0.0	2.0
9	1.00 : 0.00 : 0.00 → miskin	0.000	miskin	Ira	1700000.0	SMP	9.0	1.0	0.0	0.0	2.0
10	1.00 : 0.00 : 0.00 → miskin	0.000	miskin	Jaja	1900000.0	SMP	10.0	1.0	1.0	0.0	4.0
11	1.00 : 0.00 : 0.00 → miskin	0.000	miskin	Kaka	1700000.0	SMP	10.0	2.0	1.0	0.0	3.0
12	1.00 : 0.00 : 0.00 → miskin	0.000	miskin	Lala	3000000.0	SMP	9.0	2.0	1.0	0.0	3.0
13	1.00 : 0.00 : 0.00 → miskin	0.000	miskin	Mami	2500000.0	SMP	9.0	2.0	0.0	0.0	4.0
14	0.00 : 0.00 : 1.00 → sejahtera	0.000	sejahtera	Nani	4000000.0	SMA	11.0	2.0	1.0	1.0	0.0
15	0.00 : 0.00 : 1.00 → sejahtera	0.000	sejahtera	Opa	4200000.0	SMA	11.0	2.0	1.0	1.0	1.0
16	0.00 : 0.00 : 1.00 → sejahtera	0.000	sejahtera	Qra	3890000.0	SMA	11.0	2.0	1.0	1.0	2.0
17	1.00 : 0.00 : 0.00 → miskin	0.000	miskin	Rani	2000000.0	SMA	12.0	2.0	1.0	1.0	2.0
18	0.00 : 0.00 : 1.00 → sejahtera	0.000	sejahtera	Sasa	5000000.0	SMA	12.0	2.0	1.0	1.0	3.0
19	0.00 : 0.00 : 1.00 → sejahtera	0.000	sejahtera	Tati	4600000.0	SMA	12.0	2.0	1.0	1.0	3.0
20	0.00 : 0.00 : 1.00 → sejahtera	0.000	sejahtera	Udi	4700000.0	SMA	13.0	2.0	1.0	1.0	2.0
21	0.00 : 0.00 : 1.00 → sejahtera	0.000	sejahtera	Vava	6000000.0	UNIV	14.0	2.0	1.0	1.0	2.0
22	0.00 : 0.00 : 1.00 → sejahtera	0.000	sejahtera	Wawa	6500000.0	UNIV	14.0	2.0	1.0	1.0	2.0
23	0.00 : 0.00 : 1.00 → sejahtera	0.000	sejahtera	Xaxa	5500000.0	UNIV	18.0	2.0	1.0	1.0	1.0
24	0.00 : 0.00 : 1.00 → sejahtera	0.000	sejahtera	Yaya	6100000.0	UNIV	19.0	2.0	1.0	1.0	2.0
25	0.00 : 1.00 : 0.00 → sangatmiskin	0.000	sangatmiskin	Mamir	200000.0	NO	0.0	1.0	0.0	0.0	4.0
26	0.00 : 1.00 : 0.00 → sangatmiskin	0.000	sangatmiskin	Mbamir	300000.0	NO	1.0	1.0	0.0	0.0	6.0
27	0.00 : 1.00 : 0.00 → sangatmiskin	0.000	sangatmiskin	Mc Amir	400000.0	NO	2.0	1.0	0.0	0.0	7.0
28	0.00 : 1.00 : 0.00 → sangatmiskin	0.000	sangatmiskin	Mdamir	250000.0	NO	3.0	1.0	0.0	0.0	4.0

Gambar 8.32. Prediksi Tree dengan menambah 1 rumah setiap orang.

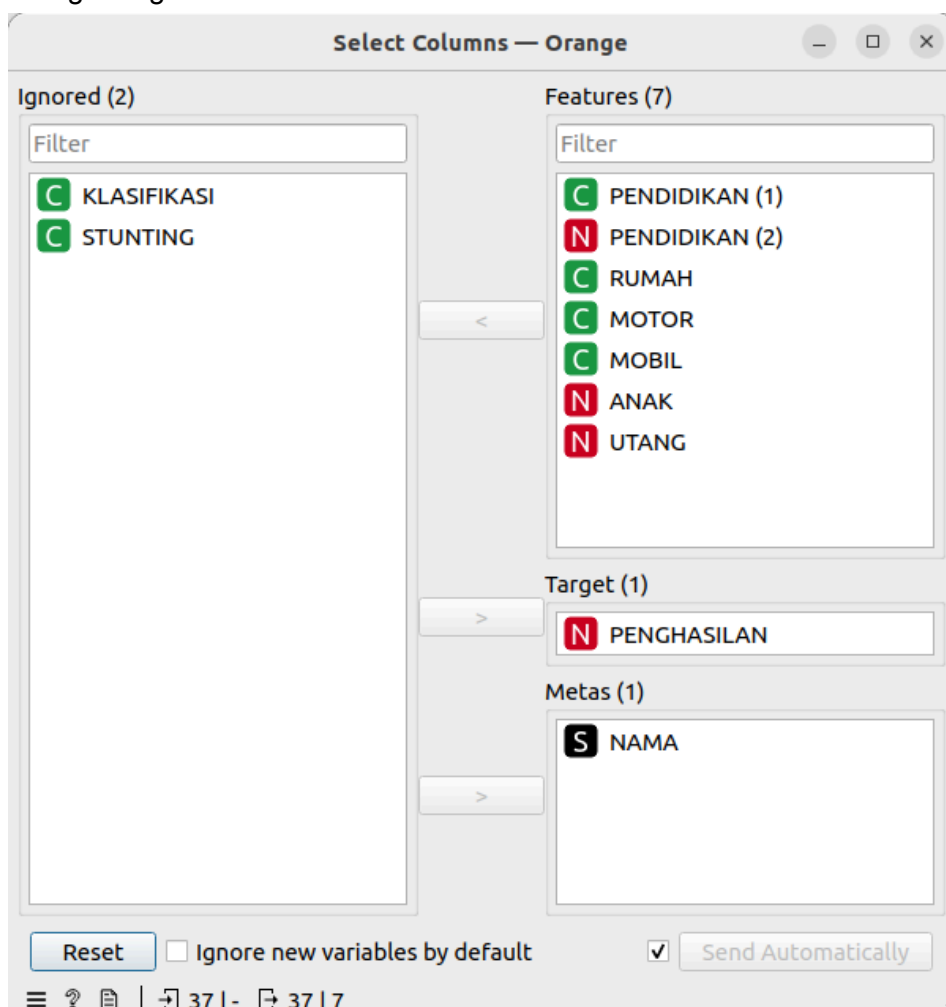
Hasil prediksi penambahan 1 rumah untuk setiap orang ternyata cukup mengejutkan. Tidak ada perubahan status dengan penambahan satu (1) rumah untuk setiap orang.

Dari percobaan yang kita lakukan terlihat jelas bahwa upaya meningkatkan penghasilan seseorang atau keluarga akan sangat terasa efeknya untuk meningkatkan status seseorang dari sangat miskin menjadi miskin atau bahkan menjadi sejahtera. Untuk itu selanjutnya kita akan melihat cara menaikkan penghasilan yang paling feasible.

Contoh Prediksi Program Peningkatan Penghasilan

Dalam analisa sebelumnya terlihat dengan jelas bahwa klasifikasi seseorang atau keluarga akan berubah jika penghasilan dapat di naikan. Tentunya ada banyak sekali variabel yang dapat berpengaruh dalam menaikkan penghasilan orang di suatu daerah. Di dunia nyata variable ini sangat banyak, untuk keperluan pembelajaran kita akan membatasi pembatasan pada data training yang kita miliki. Kita pada dasarnya

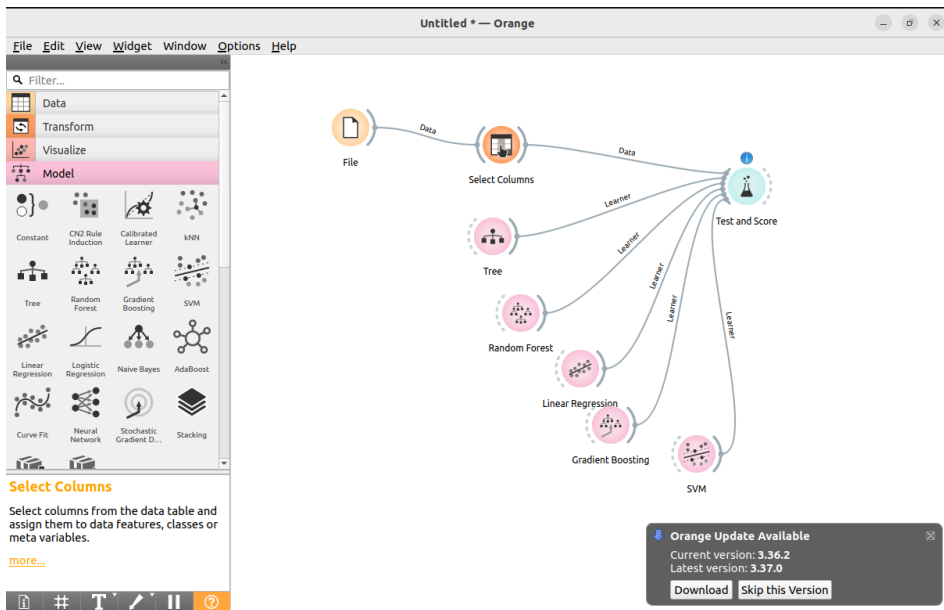
mengubah melalui widget Select Columns agar target kolom adalah penghasilan sedangkan kolom yang lain kita jadikan sebagai fitur untuk meramal target. Untuk mengingatkan kita semua, penampakan setup widget Select Columns seperti pada gambar dengan kolom Penghasilan sebagai target.



Gambar 8.33. widget Select Columns dengan target kolom Penghasilan.

Tampak pada gambar bahwa pada kesempatan ini, hampir semua kolom yang ada digunakan sebagai fitur kecuali kolom nama hanya dijadikan metadata saja, dan kolom Klasifikasi dan Stunting di ignore tidak digunakan sama sekali dalam proses perhitungan dan modeling.

Sebelum kita mulai melakukan pemodelan sebaiknya kita tes dulu algoritma mana yang akan memberikan error yang paling kecil. Untuk mengetes kita menggunakan program di bawah ini.



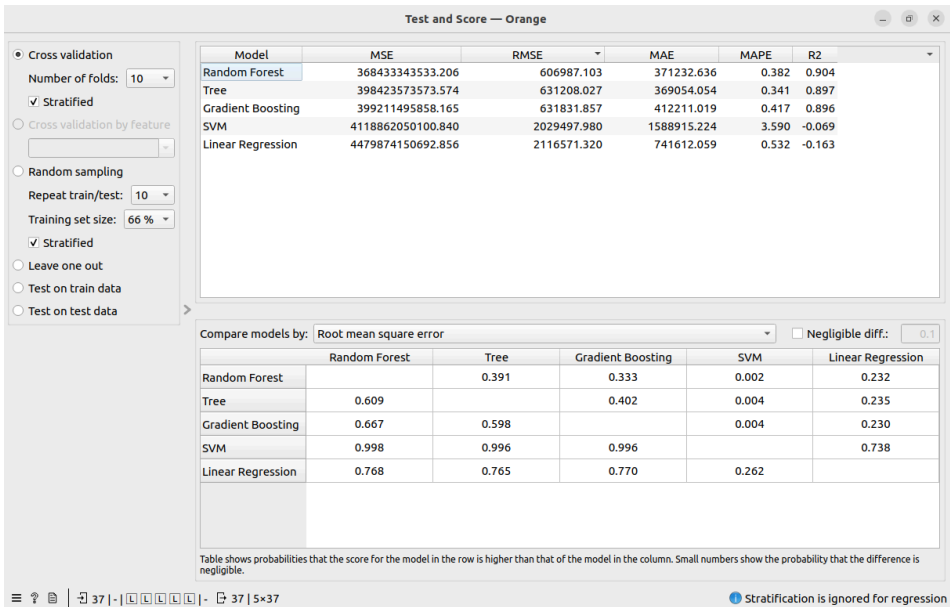
Gambar 8.34. Test Algoritma dengan widget Test & Score.

Untuk mengetes ketelitian algoritma, kita menggunakan widget Test & Score. Keluaran Data dari widget Select Columns dimasukan ke widget Test & Score, selain itu keluaran dari widget algoritma yang akan di test dimasukkan juga ke widget Test & Score. Pada kesempatan ini, karena kita tahu bahwa Penghasilan merupakan variabel yang linear, maka Algoritma yang di test adalah algoritma yang dapat menganalisis model dengan hasil linear, dalam hal ini,

- Tree
- Ransom Forst
- Linear Regression
- Gradient Boosting
- SVM

Hasil yang diperoleh diperlihatkan pada gambar berikut. Disini kita tertarik pada Error yang terkecil. Penjelasan lebih lengkap tentang masing-masing

Error dapat di baca di lampiran. Gambar hasil Test & Score diperlihatkan pada gambar berikut.



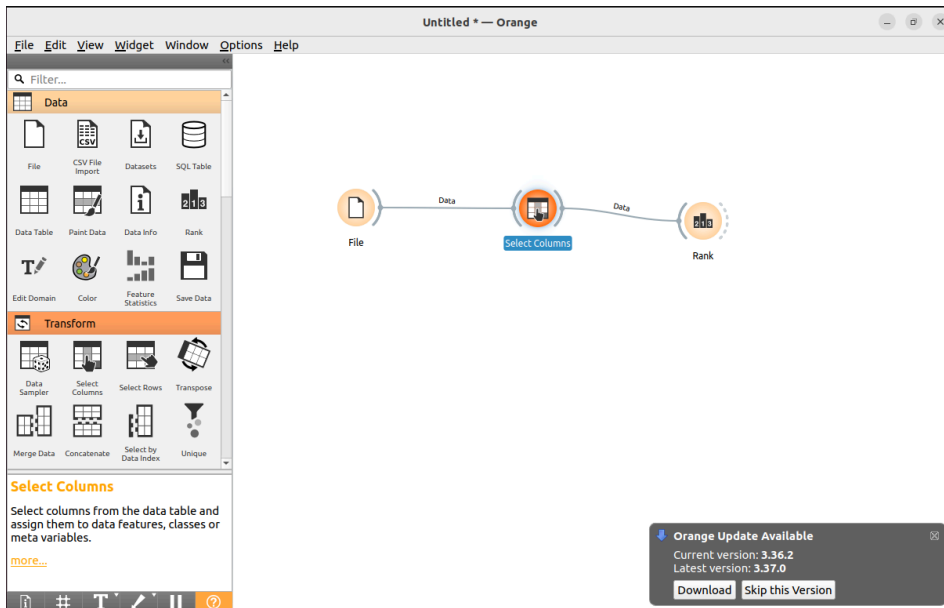
Gambar 8.35. Hasil widget Test & Score.

Terlihat dengan jelas hasil MSE, RMSE, MAE, MAPE sangat konsisten, menunjukkan urutan berikut untuk error yang paling kecil,

- Gradient Boosting
- Ransom Forst
- Tree
- Linear Regression
- SVM

Dengan Gradient Boosting menghasilkan error yang paling kecil maka sebaiknya kita menggunakan gradient boosting untuk memperoleh model dari data.

Analisa lanjutan yang mungkin akan menarik untuk diperhatikan adalah ranking dari fitur yang paling berpengaruh pada penghasilan. Cara paling sederhana untuk melakukan analisis rank adalah menyambungkan keluaran dari widget select columns ke widget rank seperti tampak pada gambar.



Gambar 8.36. Workflow untuk menganalisis Rank.

Hasilnya rank diperoleh diperlihatkan pada gambar berikut.

The screenshot shows the Orange Rank widget window. The table displays the following data:

	#	Univ...reg.	R...F
1 C PENDIDIKAN (1)	5	NA	0.867
2 C RUMAH	2	NA	0.686
3 C MOBIL	2	NA	0.674
4 C MOTOR	2	NA	0.670
5 N PENDIDIKAN (2)		NA	0.462
6 N ANAK		NA	0.227
7 N UTANG		NA	0.044

Gambar 8.37. Rank yang diperoleh.

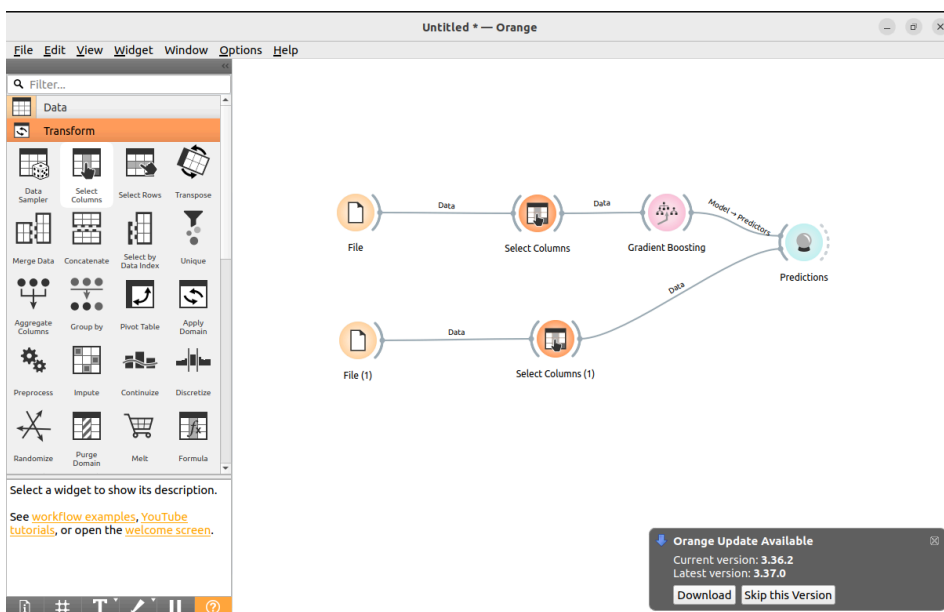
Tampak dari rank yang diperoleh bahwa urusan pengaruh terhadap penghasilan adalah,

- Pendidikan (jenis pendidikan)

- Rumah
- Mobil

Berdasarkan rank yang diperoleh, program KemenkopUKM yang paling relevan untuk meningkatkan penghasilan adalah meningkatkan pendidikan

Logika sederhana yang perlu diuji setelah memperoleh rank pengaruh berbagai fitur yang ada ke target penghasilan adalah melihat sebagai kuat pengaruh fitur tersebut kepada target penghasilan. Untuk kita kita dapat membuat perubahan dari file training data dengan menaikkan tingkat pendidikan semua orang untuk melihat bagaimana efek yang terjadi. Workflow yang digunakan tampak pada gambar berikut.



Gambar 8.38. Workflow memprediksi pengaruh kenaikan pendidikan.

Disini pada kolom lama pendidikan di file training data kita tambahkan 6 tahun untuk semua orang. Prediksi yang diperoleh cukup mengagumkan. Sebagian besar penghasilan orang ternyata akan naik dengan semakin tingginya pendidikan. Dari garis merah yang ada di hasil prediksi terlihat jelas bahwa kenaikan yang sangat signifikan terjadi bila pendidikan telah mencapai perguruan tinggi / universitas. Untuk pendidikan SD, SMP, SMA

perubahan dan kenaikan yang terjadi tidak terlalu dahsyat. Hal ini diperlihatkan pada gambar di bawah ini.

Predictions — Orange

Shown regression error: (None)

Restore Original Order

	Gradient Boosting	PENGHASILAN	NAMA	ENDIDIKAN (1)	ENDIDIKAN (2)	RUMAH	MOTOR	MOBIL	ANAK	UTANG
1	3577548.1	2000000.0	Amir	SMA	12.0	0.0	0.0	0.0	4.0	0.0
2	3283963.5	1500000.0	Badu	SMA	12.0	0.0	0.0	0.0	5.0	0.0
3	3472899.3	1700000.0	Candra	SMA	12.0	0.0	0.0	0.0	3.0	0.0
4	3993177.8	1000000.0	Dedi	UNIV	13.0	0.0	0.0	0.0	5.0	0.0
5	4181244.8	1900000.0	Eddy	UNIV	13.0	0.0	0.0	0.0	3.0	0.0
6	3472899.3	1980000.0	Fahri	SMA	12.0	0.0	0.0	0.0	3.0	0.0
7	4286762.3	1800000.0	Gagah	UNIV	13.0	0.0	0.0	0.0	4.0	0.0
8	4357834.7	1900000.0	Hadi	UNIV	15.0	0.0	0.0	0.0	2.0	0.0
9	4357834.7	1700000.0	Ira	UNIV	15.0	0.0	0.0	0.0	2.0	0.0
10	4673794.5	1900000.0	Jaja	UNIV	16.0	0.0	1.0	0.0	4.0	0.0
11	5011112.1	1700000.0	Kaka	UNIV	16.0	1.0	1.0	0.0	3.0	0.0
12	5011112.1	3000000.0	Lala	UNIV	15.0	1.0	1.0	0.0	3.0	0.0
13	5094496.6	2500000.0	Mami	UNIV	15.0	1.0	0.0	0.0	4.0	0.0
14	2820236.6	4000000.0	Nani	UNIV	17.0	1.0	1.0	1.0	0.0	100000.0
15	2923259.0	4200000.0	Opa	UNIV	17.0	1.0	1.0	1.0	1.0	1000000.0
16	3105704.0	3890000.0	Qra	UNIV	17.0	1.0	1.0	1.0	2.0	100000.0
17	3032174.5	2000000.0	Rani	UNIV	18.0	1.0	1.0	1.0	2.0	200000.0
18	6073836.2	5000000.0	Sasa	UNIV	18.0	1.0	1.0	1.0	3.0	0.0
19	6073836.2	4600000.0	Tati	UNIV	18.0	1.0	1.0	1.0	3.0	0.0
20	6101427.7	4700000.0	Udi	UNIV	19.0	1.0	1.0	1.0	2.0	0.0
21	6101427.7	6000000.0	Vava	UNIV	20.0	1.0	1.0	1.0	2.0	0.0
22	6101427.7	6500000.0	Wawa	UNIV	20.0	1.0	1.0	1.0	2.0	0.0
23	5558138.8	5500000.0	Xaxa	UNIV	24.0	1.0	1.0	1.0	1.0	0.0
24	6101427.7	6100000.0	Yaya	UNIV	25.0	1.0	1.0	1.0	2.0	0.0
25	1938424.1	200000.0	Mamir	SD	6.0	0.0	0.0	0.0	4.0	0.0
26	1109905.3	300000.0	Mbamir	SMP	7.0	0.0	0.0	0.0	6.0	0.0
27	1109905.3	400000.0	Mcamir	SMP	8.0	0.0	0.0	0.0	7.0	0.0
28	1875312.1	250000.0	Mdamir	SMP	9.0	0.0	0.0	0.0	4.0	0.0

Show performance scores

37 | 37 | 1x37

Gambar 8.39. Hasil Prediksi Penghasilan dengan kenaikan Pendidikan.

Melihat kondisi ini, langkah selanjutnya adalah menghitung investasi / anggaran untuk melakukan program perbaikan ini. Berapa besar subsidi SPP yang harus dialokasikan agar semua penduduk dapat mengenyam pendidikan setinggi mungkin, terutama untuk bisa mencapai universitas.

Komponen biaya ini akhirnya harus di optimasi agar dapat mencapai hasil semaksimal mungkin dengan biaya seminimal mungkin. Untuk bisa melakukan ini kita perlu membuat file tambahan yang berisi komponen biaya, baik subsidi SPP SD, SMP, SMA, dana operasional sekolah, beasiswa perguruan tinggi dll. Jika data ini bisa dikumpulkan dan optimasi bisa dilakukan bukan mustahil kita dapat memaksimalkan usaha peningkatan pendidikan dengan anggaran seminimal mungkin.

Contoh Sederhana Prediksi Cuaca

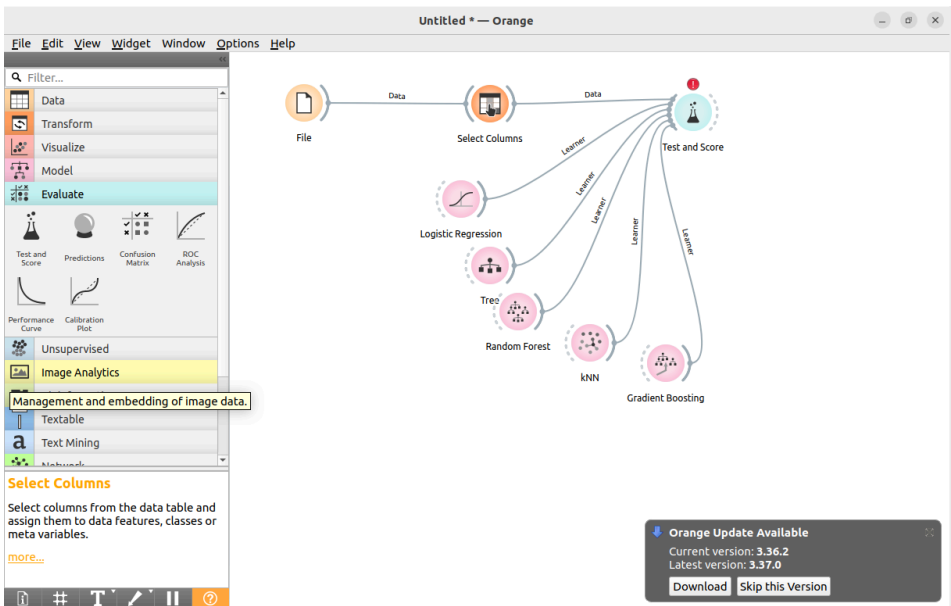
Contoh selanjutnya adalah memprediksi cuaca secara sederhana sekali. Untuk dapat melakukan prediksi yang lebih akurat perlu data yang lebih banyak dan tersebar di wilayah yang luas. Disini kita akan menggunakan data yang sangat sederhana sekali yang dapat diambil menggunakan Arduino menggunakan sensor DHT11 untuk mengukur suhu dan kelembaban. Data yang akan digunakan dalam file cuaca.xls diperlihatkan pada gambar berikut.

	A	B	C
1	<u>SUHU</u>	<u>LEMBAB</u>	<u>CUACA</u>
2	30	15	panas
3	29	17	panas
4	29.5	16	panas
5	23	80	hujan
6	24	87	hujan
7	28	20	panas
8	20	90	hujan
9	21	91	hujan
10	19	95	hujan
11	29	30	panas
12	29.5	20	panas
13	28	20	panas
14	19	100	hujan
15	20	95	hujan
16			

Gambar 8.40. Data Training Cuaca

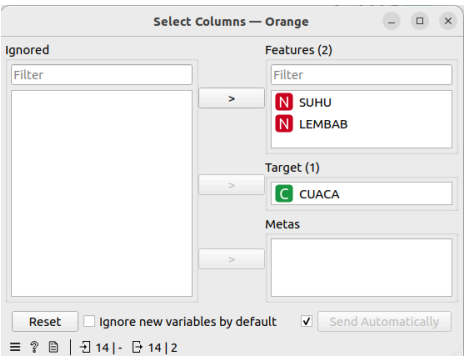
Bagi mata yang berpengalaman, terlihat bahwa dalam data training cuaca yang sederhana ini, fitur data adalah suhu dan kelembaban. Cuaca yang berisi pilihan panas atau hujan merupakan target untuk machine learning.

Langkah pertama yang perlu dilakukan adalah mengetes algoritma yang dapat kita gunakan untuk memperoleh model. Ini dapat diperoleh menggunakan workflow berikut ini.

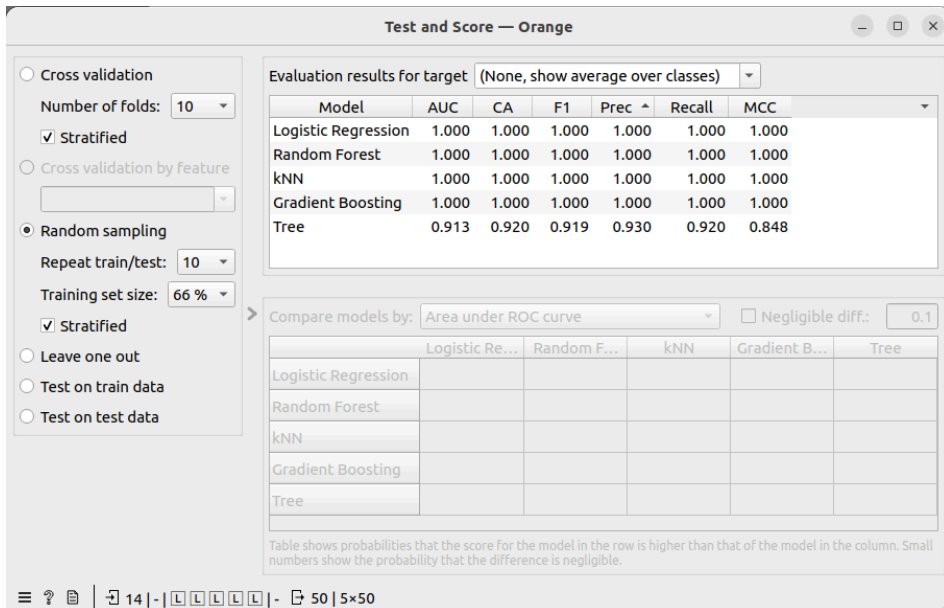


Gambar 8.41. Workflow Test Algoritma untuk Modeling Cuaca.

Disini kita perlu melakukan select column terlebih dahulu untuk memastikan bahwa target analisa adalah kolom cuaca, bukan kolom suhu atau kelembaban. Penampakan setting select columns diperlihatkan dalam gambar berikut.



Gambar 8.42. Select Columns untuk analisa cuaca. Hasil tes berbagai algoritma diperlihatkan pada gambar berikut.



Gambar 8.43. Hasil Test Algoritma untuk pemodelan Cuaca.

Pada gambar di atas di perlihatkan hasil tes dari beberapa algoritma untuk memperoleh model cuaca. Dengan kondisi data yang sangat terbatas, hasil yang diperoleh terlihat bahwa sebagian besar algoritma yang digunakan memberikan hasil yang baik. Algoritma yang di tes antara lain adalah,

- Logistic Regression
- Random Forest
- kNN
- Gradient Booting
- Tree

Semua memberikan hasil presisi 100% kecuali Tree yang memberikan akurasi hanya 93%. Untuk melakukan forecasting cuaca kita dapat menggunakan file ke dua yang berisi hasil pengukuran suhu dan kelembaban yang dapat kita kirim ke widget prediction untuk dibandingkan dengan model yang kita peroleh sebelumnya. Tampilan data suhu dan kelembaban sebagai data yang akan di prediksi diperlihatkan pada gambar berikut.

cuaca-ramalan.xls - LibreOffice Calc

File Edit View Insert Format Styles Sheet Data Tools Window Help

Arial 10 pt B I U A

C16 f_x Σ =

	A	B	C
1	<u>SUHU</u>	<u>LEMBAB</u>	<u>CUACA</u>
2	30	15	
3	28	17	
4	29.5	16	
5	22	80	
6	24	87	
7	29	20	
8	19.5	90	
9	20.5	91	
10	19	95	
11	29	30	
12	29.5	20	
13	28	20	
14	19	100	
15	20	95	
16			

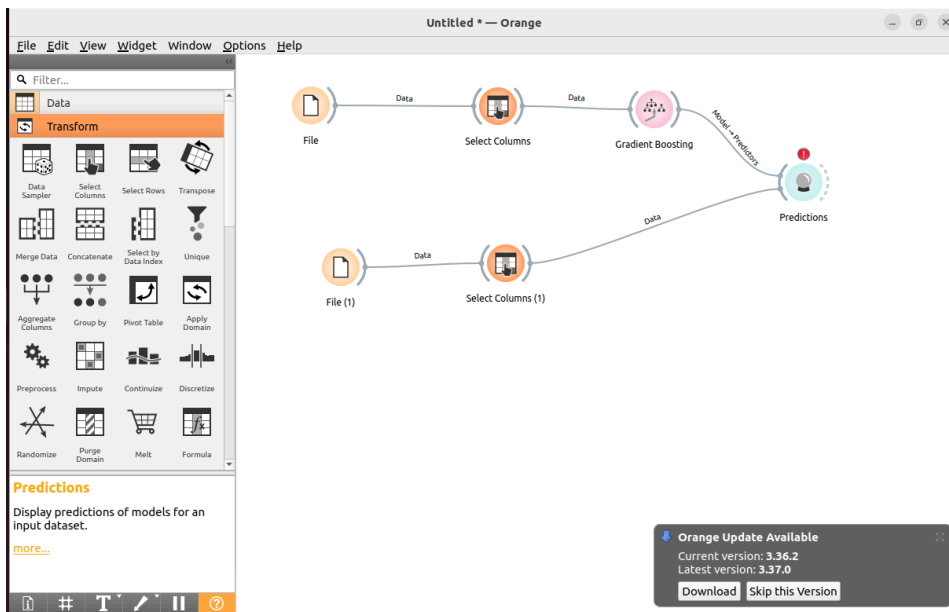
Sheet1

Find Find All Formatted Display

Sheet 1 of 1 PageStyle_Sheet1 English (USA) 200%

Gambar 8.44. Data Cuaca untuk di prediksi.

Gambar secara umum workflow untuk keperluan ini dapat dilihat berikut ini.



Gambar 8.45. Workflow untuk memprediksi cuaca.

Hasil prediksi cuaca dari data cuaca di bandingkan dengan model cuaca yang diperlihatkan pada gambar berikut.

Predictions			
Show probabilities for (None) Show results			
	Gradient Boosting	CUACA	SUHU
1	panas	?	30.0
2	panas	?	28.0
3	panas	?	29.5
4	hujan	?	22.0
5	hujan	?	24.0
6	panas	?	29.0
7	hujan	?	19.5
8	hujan	?	20.5
9	hujan	?	19.0
10	panas	?	29.0
11	panas	?	29.5
12	panas	?	28.0
13	hujan	?	19.0
14	hujan	?	20.0

Gambar 8.46. Hasil Prediksi Cuaca.

Dengan logika yang sama, kita dapat mengembangkan aplikasi untuk peramalan cuaca yang lebih baik jika kita dapat mengumpulkan data dalam jumlah besar dalam wilayah yang luas.

Contoh Analisa untuk SATU Indonesia Awards

Salah satu kegiatan yang pernah dilakukan penulis adalah melakukan modeling terhadap penjurian SATU Indonesia AWARDS. Detail teknisnya bisa dibaca di paper jurnal,

Purbo, O.W., 2021. Comparing Machine Learning and Human Judge in SATU Indonesia Awarding Processes. Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI), 7(3), pp.453-460.

Selama ini proses penjurian SATU Indonesia AWARDS dilakukan secara manual. Nominasi dimasukan ke panitia berbentuk tulisan mencakup, visi, kesulitan, keluaran (output), pemanfaatan (outreach), dan sustainability. Tulisan ini kemudian dinilai oleh juri manusia, seperti Prof. Emil Salim, Prof. Nila Moeloek, Prof. Fasli Jalal dll. Sehingga dalam bahasa yang sederhana, format data yang akan kita analisis kira-kira sebagai berikut,

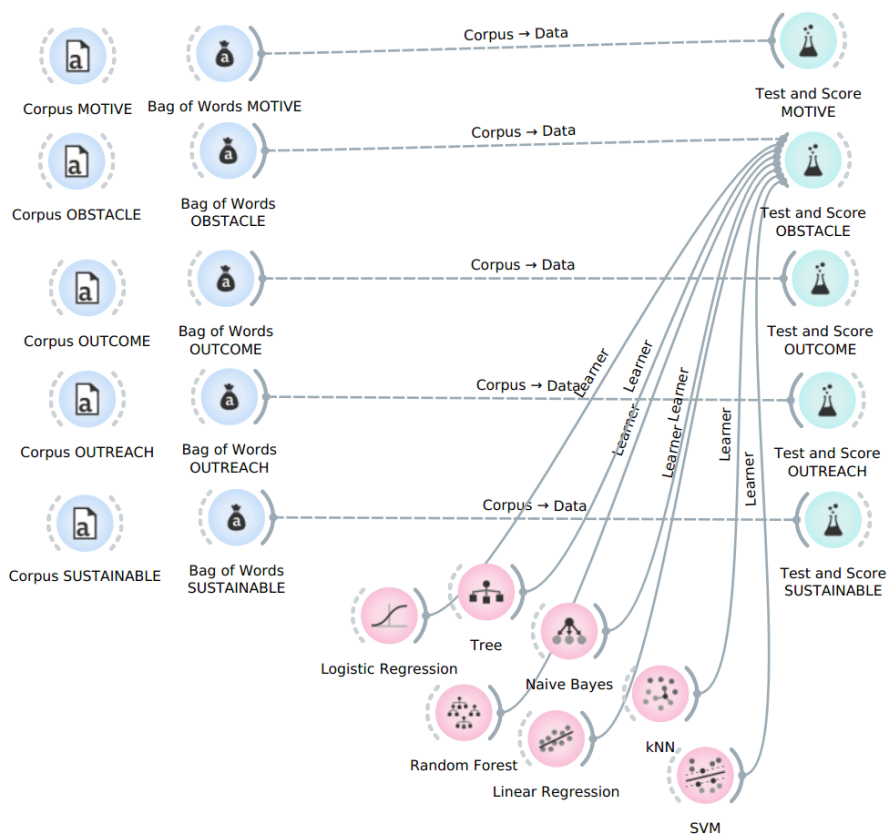
Tabel 8.2. Contoh Data Nilai SATU Indonesia Awards

Motive	Obstacle	Outcome	Outreach	Sustainable	Nilai
a sahdkj m	kjshdkj sfduy	jfsdtx hjkdsh	jdghsj jdklsj	ksuydu jdhsj	7
jksdhgj	smjdhi gjhhg	jdshj	,jdshjk hjt	,mshfdui kjdngdja	9
gdjksh hjkdsh	,dhskjljl dkjh	s,jahdkj mkjashd	jsahdkjm kjahsd	mnasgd	1
akjsdf jhasgdjh	,sajdhjk kjashjkfd	askjdh kjahsdjk	askjdh kjhasjk	kjsadhmahs	9

Bayangkan setiap tahun adalah puluhan ribu nominasi berbentuk tulisan yang harus dibaca dan di beri nilai. Tantangan terbesarnya adalah bagaimana caranya agar kita dapat menilai secara konsisten dalam

jumlah puluhan ribu dan dilakukan dalam waktu terbatas biasanya sekitar satu minggu saja. Konsekuensinya jika ini dilakukan oleh orang, maka kemungkinan besar akan sulit untuk bisa memperoleh hasil yang konsisten. Cara yang paling baik adalah menggunakan mesin untuk melakukan penilaian tersebut.

Kebetulan data nilai ini sudah terkumpul selama belasan tahun, jadi ada tulisan masing-masing nominasi beserta nilai juri-nya. Kalau kita bandingkan dengan model linear $A \times B = Y$ maka X berisi tulisan masing-masing nominasi dan nilai juri (target) adalah Y. Tugas aplikasi yang dibuat adalah mencari nilai dari model A dan B. Setelah model A dan B diperoleh, maka selanjutnya kita bisa melakukan prediksi nilai Y secara konsisten jika ada tulisan X baru yang masuk. Akibatnya, proses penjurian bisa dilakukan dengan cepat dan konsisten.



Gambar 8.47. Berbagai algoritma terhadap data yang ada.

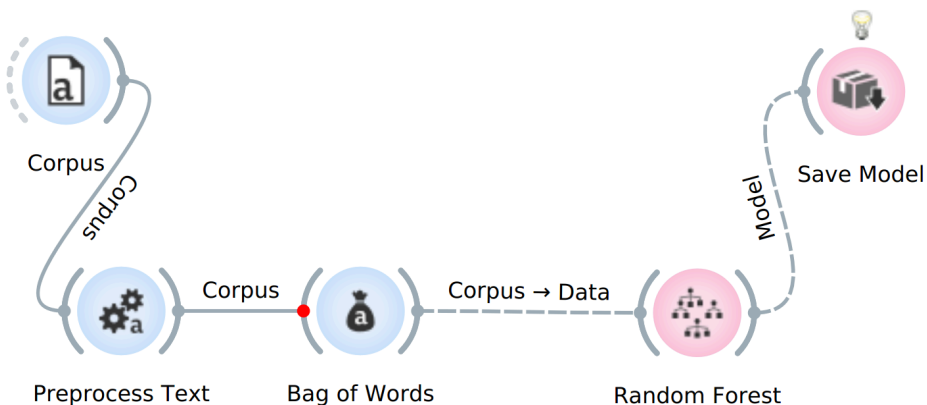
Sebelum melakukan pemodelan, kita perlu mengetes dulu algoritma mana yang paling baik. Data text di masukan ke widget bag of words, sebelum dimasukan ke widget test seperti pada gambar.

Ada 5 data berbentuk text di file excel yang dikenal sebagai corpus yang harus di test, yaitu,

- Motive
- Obstacle
- Outcome
- Outreach
- Sustainable

Dari lima (5) widget test and score yang ada kita dapat menilai algoritma mana yang baik untuk masing-masing jenis data. Disini memang ada beberapa algoritma yang di test, seperti, Tree, Logistic Regression, Random Forest, Naive Bayes, kNN, Linear Regression, SVM. Hasil dari test and score menunjukan algoritma random forest biasanya akan memberikan hasil paling presisi untuk data dari bag of words data SATU Indonesia awards.

Langkah selanjutnya adalah membuat menyimpan model untuk masing-masing corpus. Adapun penampakan program untuk masing-masing model adalah seperti tampak pada gambar.



Gambar 8.48. Program untuk Memperoleh model dari masing2 Corpus.

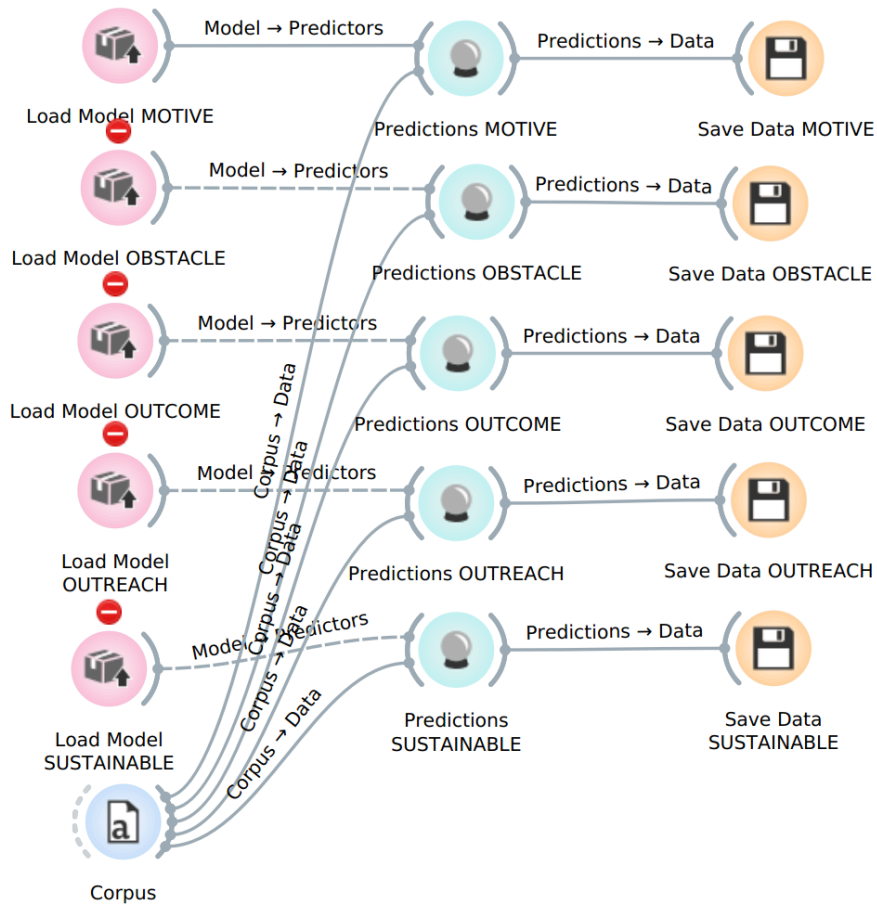
Secara sederhana dapat diceritakan, data dari corpus dimasukkan ke preprocess text agar membuang kata-kata yang tidak dibutuhkan seperti kata penghubung, yang, di, ke, dari dll.

Selanjutnya data dimasukkan ke bag of words, yang akan menghitung jumlah masing-masing kata. Dalam imajinasi yang sederhana, sebetulnya kita membangun koordinat dimana masing-masing kata menjadi sumbu koordinat dengan jumlah masing-masing kata kita dapat memperoleh vector dari sebuah text. Konsekuensinya Bag of Words dari sebuah Corpus pada dasarnya merupakan kumpulan vector dari text di corpus.

Data bag of words dari corpus kemudian dikirimkan ke algoritma random forest untuk memperoleh model dari corpus. Proses kalkulasi model biasanya merupakan proses iterasi / regresi yang akan memakan waktu lama jika kita menghadapi data yang besar. Model dari random forest dapat di rekam / di save menggunakan widget save untuk dapat digunakan kemudian. Tentunya kita perlu membuat model untuk masing-masing corpus, dalam hal ini, motive, obstacle, outcome, outreach dan sustainable.

Setelah kita memiliki model dari masing-masing jenis text, langkah selanjutnya adalah memprediksi nilai dari sebuah text yang masuk. Adapun program untuk melakukan prediksi nilai tersebut dapat kita lihat pada gambar.

Secara umum data baru berupa sebuah corpus akan dimasukkan ke widget prediksi yang memperoleh masukan dari lima model yang kita peroleh sebelumnya. Secara sederhana hasil dari widget prediksi kemudian dapat di save untuk dilakukan analisa lebih lanjut.



Gambar 8.49. Prediksi nilai dari model.

Secara umum, setelah model di peroleh, maka hanya butuh waktu beberapa menit biasanya tidak sampai 30 menit untuk menilai / menganalisa semua data nominasi yang masuk. Hal ini menyebabkan semua proses penilaian SATU Indonesia Awards menjadi sangat cepat. Tentunya hasil akhir perlu di konfirmasi ulang oleh juri manusia untuk memastikan bahwa hasil yang diperoleh baik dan sesuai dengan keinginan para juri.

BAB 9: Analisis Data Koperasi dan UKM

Data Koperasi dan UMKM

Kita cukup beruntung karena KemenkopUKM telah melakukan upaya yang signifikan dalam mengumpulkan data UMKM di Indonesia. Saat ini, terdapat lebih dari 13 juta UMKM yang telah terdata. Data yang dikumpulkan mencakup berbagai aspek penting, seperti lokasi usaha, modal awal, waktu berdirinya usaha, karakteristik pemilik usaha, skala usaha, jenis produk atau jasa yang dihasilkan, hingga data keuangan.

Lebih dari 100+ fitur yang dikumpulkan dari masing-masing UMKM, sebagian dari fitur data UMKM yang dikumpulkan adalah,

Provinsi, Kabupaten, Kecamatan, Kelurahan, Modal Pendirian, Bulan Mulai Beroperasi, Tahun Mulai Beroperasi, Jenis Kelamin, Disabilitas, Status Pengusaha, Pendidikan Pengusaha, Skala Usaha, Kategori KBLI, Metoda Pemasaran, Penggunaan Teknologi, Kepemilikan Laporan Keuangan, Kepemilikan NIB, Omzet, Tenaga Kerja Lak-Laki, Tenaga Kerja Perempuan

Data tersebut sebetulnya disimpan dalam bentuk database, penampakan data dalam aplikasi spreadsheet diperlihatkan pada gambar di bawah ini. Informasi yang begitu detail ini sangat berharga bagi pengembangan UMKM di Indonesia.

Dengan adanya data yang komprehensif ini, pemerintah, khususnya KemenkopUKM, dapat merancang kebijakan dan program yang lebih tepat sasaran untuk mendukung pertumbuhan UMKM. Misalnya, data mengenai skala usaha dapat digunakan untuk menyusun program pendampingan yang sesuai. Data mengenai penggunaan teknologi dapat menjadi acuan dalam program digitalisasi UMKM. Selain itu, data ini juga dapat dimanfaatkan oleh lembaga keuangan untuk memberikan akses pembiayaan yang lebih mudah bagi UMKM.

No	Provinsi	Kabupaten/Kecamatan	Kelurahan	Modal	Beroperasi	Mulai Beroperasi	Jenis Kelamin	Disabilitas	Status Pengusaha	Pendirian	Pengusaha	Usaha	Kategori	KBLI
1	JAWA BARAT	BOGOR	BOGOR	1,000,000	4	2019P	Pemilik	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah
2	JAWA BARAT	BOGOR	BOGOR	1,000,000	5	2020P	Pemilik	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah
3	JAWA BARAT	BOGOR	BOGOR	1,000,000	6	2019P	Pemilik	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah
4	JAWA BARAT	BOGOR	BOGOR	500,000	4	2020P	Pemilik	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah
5	JAWA BARAT	BOGOR	BOGOR	2,000,000	6	2019P	Pemilik	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah
6	JAWA BARAT	BOGOR	BOGOR	1,000,000	3	2019P	Pemilik	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah
7	JAWA BARAT	BOGOR	BOGOR	1,000,000	3	2019P	Pemilik	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah
8	JAWA BARAT	BOGOR	BOGOR	1,000,000	3	2019P	Pemilik	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah
9	JAWA BARAT	BOGOR	BOGOR	1,000,000	3	2019P	Pemilik	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah
10	JAWA BARAT	BOGOR	BOGOR	1,000,000	3	2019P	Pemilik	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah
11	JAWA BARAT	BOGOR	BOGOR	1,000,000	3	2019P	Pemilik	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah
12	JAWA BARAT	BOGOR	BOGOR	1,000,000	3	2019P	Pemilik	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah
13	JAWA BARAT	BOGOR	BOGOR	1,000,000	3	2019P	Pemilik	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah
14	JAWA BARAT	BOGOR	BOGOR	1,000,000	3	2019P	Pemilik	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah
15	JAWA BARAT	BOGOR	BOGOR	1,000,000	3	2019P	Pemilik	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah
16	JAWA BARAT	BOGOR	BOGOR	1,000,000	3	2019P	Pemilik	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah
17	JAWA BARAT	BOGOR	BOGOR	1,000,000	3	2019P	Pemilik	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah
18	JAWA BARAT	BOGOR	BOGOR	1,000,000	3	2019P	Pemilik	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah
19	JAWA BARAT	BOGOR	BOGOR	1,000,000	3	2019P	Pemilik	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah
20	JAWA BARAT	BOGOR	BOGOR	1,000,000	3	2019P	Pemilik	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah
21	JAWA BARAT	BOGOR	BOGOR	1,000,000	3	2019P	Pemilik	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah
22	JAWA BARAT	BOGOR	BOGOR	1,000,000	3	2019P	Pemilik	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah
23	JAWA BARAT	BOGOR	BOGOR	1,000,000	3	2019P	Pemilik	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah
24	JAWA BARAT	BOGOR	BOGOR	1,000,000	3	2019P	Pemilik	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah
25	JAWA BARAT	BOGOR	BOGOR	1,000,000	3	2019P	Pemilik	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah
26	JAWA BARAT	BOGOR	BOGOR	1,000,000	3	2019P	Pemilik	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah
27	JAWA BARAT	BOGOR	BOGOR	1,000,000	3	2019P	Pemilik	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah	Tanah

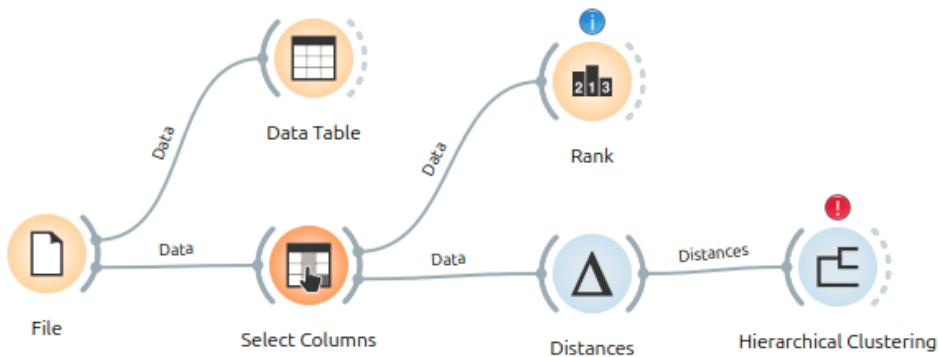
Gambar 9.1 Tampilan Data di Aplikasi Spreadsheet.

Orange Workflows untuk analisa

Orange Workflows menawarkan cara yang intuitif untuk menganalisis data koperasi dan UMKM. Gambar keseluruhan Orange Workflows yang digunakan diperlihatkan dalam Gambar di bawah ini. Prosesnya dimulai dengan mengimpor data dari file Excel menggunakan widget "File". Setelah data berhasil diimpor, kita bisa langsung melihat gambaran umum data tersebut melalui widget "Data Table". Widget ini memungkinkan kita untuk memeriksa apakah data sudah benar-benar terbaca dengan baik dan tidak ada kesalahan pada data.

Langkah selanjutnya adalah menentukan fitur dan target yang akan kita analisis. Fitur adalah karakteristik dari data yang akan kita gunakan untuk mengelompokkan data, misalnya omset, jumlah karyawan, atau jenis usaha. Sementara itu, target adalah variabel yang ingin kita prediksi atau pahami lebih dalam, seperti tingkat keberhasilan usaha. Untuk menentukan fitur dan target, kita akan menggunakan widget "Select Columns". Setelah fitur dan target ditentukan, kita dapat menggunakan widget "Rank" untuk melihat fitur mana yang paling berpengaruh terhadap

target. Informasi ini akan membantu kita dalam menyeleksi fitur yang relevan dan menghilangkan fitur yang tidak relevan.



Gambar 9.2. Orange Workflows analisa Data Koperasi dan UMKM.

Setelah proses persiapan data dan pemilihan fitur selesai, kita dapat memulai analisis menggunakan metode *unsupervised learning*. Salah satu algoritma yang sering digunakan dalam *unsupervised learning* adalah pengelompokan data (clustering). Pada Orange, kita dapat melakukan pengelompokan data menggunakan widget "Distances". Widget ini akan mengelompokkan data berdasarkan kesamaan fitur yang telah kita tentukan sebelumnya. Hasil pengelompokan ini kemudian dapat kita visualisasikan menggunakan widget "Hierarchical Clustering". Visualisasi ini akan menampilkan hierarki kelompok data yang terbentuk, sehingga kita dapat melihat hubungan antar kelompok data secara lebih jelas.

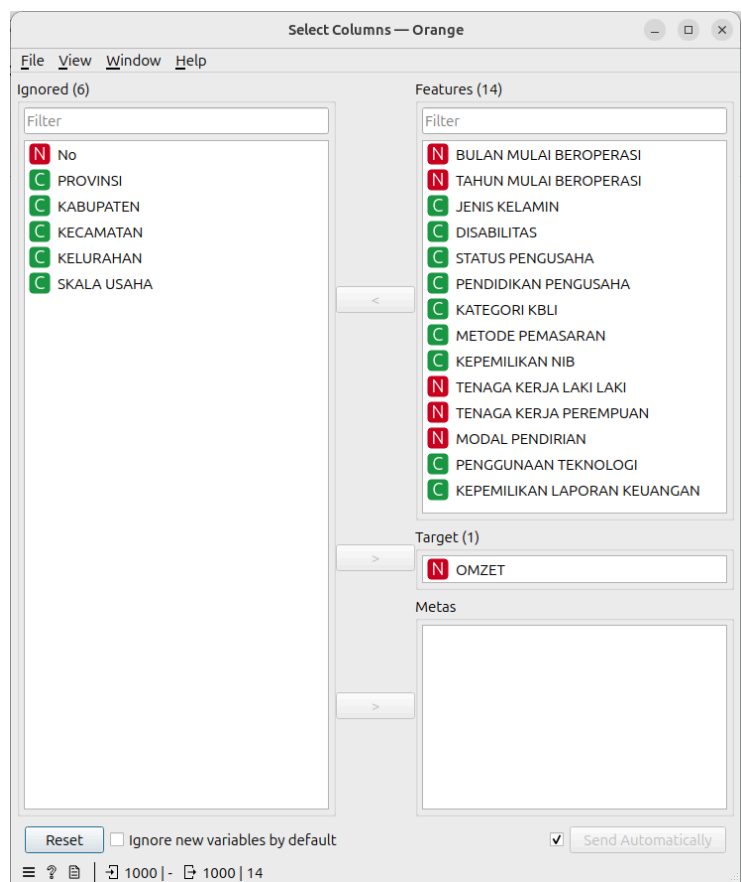
Dengan menggunakan Orange Workflows, kita dapat menemukan pola-pola menarik dalam data koperasi dan UMKM. Misalnya, kita dapat mengidentifikasi kelompok usaha yang memiliki karakteristik yang sama, seperti jenis usaha, ukuran usaha, atau lokasi geografis. Selain itu, kita juga dapat mengidentifikasi faktor-faktor yang mempengaruhi kinerja usaha, seperti pengaruh modal awal terhadap omzet atau pengaruh penggunaan teknologi terhadap produktivitas. Informasi ini sangat berharga bagi para pembuat kebijakan, pelaku usaha, dan pihak-pihak lain yang berkepentingan dalam pengembangan UMKM.

Orange Workflows merupakan tool yang sangat berguna untuk menganalisis data koperasi dan UMKM. Dengan antarmuka yang user-friendly dan berbagai widget yang tersedia, kita dapat melakukan

analisis data secara efektif dan efisien. Mulai dari persiapan data, pemilihan fitur, hingga visualisasi hasil, semua proses dapat dilakukan dengan mudah melalui Orange Workflows.

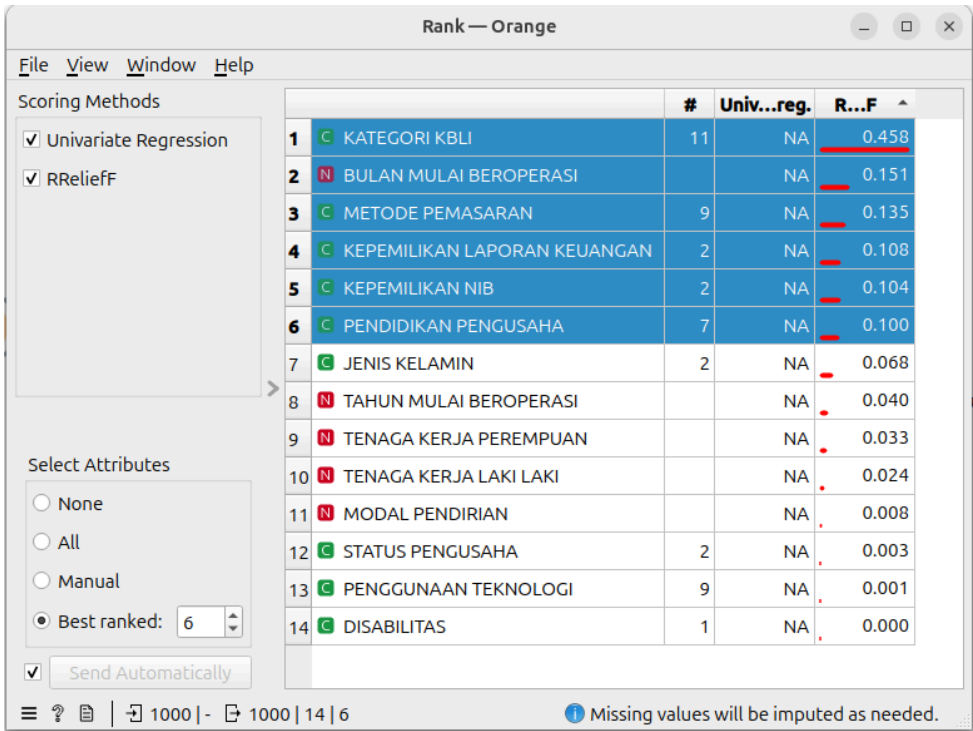
Analisa OMZET

Kita akan menggunakan perangkat lunak Orange Workflows untuk menggali lebih dalam data omzet dari koperasi dan UMKM. Metode yang akan kita gunakan adalah *unsupervised learning*, yaitu suatu teknik analisis data yang memungkinkan komputer menemukan pola-pola tersembunyi dalam data tanpa kita perlu memberitahu sebelumnya apa yang harus dicari. Dengan kata lain, kita akan membiarkan perangkat lunak mengidentifikasi kelompok-kelompok usaha yang memiliki karakteristik omzet yang serupa.



Gambar 9.3 widget Select Columns untuk analisa Omzet.

Langkah pertama yang perlu dilakukan adalah mempersiapkan data. Melalui widget "Select Columns", kita akan menentukan kolom mana yang akan menjadi fokus analisis kita. Kolom "Omzet" akan kita tetapkan sebagai target yang ingin kita prediksi atau pahami lebih dalam. Sementara itu, kolom-kolom yang tidak relevan seperti nomor urut, No, Provinsi, Kabupaten, Kecamatan, Kelurahan, dan Skala Usaha (yang dalam kasus ini semuanya mikro) akan kita abaikan atau kita kategorikan sebagai "ignored". Sisa kolom yang ada, seperti modal awal, jumlah karyawan, atau jenis usaha, akan kita jadikan sebagai fitur untuk menganalisis pola-pola yang mungkin mempengaruhi omzet.



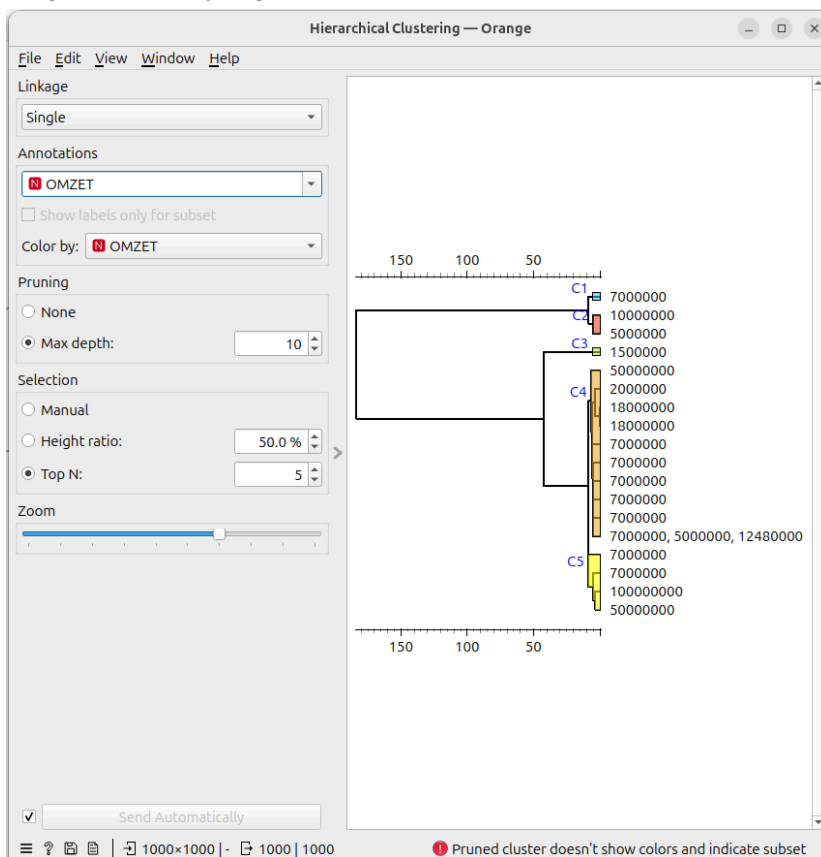
Gambar 9.4 Rank untuk target Omzet.

Dalam orange workflows ada dua tool, yaitu "widget Select Columns" dan "widget Rank". Hasil dari pemilihan kolom dari "widget Select Columns" langsung kita masukkan ke "widget Rank". Rank mempunyai banyak method untuk menentukan peringkat, jadi kita bisa pilih method mana yang paling cocok. Misalnya, kita bisa pakai method ReliefF untuk melihat kolom mana yang paling berpengaruh terhadap omzet.

Dari scoring method ReliefF, terlihat bahwa yang paling menentukan omzet berurut ke bawah sebagai berikut,

- Kategori KBLI
- Bulan Mulai Beroperasi
- Metoda Pemasaran
- Kepemilikan Laporan Keuangan
- Kepemilikan NIB
- Pendidikan Pengusaha

Uniknya, modal usaha ternyata tidak terlalu berpengaruh. Jadi, kesimpulannya, jenis usaha dan cara pemasaran yang tepat, serta pendidikan pengusaha, lebih penting untuk meningkatkan omzet dibandingkan modal yang besar.



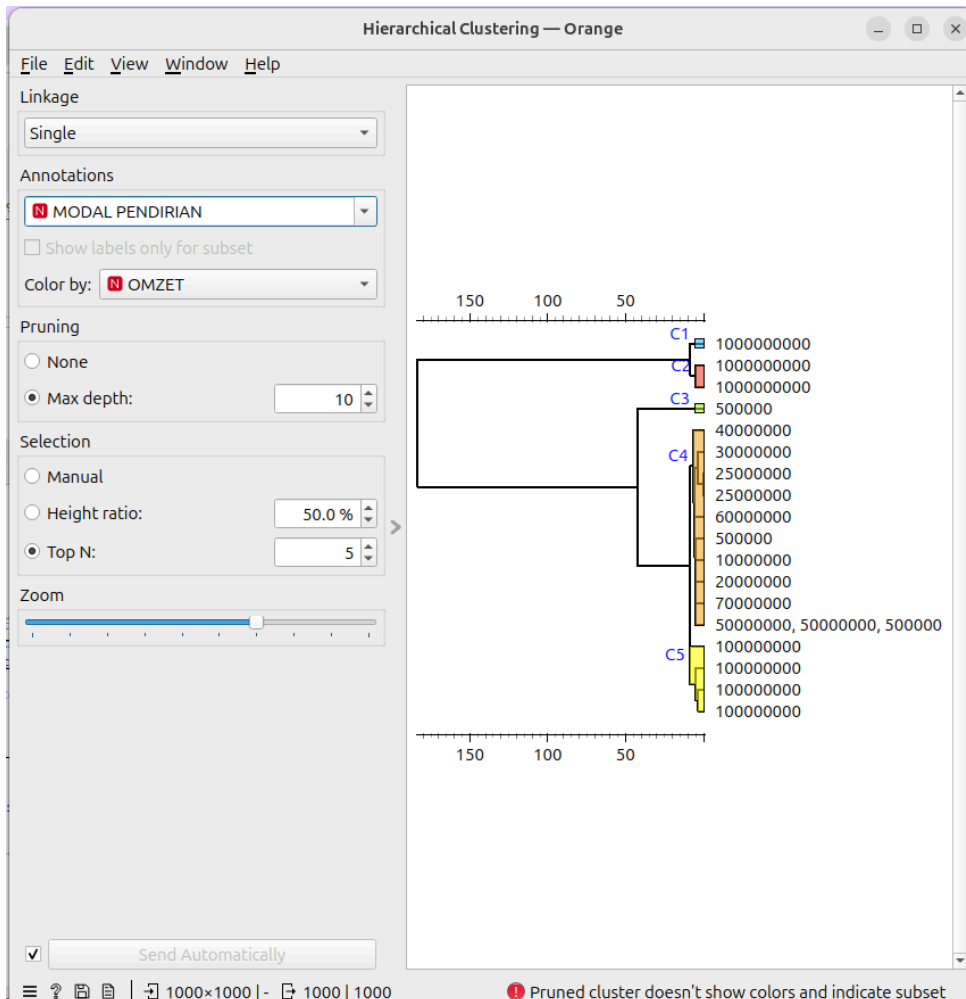
Gambar 9.5 Hasil Cluster Omzet

Hasil pemilihan kolom data menggunakan "widget Select Columns" dapat langsung kita gunakan untuk menganalisis kelompok data secara otomatis (unsupervised learning) dengan "widget Distance". Untuk melihat hasil pengelompokan secara visual, kita bisa menghubungkan lagi hasil dari "widget Distance" ke "widget Hierarchical Clustering".

Gambar Hierarchical Clustering yang ditampilkan menunjukkan hasil pengelompokan data secara hierarkis. Kita hanya menampilkan 5 kelompok utama saja untuk mempermudah analisis. Dari gambar ini, terlihat ada tiga kelompok besar,

- Kelompok atas
- Kelompok tengah
- Kelompok bawah

Kelompok data di bagian atas dan bawah memiliki nilai omzet yang hampir sama, yaitu puluhan juta. Kita perlu mencari tahu lebih dalam mengapa kedua kelompok ini terpisah. Sedangkan kelompok di tengah memiliki omzet yang biasa-biasa saja, namun bagian ujung atas dari kelompok ini memiliki omzet yang sedikit lebih besar. Kita juga perlu menyelidiki lebih lanjut apa penyebab perbedaan ini.

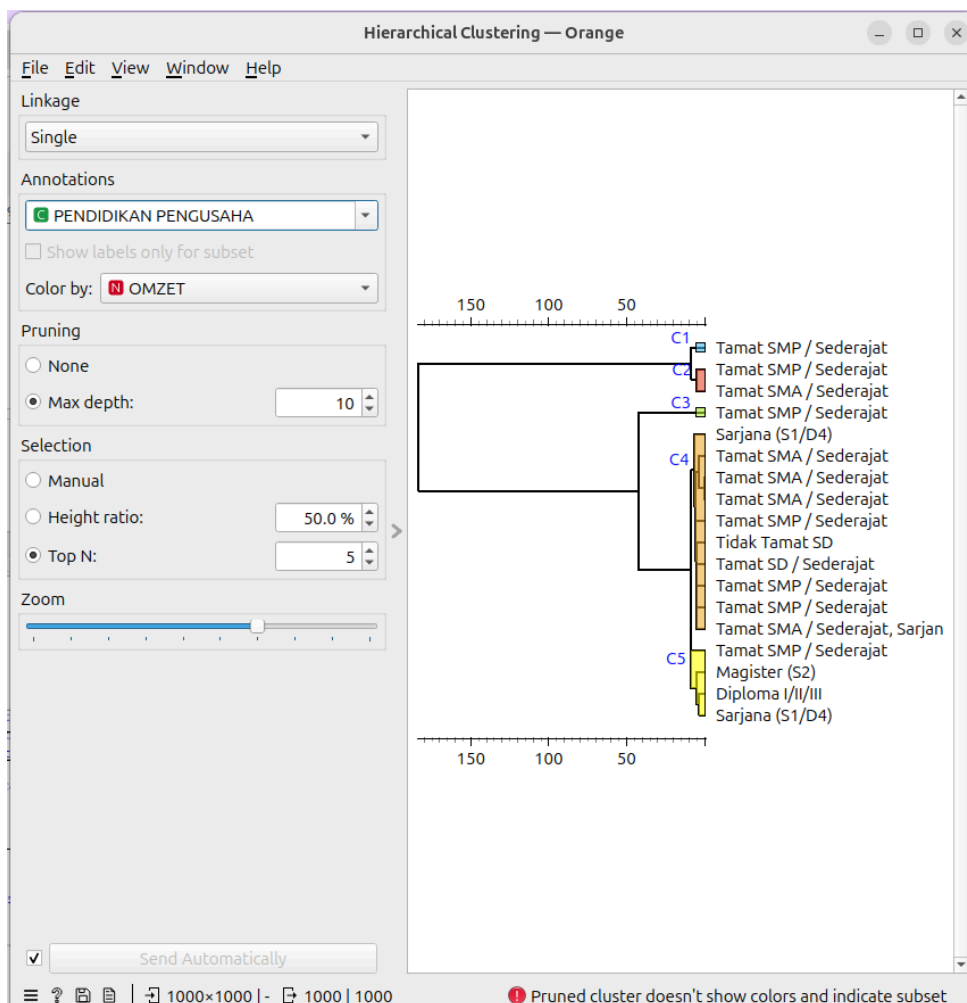


Gambar 9.6 Clustering Omzet, dengan label Modal Pendirian

Ketika kita mengganti ukuran perusahaan berdasarkan modal pendirian dalam visualisasi kelompok perusahaan (hierarchical clustering), kita melihat hal menarik. Meskipun begitu, kelompok perusahaan di bagian atas dan bawah visualisasi ini memiliki modal pendirian yang hampir sama. Jadi, pasti ada faktor lain yang membuat kedua kelompok ini terpisah dalam visualisasi ini.

Jika kita perhatikan kelompok perusahaan di tengah visualisasi, kita bisa lihat bahwa modal pendirian mereka lebih kecil dibandingkan dua kelompok di atas dan bawah. Namun, di bagian atas kelompok tengah, ada beberapa perusahaan dengan modal sedikit lebih besar. Hal ini

sepertinya berhubungan dengan omzet mereka yang juga sedikit lebih tinggi dibandingkan perusahaan lain di kelompok tengah.



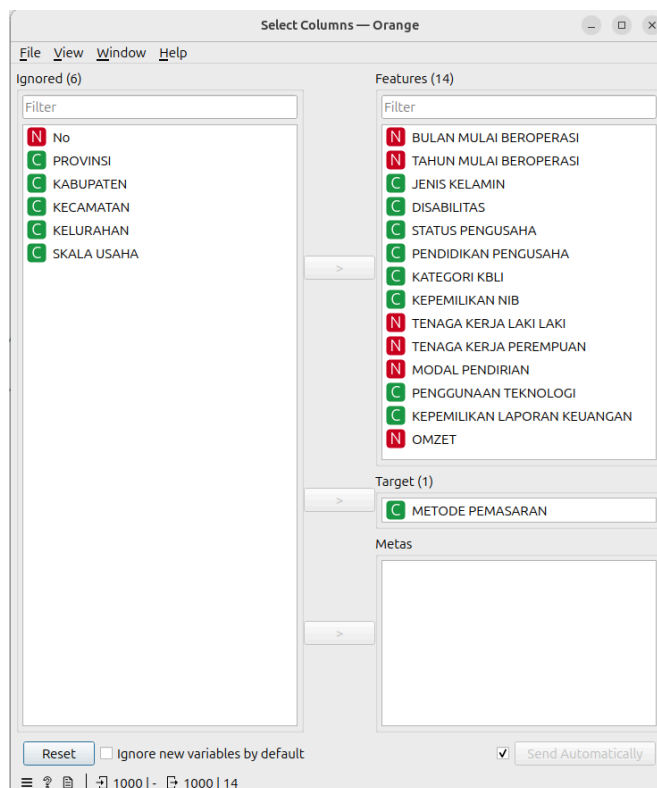
Gambar 9.7 Clustering Omzet, dengan Label Pendidikan Pengusaha

Ada hal menarik saat kita bandingkan tingkat pendidikan dengan omzet yang dihasilkan pengusaha. Ternyata, pengusaha dengan omzet besar tidak selalu memiliki pendidikan tinggi. Tampak bahwa kelompok atas dan bawah yang mempunyai omzet relatif sama, ternyata berbeda signifikan dari sisi pendidikan. Di kelompok atas, kebanyakan pendidikan SMP dan SMA, sementara kelompok bawah mempunyai pendidikan tinggi bahkan S2.

Pada kelompok pengusaha dengan modal sedang, kita menemukan pola yang berbeda. Di sini, pengusaha dengan pendidikan tinggi cenderung memiliki omzet yang lebih besar

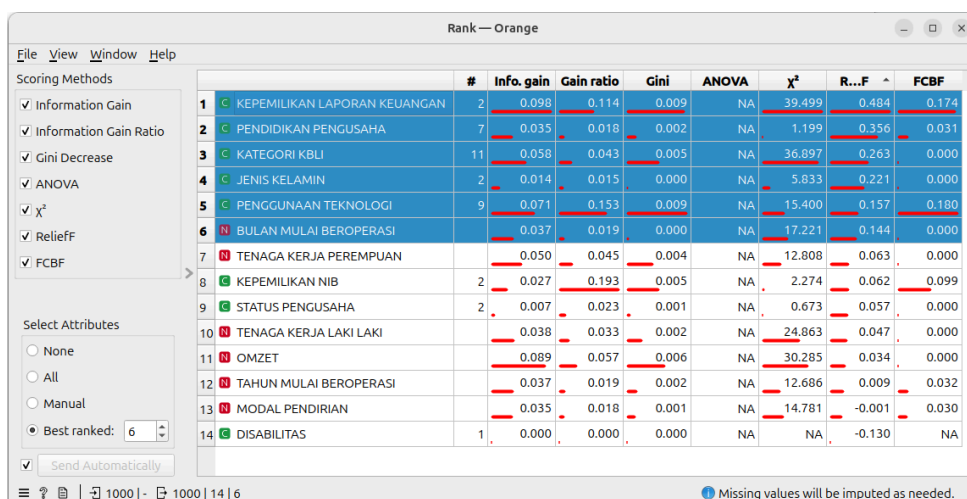
Analisa Metoda Pemasaran

Pada bagian ini, kita akan melakukan analisis mendalam terhadap berbagai metode pemasaran yang digunakan oleh koperasi dan UMKM. Dengan memanfaatkan perangkat lunak Orange Workflows, kita akan mengungkap pola-pola tersembunyi dalam data tanpa perlu menentukan secara spesifik apa yang ingin kita cari. Teknik yang akan kita gunakan disebut *unsupervised learning*, di mana komputer secara mandiri akan mengelompokkan usaha-usaha yang memiliki karakteristik pemasaran yang serupa. Hal ini memungkinkan kita untuk menemukan kesamaan dan perbedaan dalam strategi pemasaran yang diterapkan oleh berbagai jenis usaha.



Gambar 9.8 widget Select Columns untuk analisa metoda pemasaran.

Tahap awal dari analisis ini adalah persiapan data. Kita akan menggunakan widget "Select Columns" untuk memilih kolom data yang relevan. Kolom "Metode Pemasaran" akan menjadi fokus utama analisis kita, sementara kolom-kolom yang tidak berkaitan, seperti informasi administratif dan skala usaha, akan kita abaikan. Kolom-kolom lainnya, seperti modal awal, omzet, dan jumlah karyawan, akan kita gunakan sebagai variabel untuk mengidentifikasi pola-pola yang mungkin mempengaruhi pemilihan metode pemasaran. Dengan kata lain, kita akan mencoba mencari tahu faktor-faktor apa yang membuat suatu usaha memilih metode pemasaran tertentu.



Rank — Orange

File View Window Help

Scoring Methods

- ☒ Information Gain
- ☒ Information Gain Ratio
- ☒ Gini Decrease
- ☒ ANOVA
- ☒ χ^2
- ☒ ReliefF
- ☒ FCBF

Select Attributes

- ☐ None
- ☐ All
- ☐ Manual
- ☒ Best ranked: 6

☒ Send Automatically

	#	Info. gain	Gain ratio	Gini	ANOVA	χ^2	R...F	FCBF
1 C KEPEMILIKAN LAPORAN KEUANGAN	2	0.098	0.114	0.009	NA	39.499	0.484	0.174
2 C PENDIDIKAN PENGUSAHA	7	0.035	0.018	0.002	NA	1.199	0.356	0.031
3 C KATEGORI KBLI	11	0.058	0.043	0.005	NA	36.097	0.263	0.000
4 C JENIS KELAMIN	2	0.014	0.015	0.000	NA	5.833	0.221	0.000
5 C PENGGUNAAN TEKNOLOGI	9	0.071	0.153	0.009	NA	15.400	0.157	0.180
6 N BULAN MULAI BEROPERASI		0.037	0.019	0.000	NA	17.221	0.144	0.000
7 N TENAGA KERJA PEREMPUAN		0.050	0.045	0.004	NA	12.808	0.063	0.000
8 C KEPEMILIKAN NIB	2	0.027	0.193	0.005	NA	2.274	0.062	0.099
9 C STATUS PENGUSAHA	2	0.007	0.023	0.001	NA	0.673	0.057	0.000
10 N TENAGA KERJA LAKI LAKI		0.038	0.033	0.002	NA	24.863	0.047	0.000
11 N OMZET		0.089	0.057	0.006	NA	30.285	0.034	0.000
12 N TAHUN MULAI BEROPERASI		0.037	0.019	0.002	NA	12.686	0.009	0.032
13 N MODAL PENDIRIAN		0.035	0.018	0.001	NA	14.781	-0.001	0.030
14 C DISABILITAS	1	0.000	0.000	0.000	NA	NA	-0.130	NA

Missing values will be imputed as needed.

Gambar 9.9 Rank untuk target Metoda Pemasaran.

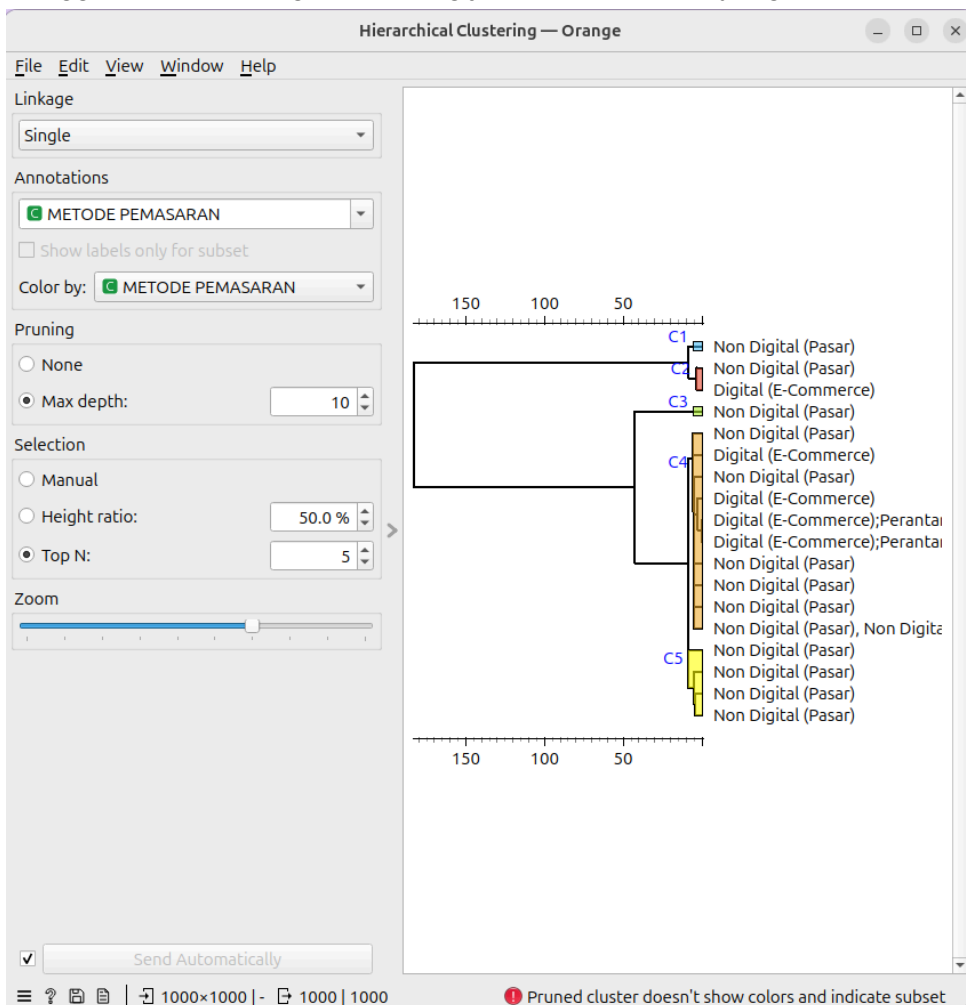
Dalam orange workflows ada dua tool, yaitu "widget Select Columns" dan "widget Rank". Hasil dari pemilihan kolom dari "widget Select Columns" langsung kita masukkan ke "widget Rank". Rank mempunyai banyak method untuk menentukan peringkat, jadi kita dapat memilih method mana yang paling cocok. Misalnya, kita bisa pakai method ReliefF untuk melihat kolom mana yang paling berpengaruh terhadap metoda pemasaran.

Dari scoring method ReliefF, terlihat bahwa yang paling menentukan omzet berurut ke bawah sebagai berikut,

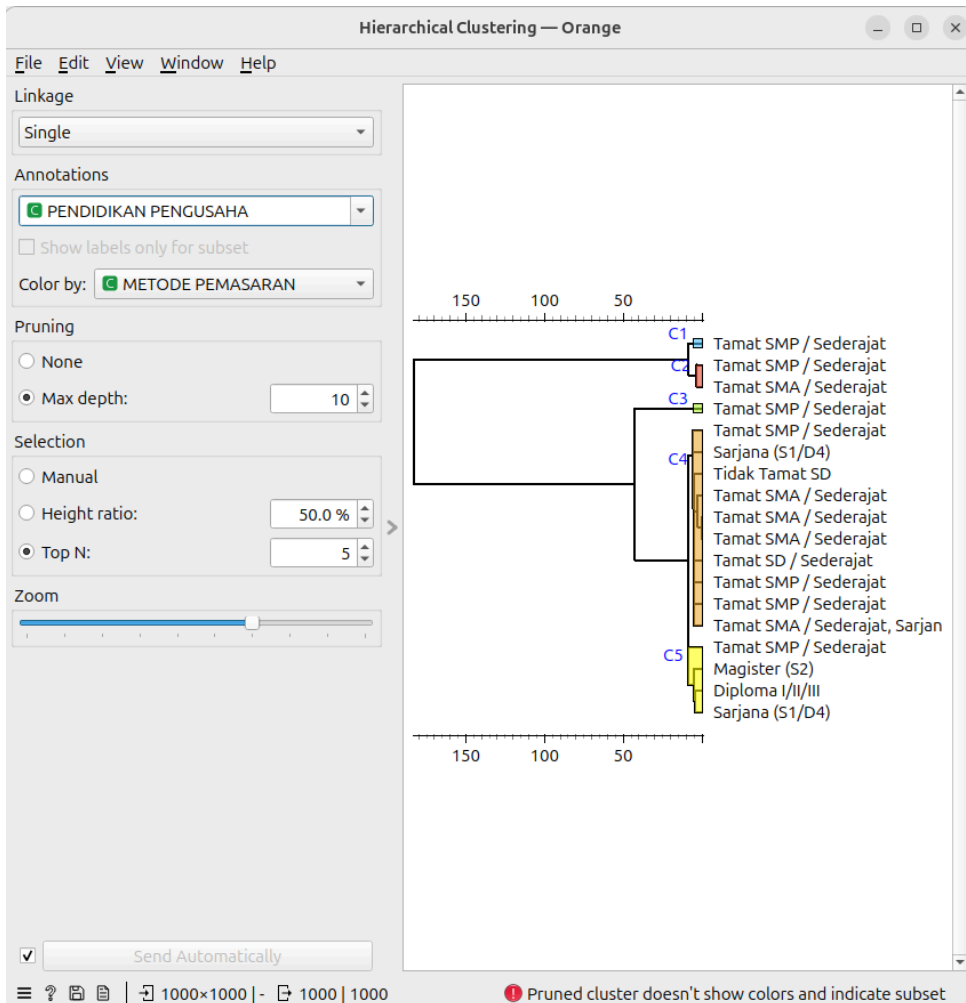
- Kepemilikan Laporan Keuangan
- Pendidikan Pengusaha

- Kategori KBLI
- Jenis Kelamin
- Penggunaan Teknologi

Uniknya, modal usaha dan omzet ternyata tidak terlalu berpengaruh. Yang menarik untuk diamati, ciri khas pengusaha yang menggunakan metoda pemasaran yang baik cenderung akan mempunyai laporan keuangan, tingkat pendidikan pengusaha yang baik dan kemauan untuk menggunakan teknologi di samping jenis usaha (KBLI) yang sesuai.



Gambar 9.10 Hasil Cluster Metoda Pemasaran



Gambar 9.11 Clustering Tingkat Pendidikan

Hasil pemilihan kolom data menggunakan "widget Select Columns" dapat langsung kita gunakan untuk menganalisis kelompok data secara otomatis (unsupervised learning) dengan "widget Distance". Untuk melihat hasil pengelompokan secara visual, kita bisa menghubungkan lagi hasil dari "widget Distance" ke "widget Hierarchical Clustering".

Gambar Hierarchical Clustering yang ditampilkan menunjukkan hasil pengelompokan data secara hierarkis. Kita hanya menampilkan 5 kelompok utama saja untuk mempermudah analisis. Dari gambar ini, terlihat ada tiga kelompok besar,

- Kelompok atas
- Kelompok tengah
- Kelompok bawah

Kelompok data di bagian atas dan bawah dari pembahasan sebelumnya memiliki nilai omzet yang hampir sama, yaitu puluhan juta. Yang menarik dari kelompok ini sebagian besar tidak menggunakan metoda pemasaran digital. Ada sedikit dari kelompok atas yang menggunakan metoda pemasaran digital.

Sedangkan kelompok di tengah memiliki omzet yang biasa-biasa saja, namun bagian ujung atas dari kelompok ini memiliki omzet yang sedikit lebih besar. Dari pembahasan sebelumnya kelompok tengah atas ini merupakan kelompok yang bermodal biasa, tapi berpendidikan, sehingga memiliki omzet yang agak lebih tinggi daripada kelompok tengah lainnya. Unik-nya ternyata kelompok ini juga yang justru aktif menggunakan metoda pemasaran digital.

Jelas temuan yang berhasil keluar dari analisa data ini adalah untuk bisa sukses sebuah usaha mikro mempunyai resep yang sederhana,

- Pendidikan tinggi
- Menggunakan Metoda Pemasaran Digital
- Modal tidak berbeda jauh dengan UMKM lainnya.
- Omzet Insya Allah lebih baik dari usaha mikro setara.

Tentunya kita dapat melakukan eksplorasi lebih lanjut terhadap data koperasi dan UMKM ini untuk memperoleh banyak hal yang menarik.

BAB 10: AI dan Digital Marketing

Kecerdasan Buatan (AI) telah menjadi kunci dalam dunia pemasaran digital yang semakin kompleks. Dengan volume data yang terus membesar, AI memungkinkan bisnis untuk lebih memahami pelanggan mereka dan memberikan penawaran yang lebih personal. Bagi koperasi dan UMKM, AI adalah alat yang sangat berguna untuk bersaing di pasar yang semakin ketat. AI dapat membantu mereka dalam mengotomatiskan tugas-tugas rutin, menganalisis data pelanggan secara mendalam, dan membuat keputusan bisnis yang lebih cerdas. Singkatnya, AI dapat menjadi pembeda antara sukses dan gagal bagi koperasi dan UMKM di era digital.

AI dan pemasaran digital saling melengkapi. AI dapat digunakan dalam berbagai aspek pemasaran digital, mulai dari analisis data pelanggan hingga pembuatan konten yang lebih relevan. Dengan AI, koperasi dan UMKM dapat:

- **Mengotomatiskan tugas:** Misalnya, AI dapat membantu dalam mengelola media sosial, mengirim email marketing, dan bahkan membuat laporan penjualan.
- **Menganalisis data:** AI dapat mengidentifikasi tren pasar, perilaku konsumen, dan preferensi pelanggan.
- **Membuat keputusan yang lebih baik:** Dengan data yang akurat dan analisis yang mendalam, AI dapat membantu koperasi dan UMKM dalam membuat keputusan strategis yang lebih baik.

AI menawarkan banyak manfaat bagi koperasi dan UMKM. Dengan memanfaatkan teknologi ini, mereka dapat meningkatkan efisiensi, efektivitas, dan daya saing bisnis mereka di pasar yang semakin kompetitif.

Penerapan AI dalam Berbagai Aspek Digital Marketing

Personalisasi

- **Rekomendasi Produk/Jasa:**

- **Contoh:** Koperasi pertanian menggunakan AI untuk merekomendasikan pupuk yang paling cocok berdasarkan data tanah dan jenis tanaman anggota.
- **Implementasi:** Menggunakan algoritma rekomendasi seperti Collaborative Filtering untuk menganalisis pola pembelian anggota dan memberikan saran produk yang relevan.
- **Konten yang Disesuaikan:**
 - **Contoh:** UMKM fashion membuat konten blog yang disesuaikan berdasarkan preferensi gaya masing-masing pelanggan.
 - **Implementasi:** Menggunakan Natural Language Processing (NLP) untuk menganalisis data sosial media pelanggan dan menghasilkan konten yang relevan.
- **Pengalaman Pelanggan yang Lebih Baik:**
 - **Contoh:** UMKM makanan menyediakan chatbot yang dapat menjawab pertanyaan pelanggan tentang menu dan nutrisi secara real-time.
 - **Implementasi:** Menggunakan chatbot berbasis AI yang dilatih dengan data FAQ dan informasi produk.

Otomatisasi

- **Chatbot:**
 - **Contoh:** Koperasi simpan pinjam menggunakan chatbot untuk menjawab pertanyaan anggota terkait saldo, transaksi, dan persyaratan pinjaman.
 - **Implementasi:** Menggunakan platform chatbot seperti Dialogflow atau Chatfuel untuk membangun chatbot yang dapat berinteraksi dengan pelanggan secara alami.
- **Pemasaran melalui Email:**
 - **Contoh:** UMKM kerajinan tangan mengirimkan email promosi yang dipersonalisasi berdasarkan riwayat pembelian pelanggan.
 - **Implementasi:** Menggunakan platform email marketing seperti Mailchimp atau HubSpot yang terintegrasi dengan AI untuk segmentasi dan personalisasi email.
- **Manajemen Media Sosial:**

- **Contoh:** Koperasi perikanan menggunakan alat manajemen media sosial berbasis AI untuk menjadwalkan postingan dan memantau interaksi dengan pengikut.
- **Implementasi:** Menggunakan alat seperti Hootsuite atau Buffer yang dilengkapi dengan fitur AI untuk analisis sentimen dan pengoptimalan konten.

Analisis Data

- **Prediksi Tren:**
 - **Contoh:** UMKM kuliner memprediksi tren makanan populer berdasarkan data pencarian dan media sosial.
 - **Implementasi:** Menggunakan alat analisis data seperti Google Analytics dan alat prediksi tren untuk mengidentifikasi peluang pasar baru.
- **Segmentasi Pasar:**
 - **Contoh:** Koperasi konsumen membagi anggota menjadi beberapa segmen berdasarkan demografi dan perilaku pembelian untuk kampanye pemasaran yang lebih efektif.
 - **Implementasi:** Menggunakan alat segmentasi pelanggan untuk mengelompokkan anggota berdasarkan karakteristik yang relevan.
- **Pengukuran Kinerja Kampanye:**
 - **Contoh:** UMKM kecantikan mengukur efektivitas kampanye iklan di media sosial dengan menggunakan metrik seperti tingkat klik-tayang (CTR) dan konversi.
 - **Implementasi:** Menggunakan alat analisis pemasaran untuk melacak kinerja kampanye dan mengoptimalkan strategi.

Periklanan

- **Penargetan Iklan yang Lebih Presisi:**
 - **Contoh:** Koperasi jasa keuangan menargetkan iklan produk keuangan kepada calon anggota yang memiliki profil demografis dan perilaku yang sesuai.
 - **Implementasi:** Menggunakan platform periklanan seperti Google Ads dan Facebook Ads yang memungkinkan penargetan audiens yang sangat spesifik.

- **Optimasi Iklan:**
 - **Contoh:** UMKM elektronik mengoptimalkan iklan berdasarkan kinerja dan anggaran yang tersedia.
 - **Implementasi:** Menggunakan fitur otomatisasi penawaran dalam platform periklanan untuk meningkatkan ROI.
- **Pengembangan Produk:**
 - **Analisis Sentimen Pelanggan:**
 - **Contoh:** UMKM fashion menganalisis sentimen pelanggan terhadap produk baru melalui ulasan online dan media sosial.
 - **Implementasi:** Menggunakan alat analisis sentimen untuk mengukur kepuasan pelanggan dan mengidentifikasi area yang perlu ditingkatkan.
 - **Pengembangan Produk Baru:**
 - **Contoh:** Koperasi pertanian mengembangkan produk baru berdasarkan tren pasar dan permintaan anggota.
 - **Implementasi:** Menggunakan data analisis untuk mengidentifikasi peluang produk baru dan melakukan riset pasar yang lebih efektif.

Contoh Implementasi Kode (Python):

Berikut adalah contoh implementasi penggunaan Sentiment Analysis menggunakan python,

```
# Contoh menggunakan library Natural Language Processing
(NLTK) untuk analisis sentimen
import nltk
from nltk.sentiment.vader import SentimentIntensityAnalyzer

# Contoh teks ulasan produk
review = "Produk ini sangat bagus, kualitasnya terbaik!"

# Inisialisasi SentimentIntensityAnalyzer
sia = SentimentIntensityAnalyzer()

# Menghitung skor sentimen
sentiment_scores = sia.polarity_scores(review)
```

```
print(sentiment_scores)
```

Penerapan AI dalam digital marketing menawarkan banyak manfaat bagi koperasi dan UMKM, termasuk personalisasi yang lebih baik, otomatisasi proses, analisis data yang lebih mendalam, dan periklanan yang lebih efektif. Dengan memanfaatkan teknologi AI, koperasi dan UMKM dapat meningkatkan efisiensi, meningkatkan pengalaman pelanggan, dan mencapai pertumbuhan bisnis yang lebih baik.

Keuntungan dan Tantangan Penggunaan AI dalam Digital Marketing

Keuntungan Penggunaan AI dalam Digital Marketing untuk Koperasi dan UMKM

1. Efisiensi:

- **Otomatisasi Tugas:** AI dapat mengotomatiskan banyak tugas rutin seperti pengelolaan media sosial, email marketing, dan layanan pelanggan. Ini memungkinkan tim Anda untuk fokus pada tugas yang lebih strategis.
- **Analisis Data Cepat:** AI mampu menganalisis data pelanggan dalam jumlah besar dengan sangat cepat, memberikan wawasan yang berharga untuk pengambilan keputusan.
- **Personalisasi Massal:** AI memungkinkan Anda untuk mengirimkan pesan yang sangat personal kepada setiap pelanggan, meningkatkan keterlibatan dan konversi.

Contoh Implementasi:

- **Chatbot:** Gunakan chatbot AI untuk menjawab pertanyaan pelanggan secara otomatis 24/7, mengurangi beban kerja tim layanan pelanggan.
- **Email Marketing:** Manfaatkan AI untuk segmentasi pelanggan dan mengirimkan email yang relevan dengan minat masing-masing.

2. Efektivitas:

- **Prediksi Perilaku Pelanggan:** AI dapat memprediksi perilaku pelanggan di masa depan, memungkinkan Anda untuk merencanakan kampanye pemasaran yang lebih efektif.
- **Optimasi Iklan:** AI dapat mengoptimalkan iklan Anda secara real-time, memastikan anggaran iklan Anda digunakan secara efisien.
- **Peningkatan Konversi:** Dengan personalisasi dan prediksi yang tepat, AI dapat meningkatkan tingkat konversi penjualan Anda.

Contoh Implementasi:

- **Google Ads:** Manfaatkan fitur Smart Bidding di Google Ads untuk mengoptimalkan penawaran iklan secara otomatis.
- **Facebook Ads:** Gunakan Lookalike Audiences untuk menjangkau pelanggan baru yang mirip dengan pelanggan setia Anda.

3. Inovasi:

- **Ide Produk Baru:** AI dapat menganalisis data pasar untuk mengidentifikasi peluang produk baru dan tren yang sedang berkembang.
- **Pengalaman Pelanggan yang Lebih Baik:** AI dapat digunakan untuk menciptakan pengalaman belanja online yang lebih personal dan menarik.

Contoh Implementasi:

- **Rekomendasi Produk:** Gunakan AI untuk merekomendasikan produk kepada pelanggan berdasarkan riwayat pembelian dan minat mereka.
- **Virtual Reality:** Buat pengalaman belanja virtual yang imersif untuk menarik pelanggan.

Tantangan Penggunaan AI dalam Digital Marketing untuk Koperasi dan UMKM

1. Biaya:

- **Investasi Awal:** Implementasi AI membutuhkan investasi awal yang cukup besar, baik untuk perangkat lunak maupun sumber daya manusia.
- **Pemeliharaan:** Biaya pemeliharaan dan pengembangan sistem AI juga perlu diperhitungkan.

2. Privasi Data:

- **Regulasi:** Perusahaan perlu mematuhi peraturan privasi data yang berlaku untuk melindungi data pelanggan.
- **Kepercayaan Pelanggan:** Pelanggaran data dapat merusak kepercayaan pelanggan terhadap merek Anda.

3. Ketergantungan pada Teknologi:

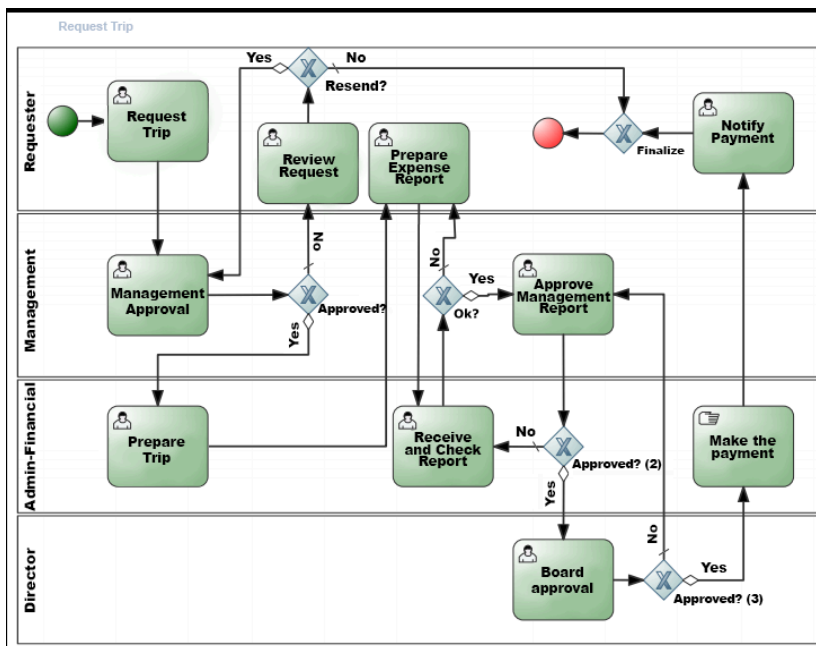
- **Keterbatasan:** AI memiliki keterbatasan dalam memahami konteks yang kompleks dan nuansa bahasa.
- **Risiko Gangguan:** Ketergantungan pada teknologi AI dapat menyebabkan gangguan bisnis jika terjadi masalah teknis.

Tips untuk Mengatasi Tantangan

- **Mulai dengan Skala Kecil:** Implementasikan AI secara bertahap untuk mengurangi risiko dan biaya.
- **Pilih Platform yang Tepat:** Pilih platform AI yang mudah digunakan dan sesuai dengan kebutuhan bisnis Anda.
- **Prioritaskan Keamanan Data:** Lindungi data pelanggan dengan menerapkan langkah-langkah keamanan yang ketat.
- **Berinvestasi dalam Pelatihan:** Latih karyawan Anda untuk menggunakan teknologi AI secara efektif.

BAB 11: Process Mining untuk Transformasi Digital

Proses bisnis adalah serangkaian langkah-langkah yang saling berhubungan dalam sebuah organisasi untuk mencapai tujuan tertentu. Mulai dari mengubah bahan mentah atau informasi menjadi produk atau layanan siap pakai, proses bisnis juga menciptakan nilai bagi pelanggan. Dengan mengidentifikasi, menganalisis, dan memperbaiki proses-proses ini, perusahaan dapat meningkatkan efisiensi, kualitas, dan keunggulan kompetitifnya. Proses bisnis juga berperan penting dalam mengelola sumber daya, mengurangi biaya, dan meningkatkan kepuasan pelanggan. Dalam era digital, teknologi semakin berperan penting dalam mengubah dan meningkatkan kompleksitas proses bisnis. Oleh karena itu, dokumentasi yang baik dan analisis mendalam terhadap proses bisnis yang sedang berjalan (process mining) menjadi kunci untuk melakukan transformasi digital dan mencapai kinerja yang lebih optimal.



Gambar 11.1. Contoh Sebuah Business Process

Transformasi digital dalam birokrasi adalah upaya modernisasi sistem pemerintahan dengan memanfaatkan teknologi informasi. Tujuan utama dari transformasi ini adalah untuk meningkatkan kualitas pelayanan publik, membuat proses pemerintahan lebih efisien, dan memperkuat

akuntabilitas. Dalam praktiknya, transformasi digital melibatkan digitalisasi berbagai layanan pemerintah, seperti perizinan dan pembayaran pajak, serta pemanfaatan data untuk pengambilan keputusan yang lebih baik. Selain itu, transformasi ini juga mendorong perubahan budaya kerja di lingkungan pemerintahan agar lebih fleksibel dan responsif terhadap kebutuhan masyarakat.

Transformasi digital diyakini akan membawa banyak manfaat bagi berbagai pihak, termasuk masyarakat yang akan mendapatkan layanan publik yang lebih mudah dan transparan, serta pemerintah yang akan menjadi lebih efisien dalam menjalankan tugasnya. Namun, proses transformasi ini juga dihadapkan pada sejumlah tantangan, seperti ketersediaan infrastruktur digital yang memadai dan kemampuan sumber daya manusia dalam mengoperasikan teknologi. Meskipun demikian, pemerintah Indonesia telah menunjukkan komitmennya untuk mewujudkan transformasi digital ini. Dengan dukungan dari seluruh lapisan masyarakat, diharapkan transformasi digital dapat berhasil dilaksanakan dan membawa Indonesia menuju pemerintahan yang lebih modern dan melayani.

Peran Process Mining dalam Transformasi Digital

Analisis proses bisnis atau *process mining* menjadi kunci dalam transformasi birokrasi. Dengan *process mining*, kita bisa mendapatkan gambaran jelas tentang bagaimana suatu proses berjalan di dalam pemerintahan. Ini memungkinkan kita untuk menemukan titik-titik yang kurang efisien, seperti perizinan yang berbelit-belit. Pemahaman yang mendalam ini menjadi landasan penting untuk melakukan perbaikan dan perubahan.

Process mining tidak hanya untuk menganalisis, tetapi juga untuk memantau perkembangan transformasi. Seperti seorang dokter yang memantau kondisi pasien, *process mining* membantu kita melihat sejauh mana perubahan yang sudah dilakukan berhasil meningkatkan kinerja. Dengan data yang akurat, kita bisa mengukur efektivitas dari berbagai program reformasi dan memastikan bahwa kebijakan yang diambil sesuai dengan tujuan yang ingin dicapai.

Manfaat *process mining* dalam transformasi birokrasi sangat luas. Selain membantu mengidentifikasi masalah dan memantau kemajuan, *process mining* juga meningkatkan transparansi dan akuntabilitas. Ketika masyarakat bisa melihat secara jelas bagaimana suatu proses berjalan, kepercayaan mereka terhadap pemerintah akan meningkat. Dengan demikian, *process mining* menjadi tool yang sangat berharga untuk mewujudkan birokrasi yang lebih efisien, efektif, dan terpercaya.

Langkah Implementasi Process Mining

Process mining adalah sekumpulan teknik yang menghubungkan bidang ilmu data dan manajemen proses untuk mendukung analisis proses operasional berdasarkan log peristiwa. Tujuan dari process mining adalah untuk mengubah data peristiwa menjadi wawasan dan tindakan. Process mining merupakan bagian integral dari ilmu data, didorong oleh ketersediaan data peristiwa dan keinginan untuk meningkatkan proses. Teknik process mining menggunakan data peristiwa untuk menunjukkan apa yang sebenarnya dilakukan orang, mesin, dan organisasi. Penambangan proses memberikan wawasan baru yang dapat digunakan untuk mengidentifikasi jalur eksekusi yang diambil oleh proses operasional dan mengatasi masalah kinerja dan kepatuhannya. Ada tiga kelas utama teknik process mining:

- penemuan proses,
- pemeriksaan kesesuaian, dan
- peningkatan proses.

Di masa lalu istilah seperti Workflow Mining dan Automated Business Process Discovery (ABPD) digunakan. Ikuti langkah-langkah berikut untuk memulai:

- Persiapan:
 - Kumpulkan data: Kumpulkan data yang relevan dengan tujuan analisis Anda. Pastikan data berkualitas baik, lengkap, dan konsisten. Input untuk process mining adalah log peristiwa. Log peristiwa melihat proses dari sudut tertentu. Setiap peristiwa dalam log harus berisi
 - pengidentifikasi unik untuk instance proses tertentu (disebut id kasus),

- aktivitas (deskripsi peristiwa yang sedang terjadi), dan
 - stempel waktu.
 - Mungkin ada atribut peristiwa tambahan yang mengacu pada sumber daya, biaya, dll., tetapi ini bersifat opsional. Dengan sedikit usaha, data tersebut dapat diambil dari sistem informasi yang mendukung proses operasional. Process mining menggunakan data peristiwa ini untuk menjawab berbagai pertanyaan terkait proses.
 - Pilih tool yang tepat: Beragam tool tersedia, seperti RapidMiner, Weka, dan Knime. Pilihlah yang sesuai dengan kebutuhan dan keahlian Anda.
- **Pengolahan Data:**
 - Pembersihan data: Bersihkan data dari noise, nilai yang hilang, dan inkonsistensi. Gunakan teknik seperti imputation dan data wrangling.
 - Transformasi data: Ubah format data jika diperlukan. Lakukan normalisasi, scaling, dan encoding untuk mempermudah analisis.
 - **Penambahan Data:**
 - Pilih metode: Pilih metode data mining yang sesuai dengan tujuan Anda, seperti klasifikasi, clustering, dan association rule mining.
 - Terapkan metode: Gunakan tool yang dipilih untuk menerapkan metode data mining pada data yang telah diolah.
 - **Interpretasi dan Penerapan:**
 - Evaluasi hasil: Analisis hasil yang diperoleh dan pastikan validitas dan akurasi.
 - Interpretasi: Temukan pola dan wawasan tersembunyi dalam data.
 - Penerapan: Gunakan pengetahuan yang diperoleh untuk pengambilan keputusan yang lebih baik.

Perhatian Penting:

- Pahami tujuan: Pastikan Anda memahami tujuan data mining sebelum memulai proses.
- Kualitas data: Data yang berkualitas merupakan kunci utama keberhasilan data mining.
- Pemilihan metode: Pilihlah metode yang tepat untuk tujuan dan jenis data Anda.
- Interpretasi yang tepat: Interpretasikan hasil dengan hati-hati dan pertimbangkan konteksnya.

Tool dan Teknologi Process Mining

Berikut beberapa tool open source yang populer untuk melakukan process mining:

Apromore

Salah satu tool process mining open source yang paling populer <https://apromore.com/> . Menawarkan berbagai fitur, termasuk:

- Penemuan proses: Secara otomatis dapat menemukan model proses dari data log.
- Analisis konformitas: Membandingkan model proses dengan data log untuk mengidentifikasi penyimpangan.
- Simulasi proses: Mensimulasikan model proses untuk memprediksi kinerja.
- Peningkatan proses: Memberikan rekomendasi untuk meningkatkan proses.
- Mudah digunakan, dengan antarmuka pengguna grafis yang intuitif.
- Tersedia dokumentasi dan komunitas yang luas.
- Kekurangan:
 - Kurang kuat dibandingkan beberapa tool komersial.
 - Dapat lambat pada dataset besar.

PM4Py

Perpustakaan Python untuk process mining

<https://pm4py.fit.fraunhofer.de/> . Menawarkan berbagai fitur, termasuk:

- Penemuan proses: Secara otomatis dapat menemukan model proses dari data log.

- Analisis konformitas: Membandingkan model proses dengan data log untuk mengidentifikasi penyimpangan.
- Simulasi proses: Mensimulasikan model proses untuk memprediksi kinerja.
- Peningkatan proses: Memberikan rekomendasi untuk meningkatkan proses.
- Fleksibel dan dapat disesuaikan.
- Kekurangan:
 - Membutuhkan pengetahuan pemrograman Python untuk digunakan.
 - Tidak memiliki antarmuka pengguna grafis.

ProM

Platform process mining open source yang komprehensif <https://promtools.org/> . Menawarkan berbagai fitur, termasuk:

- Penemuan proses: Secara otomatis dapat menemukan model proses dari data log.
- Analisis konformitas: Membandingkan model proses dengan data log untuk mengidentifikasi penyimpangan.
- Simulasi proses: Mensimulasikan model proses untuk memprediksi kinerja.
- Peningkatan proses: Memberikan rekomendasi untuk meningkatkan proses.
- Penambangan data: Menambang data log untuk informasi tentang proses.
- Kuat dan serbaguna.
- Kekurangan:
 - Kurang user-friendly dibandingkan beberapa tool lainnya.
 - Membutuhkan pengetahuan teknis untuk digunakan.

bupaR

Paket R untuk process mining

<https://cran.r-project.org/web/packages/bupaR/index.html> .

Menawarkan berbagai fitur, termasuk:

- Penemuan proses: Secara otomatis dapat menemukan model proses dari data log.
- Analisis konformitas: Membandingkan model proses dengan data log untuk mengidentifikasi penyimpangan.

- Visualisasi proses: Menampilkan model proses dalam format yang mudah dipahami.
- Mudah digunakan, dengan antarmuka pengguna R.
- Kekurangan:
 - Kurang fitur dibandingkan beberapa tool lainnya.
 - Membutuhkan pengetahuan R untuk digunakan.

Tool terbaik untuk anda akan bergantung pada kebutuhan dan keahlian Anda. Berikut beberapa faktor yang perlu dipertimbangkan:

- Fitur: Fitur apa yang Anda butuhkan? Apakah Anda membutuhkan penemuan proses, analisis konformitas, simulasi proses, peningkatan proses, atau penambahan data?
- Kemudahan penggunaan: Seberapa mudah Anda ingin menggunakan tool ini? Apakah Anda memiliki pengetahuan pemrograman?
- Dukungan komunitas: Seberapa besar komunitas yang mendukung tool ini? Apakah ada dokumentasi dan tutorial yang tersedia?
- Harga: Apakah Anda mencari tool gratis atau berbayar?

Sumber Tambahan

- <https://github.com/topics/process-mining>
- <https://www.processmining.org/>
- <https://towardsdatascience.com/introduction-to-process-mining-5f4ce985b7e5>
- <https://github.com/nhs-pycom/dev-nhs.pycom>

Contoh Implementasi Process Mining

pm4py adalah sebuah pustaka Python yang dirancang untuk mendukung algoritma penambahan proses mutakhir. Pustaka ini bersifat open-source dan dapat digunakan dalam berbagai proyek akademis dan industri. pm4py diciptakan oleh Fraunhofer Institute for Applied Information Technology. Pada kesempatan ini akan diperlihatkan proses instalasi dan penggunaan pm4py.

PM4PY: Penyiapan data

Data / catatan proses yang terjadi menjadi penting untuk dapat melakukan process mining. Sebaiknya kita mempunyai catatan data process dalam jangka panjang untuk memperoleh gambaran utuh bisnis process yang terjadi. Format Data untuk Process Mining dengan pm4py dapat menggunakan beberapa format data untuk event log, yaitu,

Comma Separated Values (CSV)

- Format CSV adalah format data yang paling umum dan mudah digunakan.
- Format ini menggunakan koma untuk memisahkan kolom data.
- Setiap baris dalam file CSV mewakili satu peristiwa dalam event log.
- Contoh format CSV untuk event log:

```
timestamp,case_id,activity,resource
2023-10-04 10:00:00,1,Order Created,user1
2023-10-04 10:15:00,1,Payment Received,user2
2023-10-04 11:00:00,1,Order Shipped,user3
2023-10-04 11:30:00,2,Order Created,user4
2023-10-04 11:45:00,2,Payment Received,user5
2023-10-04 12:00:00,2,Order Shipped,user6
```

Extensible Event Stream (XES)

- XES adalah format data XML yang dirancang khusus untuk event log.
- Format ini lebih kompleks daripada CSV, tetapi menawarkan lebih banyak fleksibilitas dan fitur.
- XES dapat menyimpan informasi tambahan tentang peristiwa, seperti atribut dan data lampiran.
- Contoh format XES untuk event log:

```
<log xes:file="example.xes" xes:version="1.0"
xmlns:xes="http://www.xesformat.org/standard/XES.xsd">
  <trace id="1">
    <event id="e1" timestamp="2023-10-04T10:00:00">
      <data name="case_id" value="1"/>
      <data name="activity" value="Order Created"/>
    </event>
  </trace>
</log>
```

```

    <data name="resource" value="user1"/>
  </event>
  <event id="e2" timestamp="2023-10-04T10:15:00">
    <data name="case_id" value="1"/>
    <data name="activity" value="Payment Received"/>
    <data name="resource" value="user2"/>
  </event>
  <event id="e3" timestamp="2023-10-04T11:00:00">
    <data name="case_id" value="1"/>
    <data name="activity" value="Order Shipped"/>
    <data name="resource" value="user3"/>
  </event>
</trace>
<trace id="2">
  <event id="e4" timestamp="2023-10-04T11:30:00">
    <data name="case_id" value="2"/>
    <data name="activity" value="Order Created"/>
    <data name="resource" value="user4"/>
  </event>
  <event id="e5" timestamp="2023-10-04T11:45:00">
    <data name="case_id" value="2"/>
    <data name="activity" value="Payment Received"/>
    <data name="resource" value="user5"/>
  </event>
  <event id="e6" timestamp="2023-10-04T12:00:00">
    <data name="case_id" value="2"/>
    <data name="activity" value="Order Shipped"/>
    <data name="resource" value="user6"/>
  </event>
</trace>
</log>

```

Object-Centric Event Log (OCEL)

- OCEL adalah format data yang dirancang untuk event log yang berpusat pada objek.
- Format ini digunakan untuk event log yang melibatkan beberapa objek, seperti pelanggan, pesanan, atau produk.

- OCEL dapat menyimpan informasi tentang objek-objek ini, seperti atribut dan hubungan antar objek.

Memilih Format Data:

- Format CSV direkomendasikan untuk event log sederhana dengan sedikit informasi tentang resource / metadata lainnya.

PM4PY: Instalasi

Di Ubuntu, Instalasi pm4py cukup sederhana melalui langkah berikut. Pertama-tama, instalasi aplikasi pendukung,

```
sudo apt update
sudo apt -y install libgl1-mesa-glx libegl1-mesa libxrandr2 \
libxrandr2 libxss1 libxcursor1 libxcomposite1 libasound2 \
libxi6 libxtst6 graphviz python3-tk libblas-dev liblapack-dev \
libsuitesparse-dev gcc g++
```

Sebagai user biasa, bukan root lakukan download dan instal anaconda

```
wget -P /tmp
https://repo.anaconda.com/archive/Anaconda3-2023.09-0-Linux-x86_64.sh
bash /tmp/Anaconda3-2023.09-0-Linux-x86_64.sh
source ~/.bashrc
~/anaconda3/bin/conda config --set auto_activate_base false
~/anaconda3/bin/conda update --all
```

Setelah anaconda tersedia, Instalasi pm4py cukup mudah installnya melalui perintah,

```
sudo apt -y install python3-pip
pip install pm4py
PATH="$PATH:/home/onno/.local/bin"
```

PM4PY: Contoh Source Code

Berikut adalah beberapa contoh source code menggunakan process mining for python PM4Py. Yang penting, pastikan data masukan

menggunakan format yang dimengerti oleh PM4Py. Source yang pertama akan kita lihat adalah sebagai berikut,

```
import pandas as pd
import pm4py

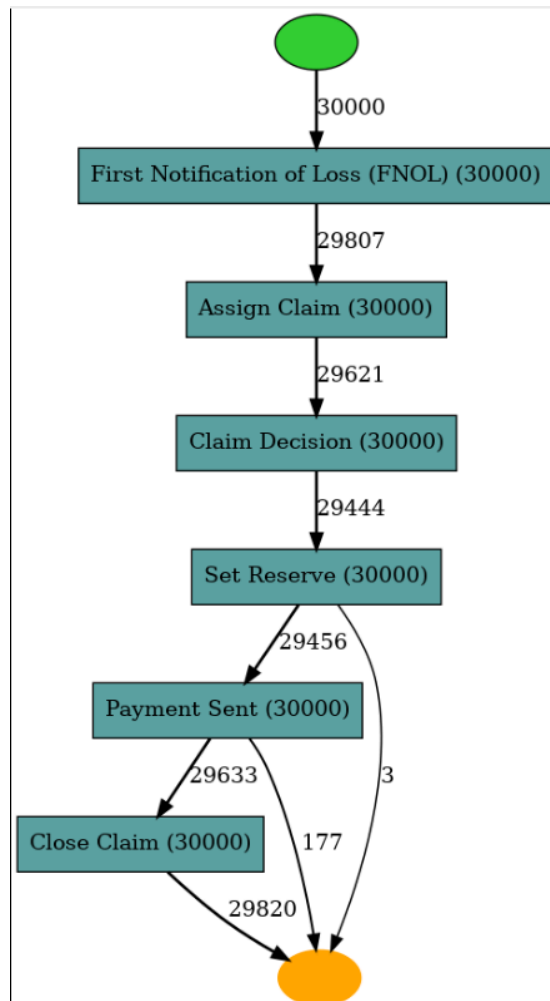
if __name__ == "__main__":
    dataframe =
pd.read_csv('/home/onno/ORANGE/PROCESS-MINING/Insurance
_claims_event_log.csv', sep=',')
    dataframe = pm4py.format_dataframe(dataframe,
case_id='case_id', activity_key='activity_name',
timestamp_key='timestamp')
    event_log = pm4py.convert_to_event_log(dataframe)

    map = pm4py.discover_heuristics_net(event_log)
    pm4py.view_heuristics_net(map)
```

Dalam bahasa manusia, kita dapat membaca source tersebut sebagai berikut,

- File data dalam format .csv dibaca menggunakan read_csv.
- Dalam format .csv tersebut, melalui format dataframe kita beritahukan kolom mana yang menjadi case_id, activity_key dan timestamp_key agar dapat dianalisa dengan baik.
- file .csv yang berhasil di baca di konversikan ke event log, oleh convert_to_event_log. Event log sebetulnya menggunakan format XML yang dikenal dengan format XES.
- Setelah data berhasil di convert ke format event log, maka kita data menganalisa bisnis process. Disini kita menggunakan discover_heuristics_net.
- Terakhir, hasil temuan bisnis process yang ditemukan di gambarkan menggunakan view_heuristics_net.

Untuk contoh data yang kita miliki maka tampilan bisnis process yang berhasil di ekstrak oleh PM4Py adalah seperti gambar di bawah ini.



Gambar 11.2. Hasil Process Mining.

Source code selanjutnya lebih sederhana, disini kita akan menganalisa data dalam format XES. Penggunaan format XES menyebabkan kita tidak perlu melakukan konversi format.

```

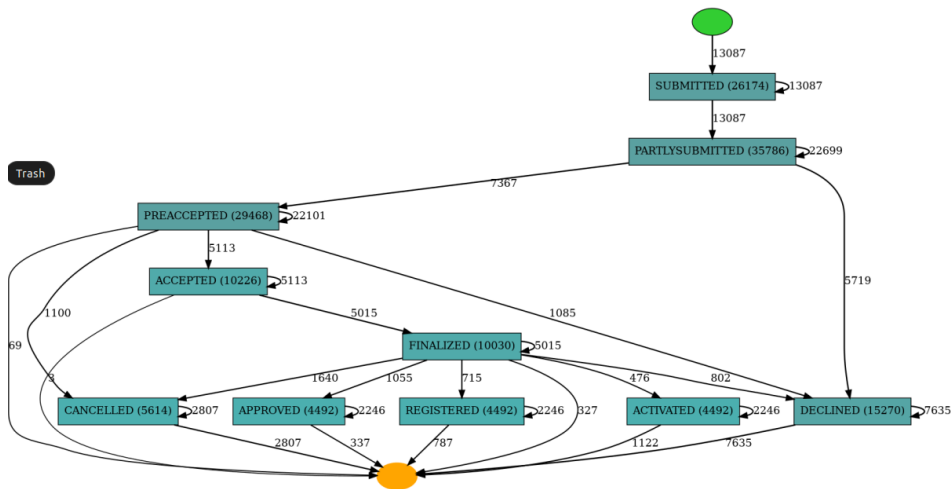
import pm4py
if __name__ == "__main__":
    log =
    pm4py.read_xes('/home/onno/ORANGE/PROCESS-MINING/BPIC
    _2012_A.xes')
  
```

```
map = pm4py.discover_heuristics_net(log)
pm4py.view_heuristics_net(map)
```

Dalam bahasa manusia, kita dapat membaca source tersebut sebagai berikut,

- File data dalam format .xes dibaca menggunakan read_xes.
- Setelah data berhasil di load, maka kita data menganalisa bisnis process. Disini kita menggunakan discover_heuristics_net.
- Terakhir, hasil temuan bisnis process yang ditemukan di gambarkan menggunakan view_heuristics_net.

Untuk contoh data yang kita miliki maka tampilan bisnis process yang berhasil di ekstrak oleh PM4Py adalah seperti gambar di bawah ini.



Gambar 11.3. Hasil Process Mining.

Berbagai hasil process mining, memungkinkan kita mendeteksi beberapa hal penting seperti,

- Dimana saja bypass process terjadi?
- Bagian mana yang menjadikan process terhambat?
- Jika kita menganalisa timestamp mulai dan selesai setiap tahap, maka kita dapat melihat titik mana yang menyebabkan bottleneck.
- dan masih banyak lagi.

Berdasarkan berbagai hasil process mining yang kita lihat, maka kita akan bisa secara intuitif melakukan transformasi birokrasi agar menjadikan bisnis process menjadi lebih baik / lebih lancar.

BAB 12: ChatGPT dan Generative AI

Generative AI, atau kecerdasan buatan generatif, merupakan sub-bidang AI yang berfokus pada pembuatan konten baru seperti teks, gambar, video, atau data lainnya dengan mempelajari pola dan struktur dari data yang diberikan. Model ini dilatih dengan sejumlah besar data, mirip dengan cara manusia belajar dari melihat banyak contoh. Dengan menggunakan model pembelajaran mesin, Generative AI mempelajari pola dan hubungan dalam data untuk kemudian menghasilkan konten baru yang memiliki karakteristik serupa dengan data latihannya.

ChatGPT, Bard, Gemini barangkali merupakan contoh nyata teknologi Generative AI yang sangat viral di dunia saat ini. Platform ini banyak digunakan untuk membantu penulis skripsi, thesis, paper jurnal bahkan berita oleh para wartawan.

Sebagai contoh, model Generative AI yang dilatih dengan gambar kucing, dapat kemudian menghasilkan gambar kucing baru yang realistis meskipun model tersebut belum pernah melihat gambar tersebut sebelumnya. Cara kerja Generative AI bisa dibilang rumit, tetapi pada dasarnya, ia menggunakan model pembelajaran mesin untuk mempelajari pola dan hubungan dalam kumpulan data dan kemudian menggunakan pola yang telah dipelajari untuk membuat konten baru.

Beberapa jenis model Generative AI yang umum digunakan termasuk Jaringan Saraf Tiruan Generatif - Generative Artificial Neuralnetwork (GAN), Model Bahasa Besar - Large Language Models (LLM), dan Autoencoders. GAN terdiri dari dua jaringan saraf: generator yang menghasilkan data baru dan discriminator yang mencoba menentukan apakah data tersebut nyata atau palsu. LLM adalah model saraf yang dilatih dengan jumlah besar data teks dan dapat digunakan untuk menghasilkan teks baru, menerjemahkan bahasa, dan menulis berbagai jenis konten kreatif. Autoencoders adalah jaringan saraf yang dilatih untuk merekonstruksi inputnya, membantu model untuk belajar pola dan fitur yang paling penting dalam data.

Generative AI mengikuti proses dasar yang sama dalam semua jenis modelnya: pelatihan, pembuatan, dan evaluasi. Model dilatih dengan

jumlah besar data, menghasilkan data baru, dan data baru tersebut dievaluasi untuk memastikan kualitasnya. Proses ini diulang hingga model dapat menghasilkan data yang memuaskan. Meskipun masih dalam tahap awal pengembangan, Generative AI memiliki potensi besar untuk menciptakan konten baru dan inovatif, tetapi juga memiliki keterbatasan dan potensi risiko yang harus ditangani dengan bijaksana.

Contoh aplikasi Generative AI mencakup berbagai bidang seperti seni dan desain, sains dan penelitian, bisnis dan industri, serta kehidupan sehari-hari. Dalam seni dan desain, Generative AI dapat digunakan untuk membuat gambar, musik, dan menulis cerita. Dalam sains, ia dapat digunakan untuk menemukan obat baru, membuat bahan baru, dan memprediksi struktur protein. Dalam bisnis, Generative AI dapat digunakan untuk membuat konten pemasaran, mengembangkan produk baru, dan meningkatkan layanan pelanggan. Dalam kehidupan sehari-hari, Generative AI dapat digunakan untuk membuat filter media sosial, game, dan asisten virtual.

Dengan terus berkembangnya teknologi, kita dapat mengharapkan lebih banyak aplikasi inovatif dan kreatif dari Generative AI di masa depan. Model ini akan terus berkembang dan mungkin akan memainkan peran yang semakin penting dalam berbagai aspek kehidupan kita.

Ollama Untuk Generative AI Lokal

Ollama adalah sebuah platform atau inisiatif yang berfokus pada pengembangan kecerdasan buatan generatif (Generative AI) yang disesuaikan untuk kebutuhan lokal. Platform ini berusaha untuk menciptakan solusi AI yang lebih relevan dan efektif dengan memperhatikan karakteristik dan kebutuhan masyarakat setempat. Ollama bertujuan untuk menyediakan tool dan teknologi AI yang dapat diaplikasikan di berbagai sektor, seperti bisnis, pendidikan, seni, budaya, serta sektor-sektor lainnya yang memerlukan personalisasi berdasarkan konteks lokal.

Beberapa fitur dan tujuan utama dari Ollama antara lain:

- **Pemahaman Data Lokal:** Ollama mengutamakan penggunaan data lokal untuk melatih model AI-nya. Dengan demikian, hasil yang

dihasilkan oleh Ollama akan lebih relevan dan kontekstual sesuai dengan budaya, bahasa, dan preferensi masyarakat setempat.

- Inovasi dan Kreativitas: Ollama menyediakan tool bagi para kreator dan inovator lokal untuk mengeksplorasi dan mengembangkan karya mereka menggunakan teknologi AI generatif. Ini termasuk pembuatan gambar, musik, puisi, dan berbagai jenis konten digital lainnya.
- Etika dan Tanggung Jawab: Penggunaan Ollama harus dilakukan dengan penuh tanggung jawab. Penting untuk memastikan bahwa data yang digunakan valid dan tidak melanggar privasi, serta ada pengawasan untuk mencegah penyalahgunaan teknologi ini.

Ollama, dengan pendekatan yang fokus pada lokalitas, berpotensi besar untuk memperkuat ekosistem AI di daerah-daerah tertentu, memberikan dampak positif yang signifikan bagi masyarakat lokal, dan membantu dalam menghadapi tantangan-tantangan unik yang dihadapi oleh komunitas tersebut. Melalui kolaborasi antara berbagai pihak seperti KemenkopUKM, industri, dan komunitas, Ollama dapat menjadi pilar utama dalam pengembangan AI yang inklusif dan bermanfaat bagi semua.

Ollama: Instalasi dan penggunaan yang mudah.

Untuk menjalankan Ollama dengan data lokal, pastikan hal berikut ini,

- ada akses internet untuk download Ollama dan modelnya.
- Ollama bisa di install di desktop, tapi akan lebih nyaman di server.
- Memory Server menjadi kunci, semakin besar model yang digunakan semakin besar memory RAM yang dibutuhkan. Untuk model llama3 7B membutuhkan paling tidak 8GB RAM.
- Penggunaan nvidia GPU tampaknya akan sangat membantu percepatan proses komputasi.

Pertama-tama instal aplikasi pendukung seperti curl dan net-tools menggunakan perintah `sudo su` dan `apt install`.

Kemudian, download Ollama dengan perintah

```
curl -fsSL https://ollama.com/install.sh | sh.
```

Setelah terdownload, jalankan Ollama dan download model yang diinginkan, seperti llama3, dengan perintah

```
ollama serve
ollama pull llama3
```

Bagi anda yang ingin menjalankan ollama dengan model bahasa Indonesia, dapat pull beberapa model bahasa Indonesia yang tersedia di Internet, yaitu,

```
ollama pull bangundwir/bahasa-4b-v2
ollama pull MarcoAland/llama3.1-rag-indo
ollama pull jayadana/jayacode
```

Untuk membuat web interface Ollama, kamu perlu beralih ke mode super user dengan sudo su. Lalu, install Open Web UI menggunakan perintah

```
sudo snap install --beta open-webui.
```

Karena data yang ada di llama3 sebetulnya terbatas, maka hasil yang di generate AI ini tidak presisi, dan terkesan halusinasi. Ollama lebih nyaman menggunakan bahasa Inggris, kita perlu menggunakan model yang berisi bahasa Indonesia jika ingin menghasilkan bahasa Indonesia yang baik. Tampilan menu CLI ollama sebagai berikut,

```
onno@llm:~$ ollama run llama3
>>> siapakah onno w. purbo
A great question!
```

Onno W. Purbo (born 1949) is an Indonesian engineer, scientist, and educator who has made significant contributions to the development of Indonesia's telecommunications industry.

Throughout his career, Purbo has received numerous awards and honors for his contributions to the field of telecommunication engineering in Indonesia. He is widely recognized as one of the most influential figures in Indonesian science and technology.

Ollama akan lebih menarik untuk melakukan paraphrase point-point penting menjadi kalimat yang lebih baik, sebagai berikut,

>>> Lakukan paraphrase tips agar aman berinternet berikut:
... - jangan kepo, semua link di klik tanpa cek terlebih dulu.
... - pikir baik-baik sebelum melakukan posting di media sosial.
... - pastikan semua teman di media sosial anda, betul-betul anda kenal.
... - jangan pernah langsung percaya hal-hal yang terlalu berlebihan, misalnya iklan harga murah.

Here are some paraphrased tips for safe internet use:

****1. Be cautious when clicking links****

Don't click on every link without checking it out first. Make sure the link is legitimate and trustworthy.

****2. Think twice before posting online****

Before sharing something on social media, think carefully about what you're posting and consider its potential consequences.

****3. Verify your connections on social media****

Make sure you really know the people you're connected with on social media. Don't accept friend requests from unknown individuals.

****4. Be skeptical of exaggerated claims****

Don't believe everything you read or hear, especially if it sounds too good to be true (e.g., extremely low prices). Always research and verify information before taking action.

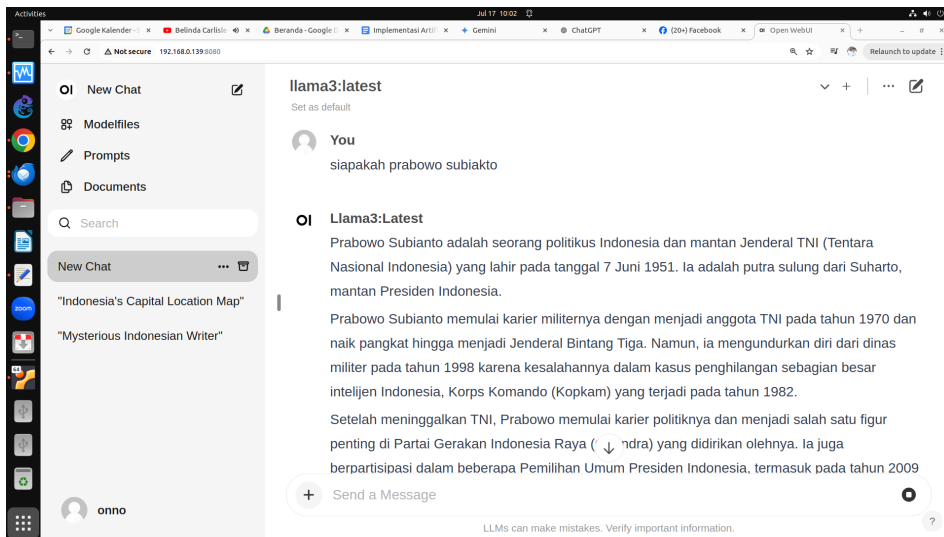
Jika ingin menjalankan ollama dalam bahasa Indonesia lebih disarankan untuk menjalankan perintah berikut,

```
ollama pull bangundwir/bahasa-4b-v2
ollama pull MarcoAland/llama3.1-rag-indo
ollama pull jayadana/jayacode
```

Untuk menjalankan ollama dalam bahasa Indonesia dapat menggunakan salah satu dari perintah berikut,

ollama run bangundwir/bahasa-4b-v2
ollama run MarcoAland/llama3.1-rag-indo
ollama run jayadana/jayacode

Ollama juga dapat diakses melalui web jika menjalankan open-webui yang bisa di akses pada port 8080 sebagai berikut,



Gambar 12.1. Tampilan open-WebUI berinteraksi dengan Ollama.

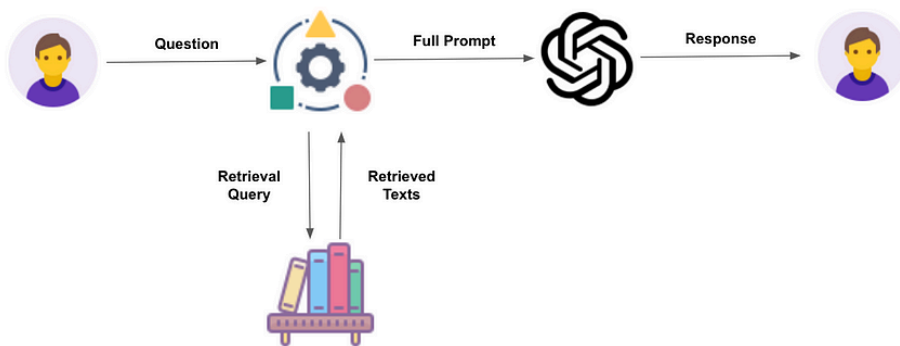
Untuk informasi lebih lengkap mengenai instalasi Ollama, bisa mengunjungi situs web resmi Ollama: <https://ollama.com/>. Tutorial dan panduan lainnya di komunitas Ollama, seperti di GitHub: <https://github.com/ollama/ollama>.

Retrieval Augmented Generation (RAG) memperkaya Ollama lokal

RAG (Retrieval-Augmented Generation) menarik untuk Ollama dengan data lokal karena berbagai alasan. Pertama, privasi ditingkatkan karena LLM (Large Language Model) dapat dijalankan secara lokal, memungkinkan pengguna mengontrol data yang diakses dan diproses, serta menyimpan data sensitif secara aman. Kedua, keamanan ditingkatkan dengan menjalankan LLM secara lokal, mengurangi risiko

kebocoran data dan memungkinkan verifikasi sumber informasi. Ketiga, pengguna memiliki kontrol penuh atas LLM dan data, memungkinkan penyesuaian model sesuai kebutuhan spesifik dan memastikan penggunaan data yang disetujui. Selain itu, biaya operasional dapat ditekan dengan mengurangi ketergantungan pada komputasi awan, dan kinerja ditingkatkan melalui pengurangan latensi. Secara keseluruhan, RAG memberikan manfaat signifikan dalam privasi, keamanan, kontrol, biaya, dan kecepatan.

Gambaran umum cara kerja RAG dapat dilihat pada gambar berikut,



Gambar 11.2. Cara kerja Retrieval-Augmented Generation (RAG).

Dalam bahasa yang sederhana, RAG akan mengubah prompt dari user yang dimasukan ke LLM. Prompt tersebut diisi informasi / konteks yang lebih baik agar responds LLM ke user menjadi lebih baik. Konteks tersebut diperoleh dari masukan file PDF, txt dll ke RAG dari admin sistem.

Contoh sebuah kantor / kementerian tentunya banyak sekali mempunya aturan, keputusan yang harus dipenuhi. Informasi ini biasanya tidak tersedia dalam LLM, oleh karena itu agar LLM dapat memberikan respon dengan baik maka pertanyaan / prompt dari user dimasukan dulu ke RAG Untuk diberikan konteks yang lebih baik agar respon LLM menjadi lebih baik.

Beberapa referensi menarik tentang RAG dapat dibaca melalui referensi berikut,

- Membangun Sistem RAG Lokal untuk Pelestarian Privasi dengan Ollama dan Weaviate:
<https://weaviate.io/developers/weaviate/starter-guides/generative>
- Tutorial RAG Lokal Mudah (Ollama) + Kode Lengkap: YouTube
- Membangun Aplikasi RAG Menggunakan LLM yang Berjalan di Komputer Lokal dengan Ollama dan Langchain:
<https://medium.com/@mohsinrashid64/rag-with-ollama-3a456147610e>
- Mempertahankan Privasi: Membangun Agen RAG 100% Lokal Menggunakan Ollama:
<https://medium.com/@muzammila784/preserve-privacy-building-100-local-rag-agent-using-ollama-b0b66c212f40>
- RAG Similarity Search:
<https://www.youtube.com/watch?v=whu7SHx5cko>

BAB 13: Sepintas Tensorflow untuk Neural Network

TensorFlow adalah library open-source yang populer di bidang machine learning, terkenal karena kemampuannya menangani dataset besar dan skalabilitas tinggi. Dengan fitur komputasi terdistribusi, TensorFlow memungkinkan pelatihan model kompleks menggunakan banyak GPU atau mesin paralel, mempercepat proses pelatihan. Fleksibilitasnya tercermin dalam penggunaan grafik komputasi untuk membangun model yang sangat kustom, mendukung berbagai jenis model termasuk deep learning. Selain itu, dukungannya terhadap beberapa bahasa pemrograman seperti Python, C++, dan Java membuat TensorFlow mudah diintegrasikan dengan tool lain dalam ekosistem data science.

Didukung oleh komunitas besar dan pengembangan aktif dari Google, TensorFlow menawarkan ekosistem yang kaya dengan model pre-trained dan integrasi dengan tool populer seperti Pandas dan Keras. Untuk menangani data besar, TensorFlow menyediakan fitur seperti TensorFlow Data dan TensorFlow Extended (TFX), yang membantu mengelola dataset secara efisien dan membangun pipeline machine learning skala besar. Kemampuannya dalam mendistribusikan komputasi dan memori mengoptimalkan penggunaan sumber daya dan mempercepat waktu pelatihan, menjadikannya pilihan utama untuk aplikasi seperti pemrosesan bahasa alami (NLP), pengenalan gambar, dan sistem rekomendasi. Untuk belajar lebih lanjut, pengguna dapat mengakses berbagai tutorial di situs resmi TensorFlow <https://www.tensorflow.org/tutorials/>

Sepintas Machine Learning dengan Neural Network

Neural Network adalah model komputasi yang terinspirasi oleh struktur dan fungsi otak manusia. Model ini terdiri dari banyak unit sederhana yang saling terhubung, disebut neuron. Neuron-neuron ini bekerja bersama untuk memproses informasi dan membuat keputusan.

A. Proses Pembelajaran Neural Network

- **Backpropagation: Algoritma Inti Pembelajaran**

- Tujuan: Menyesuaikan bobot dan bias dalam jaringan neural agar dapat meminimalkan kesalahan prediksi.
- Cara Kerja:
 - Hitung error antara prediksi jaringan dengan nilai sebenarnya.
 - Sebarkan error ini kembali ke lapisan-lapisan sebelumnya secara berulang.
 - Sesuaikan bobot dan bias setiap neuron secara proporsional terhadap kontribusinya terhadap error.
- Visualisasi: Bayangkan Anda sedang mencoba menebak angka yang tersembunyi. Backpropagation seperti proses Anda menyesuaikan tebakan Anda setelah mengetahui jawaban yang benar.
- **Optimasi: Menemukan Bobot Terbaik**
 - Tujuan: Menemukan kombinasi bobot dan bias yang menghasilkan kinerja terbaik.
 - Algoritma Optimasi:
 - Gradient Descent: Algoritma paling dasar, bergerak searah gradien penurunan error.
 - Stochastic Gradient Descent: Varian dari Gradient Descent yang menggunakan sampel data acak untuk mempercepat proses.
 - Adam, RMSprop, Adagrad: Algoritma optimasi adaptif yang menyesuaikan learning rate secara otomatis.
 - Visualisasi: Bayangkan Anda sedang mencari titik terendah dalam sebuah lembah. Optimasi adalah proses Anda mencari jalan menuju titik terendah itu.

B. Metrik Evaluasi Kinerja Neural Network

- Regresi:

- Mean Squared Error (MSE): Rata-rata kuadrat selisih antara nilai prediksi dan nilai sebenarnya.
- Root Mean Squared Error (RMSE): Akar kuadrat dari MSE.
- Mean Absolute Error (MAE): Rata-rata absolut selisih antara nilai prediksi dan nilai sebenarnya.
- Klasifikasi:
 - Accuracy: Persentase prediksi yang benar.
 - Precision: Proporsi prediksi positif yang benar.
 - Recall: Proporsi data positif yang berhasil diprediksi sebagai positif.
 - F1-Score: Harmonic mean dari precision dan recall.
 - Confusion Matrix: Tabel yang menunjukkan jumlah prediksi benar dan salah untuk setiap kelas.

C. Overfitting dan Regularisasi

- Overfitting: Kondisi di mana model terlalu kompleks dan menghafal data pelatihan, sehingga performanya buruk pada data uji.
- Regularisasi: Teknik untuk mencegah overfitting dengan menambahkan penalti pada kompleksitas model.
- Jenis Regularisasi:
 - L1 Regularization (Lasso): Menambahkan penalti pada nilai absolut bobot.
 - L2 Regularization (Ridge): Menambahkan penalti pada kuadrat bobot.
- Dropout: Secara acak menonaktifkan neuron selama pelatihan.
- Early Stopping: Menghentikan pelatihan sebelum model terlalu overfit.

Contoh Penerapan Neural Network:

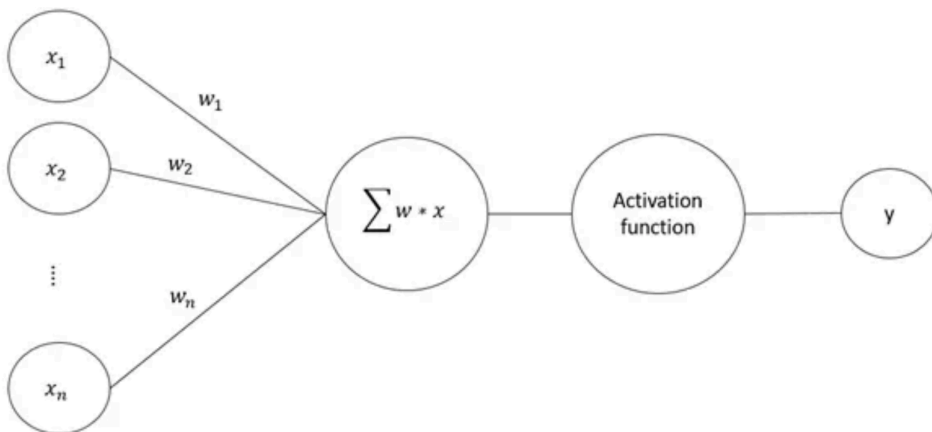
- Pengenalan Gambar: Mengidentifikasi objek dalam gambar (misalnya, wajah, mobil).
- Pengolahan Bahasa Alami: Terjemahan mesin, analisis sentimen, chatbot.
- Prediksi Deret Waktu: Memprediksi harga saham, cuaca.
- Rekomendasi Sistem: Menyarankan produk atau konten kepada pengguna.

Neural Network adalah tool yang sangat kuat untuk menyelesaikan berbagai masalah kompleks. Dengan memahami dasar-dasar proses

pembelajaran, metrik evaluasi, dan teknik mengatasi overfitting, kita dapat membangun model neural network yang efektif.

Neural Network sebagai Ekspansi Regresi Linier

Pada pandangan pertama, neural network mungkin tampak sangat berbeda dari regresi linier. Namun, jika kita melihat lebih dekat, kita akan menemukan bahwa neural network sebenarnya adalah generalisasi atau ekspansi dari regresi linier.



Gambar 13.1. Neuron Tunggal sebagai regresi linier.

Neuron Tunggal sebagai Regresi Linier:

- **Bobot dan Bias:** Neuron tunggal dalam neural network melakukan operasi dasar yang sama dengan regresi linier, yaitu mengalikan fitur input dengan bobot dan menambahkan bias. Hasilnya kemudian diaktivasi menggunakan fungsi aktivasi.
- **Fungsi Aktivasi:** Fungsi aktivasi ini bisa dianggap sebagai transformasi non-linear yang membuat model lebih fleksibel dibandingkan regresi linier murni.

Lapisan Tersembunyi:

- **Kombinasi Linier:** Lapisan tersembunyi dalam neural network melakukan kombinasi linier dari output neuron pada lapisan sebelumnya. Ini mirip dengan regresi linier, tetapi dilakukan secara berulang-ulang.

- Non-Linearitas: Fungsi aktivasi pada setiap neuron di lapisan tersembunyi memungkinkan model untuk mempelajari pola yang lebih kompleks dan non-linear.

Multiple Layers:

- Hierarki Fitur: Dengan menambahkan lebih banyak lapisan, neural network dapat belajar representasi fitur yang semakin abstrak dan kompleks. Lapisan pertama mungkin belajar fitur sederhana, sedangkan lapisan terakhir belajar fitur yang lebih kompleks.
- Fleksibilitas: Arsitektur multi-layer ini memberikan neural network kemampuan untuk memodelkan fungsi yang sangat kompleks, jauh melampaui kemampuan regresi linier.

Intinya:

- Neuron tunggal adalah unit dasar dalam neural network yang melakukan operasi yang mirip dengan regresi linier.
- Lapisan tersembunyi memungkinkan neural network untuk mempelajari pola yang lebih kompleks dan non-linear.
- Multiple layers memberikan neural network fleksibilitas yang jauh lebih besar dibandingkan regresi linier.

Analogi Sederhana:

Bayangkan regresi linier sebagai sebuah rumah kecil dengan satu ruangan. Kita hanya bisa mengatur furnitur dalam satu ruangan itu. Sedangkan neural network adalah sebuah istana dengan banyak ruangan dan lantai. Kita memiliki kebebasan untuk mengatur furnitur di setiap ruangan dan lantai sesuai dengan kebutuhan kita.

Neural network dapat dianggap sebagai perluasan dari regresi linier. Dengan menambahkan non-linearitas, lapisan tersembunyi, dan arsitektur yang lebih kompleks, neural network mampu menyelesaikan masalah yang jauh lebih kompleks dibandingkan regresi linier.

Contoh Neural Network Expansi dari Regresi Linier dalam Python

Konsep Dasar:

- **Regresi Linier:** Model sederhana yang memodelkan hubungan linear antara variabel independen dan dependen.
- **Neural Network (satu lapisan):** Dapat dianggap sebagai regresi linier yang diikuti oleh fungsi aktivasi. Fungsi aktivasi ini memungkinkan model untuk menangkap hubungan non-linear.

Contoh Kode:

```
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.datasets import make_regression
import tensorflow as tf

# Generate sample data
X, y = make_regression(n_samples=100, n_features=1, noise=20,
random_state=42)

# Split data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)

# Create a simple neural network model
with one layer
model = tf.keras.Sequential([
    tf.keras.layers.Dense(units=1, input_shape=[1])
])

# Compile the model
model.compile(optimizer='sgd', loss='mean_squared_error')

# Train the model
model.fit(X_train, y_train, epochs=100,
verbose=0)

# Evaluate the model
test_loss = model.evaluate(X_test, y_test)
print('Test loss:', test_loss)
```

Penjelasan Kode:

- **Import Library:** Mengimport library yang diperlukan, yaitu NumPy untuk operasi numerik, scikit-learn untuk pembagian data, dan TensorFlow untuk membangun model neural network.
- **Generate Data:** Membuat data sampel dengan satu fitur independen dan satu fitur dependen.
- **Split Data:** Membagi data menjadi data latih (training data) dan data uji (test data).
- **Buat Model:** Membuat model neural network sederhana dengan satu lapisan. Lapisan ini memiliki satu neuron dengan satu input.
- **Kompilasi Model:** Mengkompilasi model dengan menggunakan optimizer Stochastic Gradient Descent (SGD) dan loss function Mean Squared Error (MSE).
- **Latih Model:** Melatih model dengan data latih (training data) selama 100 epoch.
- **Evaluasi Model:** Mengevaluasi model dengan data uji dan mencetak nilai loss.

Analisis:

Model neural network ini sebenarnya sangat mirip dengan model regresi linier sederhana. Perbedaan utama adalah penggunaan fungsi aktivasi (yang dalam kasus ini adalah fungsi linear default dari layer Dense).

Jika kita menghapus fungsi aktivasi atau menggunakan fungsi aktivasi linear, maka model ini akan menjadi persis sama dengan model regresi linier.

Ekspansi ke Model yang Lebih Kompleks:

- **Tambahkan Lapisan Tersembunyi:** Dengan menambahkan lapisan tersembunyi, model dapat belajar representasi fitur yang lebih kompleks dan non-linear.
- **Gunakan Fungsi Aktivasi Non-Linear:** Fungsi aktivasi seperti ReLU, tanh, atau sigmoid memungkinkan model untuk menangkap hubungan non-linear yang lebih kompleks.

- **Tingkatkan Jumlah Neuron:** Meningkatkan jumlah neuron pada setiap lapisan dapat meningkatkan kapasitas model untuk belajar pola yang lebih kompleks.

Contoh kode di atas menunjukkan bahwa neural network dengan satu lapisan sebenarnya adalah ekspansi dari regresi linier. Dengan menambahkan lapisan tersembunyi dan menggunakan fungsi aktivasi non-linear, kita dapat membangun model neural network yang jauh lebih kompleks dan powerful.

Pengantar Tensorflow untuk Pemula

Bagi para pemula yang ingin mencoba menjalankan tensorflow, berikut adalah intro penggunaan tensorflow Keras untuk:

- Load prebuilt dataset.
- Membangun model machine learning neural network yang mengklasifikasikan image.
- Melatih neural network ini.
- Mengevaluasi akurasi model.

Tutorial ini sebetulnya dirancang untuk dijalankan di notebook Google Collab. Program Python dijalankan langsung di browser—cara yang bagus untuk belajar dan menggunakan TensorFlow.

Di Collab, sambungkan ke runtime Python: Pada bagian kanan atas menu bar, pilih CONNECT. Untuk menjalankan semua kode dalam notebook, pilih Runtime > Run all. Untuk menjalankan code cell satu per satu, arahkan cursor ke setiap sel dan pilih ikon Run cell.

Set up TensorFlow

Import TensorFlow ke program untuk memulai:

```
import tensorflow as tf
print("TensorFlow version:", tf.__version__)
```

Hasilnya,

```
TensorFlow version: 2.17.0
```

Jika kita ingin menjalankan di lingkungan development sendiri, tidak menggunakan Google collab, sebaiknya membaca panduan penginstalan untuk menyiapkan TensorFlow for development.

Catatan: Pastikan kita telah memutakhirkan ke pip terbaru untuk menginstal paket TensorFlow 2 jika kita menggunakan lingkungan development kita sendiri. Lihat panduan penginstalan untuk detailnya.

Load dataset

Load dan siapkan MNIST dataset. Nilai piksel gambar berkisar dari 0 hingga 255. Skalikan nilai ke rentang 0 hingga 1 dengan membagi nilai dengan 255.0. Ini juga mengubah data sampel dari bilangan bulat menjadi bilangan floating point.

```
mnist = tf.keras.datasets.mnist

(x_train, y_train), (x_test, y_test) = mnist.load_data()
x_train, x_test = x_train / 255.0, x_test / 255.0
```

Built machine learning model

Membuat sebuah `tf.keras.Sequential` model:

```
model = tf.keras.models.Sequential([
    tf.keras.layers.Flatten(input_shape=(28, 28)),
    tf.keras.layers.Dense(128, activation='relu'),
    tf.keras.layers.Dropout(0.2),
    tf.keras.layers.Dense(10)
])
```

Kode di atas mendefinisikan sebuah model neural network sederhana menggunakan TensorFlow Keras. Mari kita breakdown setiap bagian:

```
model = tf.keras.models.Sequential([
```

Ini memulai definisi sebuah model sequential, artinya lapisan-lapisan dalam model akan disusun secara berurutan.

```
tf.keras.layers.Flatten(input_shape=(28, 28)),
```

Lapisan pertama ini mengubah data input menjadi bentuk satu dimensi. Dalam kasus ini, input adalah gambar dengan ukuran 28x28 piksel. Lapisan ini meratakan gambar menjadi sebuah vektor dengan panjang 784.

```
tf.keras.layers.Dense(128, activation='relu'),
```

Ini adalah lapisan terkoneksi penuh (dense layer) pertama dengan 128 neuron. Fungsi aktivasi 'relu' (Rectified Linear Unit) digunakan untuk menambahkan non-linearitas ke dalam model.

```
tf.keras.layers.Dropout(0.2),
```

Lapisan dropout ini secara acak menonaktifkan 20% neuron selama pelatihan untuk mencegah overfitting.

```
tf.keras.layers.Dense(10)
```

Lapisan output dengan 10 neuron, sesuai dengan jumlah kelas dalam dataset MNIST (angka dari 0 hingga 9). Tidak ada fungsi aktivasi yang ditentukan di sini karena lapisan output biasanya menggunakan softmax untuk klasifikasi multi-kelas.

Untuk setiap data, model menghasilkan sebuah daftar nilai mentah (logits). Daftar ini berisi prediksi awal tanpa normalisasi. Nilai-nilai ini biasanya diproses lebih lanjut untuk menghasilkan probabilitas. Jika model digunakan untuk memilih satu dari beberapa pilihan (klasifikasi multi-kelas), nilai-nilai mentah ini biasanya diberikan ke fungsi softmax. Fungsi ini mengubah nilai-nilai mentah menjadi probabilitas untuk setiap pilihan, sehingga jumlah total probabilitas adalah 1. Nilai-nilai mentah ini juga bisa disebut sebagai log-odds, yaitu logaritma dari peluang suatu kejadian. Contoh coding,

```
predictions = model(x_train[:1]).numpy()  
predictions
```

```
array([[ 0.8345982 ,  0.02583528,  0.02325563, -0.13725212,
        -0.4650353 , -0.04670947, -0.8046191 , -0.91970056,
        -0.60336345,  0.06968632]], dtype=float32)
```

Fungsi `tf.nn.softmax` mengubah logits ke probability dari setiap class:

```
tf.nn.softmax(predictions).numpy()
```

```
array([[0.24845515, 0.1106641 , 0.110379 , 0.09401102,
        0.06773675, 0.10292028, 0.04823308, 0.04298984, 0.05898604,
        0.1156248 ]], dtype=float32)
```

Catatan: Dimungkinkan untuk memasukkan fungsi `tf.nn.softmax` ke dalam fungsi aktivasi lapisan terakhir jaringan. Meskipun ini dapat membuat output model lebih mudah dipahami secara langsung, pendekatan ini tidak disarankan karena tidak mungkin memberikan perhitungan loss yang tepat dan stabil secara numerik untuk semua model saat menggunakan output softmax.

Loss Function

Loss function menerima daftar nilai sebenarnya (ground truth) dan daftar nilai prediksi mentah (logits) lalu menghasilkan satu nilai loss untuk setiap data. Nilai kehilangan ini sama dengan negatif dari logaritma probabilitas kelas yang benar. Nilai kehilangan ini akan nol jika model sangat yakin dengan kelas yang benar.

Kita dapat mendefinisikan loss function untuk training menggunakan, misalnya, `losses.SparseCategoricalCrossentropy`:

```
loss_fn =
tf.keras.losses.SparseCategoricalCrossentropy(from_logits=True)
```

Untuk model yang belum di training akan memberikan probabilitas random (1/10 untuk setiap kelas), sehingga loss awal akan mendekati $-\text{tf.math.log}(1/10) \approx 2.3$.

```
loss_fn(y_train[:1], predictions).numpy()
```

2.2738006

Sebelum kita mulai training, konfigurasi dan kompilasi model menggunakan `Keras Model.compile`. Atur kelas pengoptimal ke `adam`, atur loss ke fungsi `loss_fn` yang telah kita definisikan sebelumnya, dan tentukan metrik yang akan dievaluasi untuk model dengan mengatur parameter `metrics` ke `accuracy`.

```
model.compile(optimizer='adam',  
              loss=loss_fn,  
              metrics=['accuracy'])
```

Train & evaluate model

Gunakan metode `Model.fit` untuk menyesuaikan parameter model Anda dan meminimalkan loss:

```
model.fit(x_train, y_train, epochs=5)
```

Metode `Model.evaluate` memeriksa kinerja model, biasanya pada set validasi atau set pengujian.

```
model.evaluate(x_test, y_test, verbose=2)
```

Hasilnya,

```
313/313 - 1s - 3ms/step - accuracy: 0.9765 - loss: 0.0736  
[0.0735689178109169, 0.9764999747276306]
```

Classifier gambar sekarang telah dilatih dengan akurasi ~98% pada dataset ini. Untuk mempelajari lebih lanjut, baca tutorial TensorFlow.

Jika kita ingin model mengembalikan probabilitas, kita dapat membungkus model yang telah dilatih dan menambahkan softmax padanya:

```
probability_model = tf.keras.Sequential([  
    model,  
    tf.keras.layers.Softmax()  
])
```

```
probability_model(x_test[:5])
```

Selamat! Anda berhasil melatih model pembelajaran mesin menggunakan dataset yang sudah tersedia dengan Keras API. Masih banyak hal menarik yang bisa Anda pelajari! Mulai dari memuat dan mempersiapkan data dengan tutorial pemuatan data gambar atau data CSV, hingga membangun model yang lebih kompleks dengan Keras. Ayo, lanjutkan eksplorasi dan perluas pengetahuan Anda!

BAGIAN 3: TANTANGAN DAN RISIKO IMPLEMENTASI AI, MACHINE LEARNING DAN BIG DATA

BAB 14: Etika dan Privasi Data

Mari kita bahas secara mendalam tentang Etika dan Privasi Data, dengan fokus pada tiga (3) aspek penting, yaitu:

Perlindungan, Keadilan, dan Tanggung Jawab

Perlindungan Data Pribadi dan Hak Individu

Di era digital saat ini, data pribadi telah menjadi komoditas yang sangat berharga. Setiap individu memiliki hak atas privasi dan keamanan data pribadinya. Prinsip-prinsip dasar perlindungan data meliputi:

- **Kesepakatan (Consent):** Pengumpulan dan penggunaan data pribadi harus didasarkan pada persetujuan yang jelas dan informatif dari individu yang bersangkutan.
- **Tujuan yang Jelas:** Data pribadi hanya boleh dikumpulkan dan digunakan untuk tujuan yang spesifik dan sah.
- **Akurasi:** Data pribadi harus akurat, lengkap, dan selalu diperbarui.
- **Keamanan:** Data pribadi harus dilindungi dari akses, penggunaan, pengungkapan, atau perubahan yang tidak sah.
- **Transparansi:** Individu harus memiliki akses untuk mengetahui data pribadi mereka yang dimiliki oleh pihak lain, serta bagaimana data tersebut digunakan.
- **Akuntabilitas:** Pengendali data bertanggung jawab atas segala bentuk pemrosesan data pribadi dan harus dapat membuktikan bahwa mereka telah memenuhi kewajiban perlindungan data.

Algoritma yang Adil dan Transparan

Algoritma semakin banyak digunakan dalam berbagai aspek kehidupan, mulai dari rekomendasi produk hingga pengambilan keputusan yang

berdampak signifikan pada kehidupan individu. Agar algoritma dapat digunakan secara etis, perlu dipenuhi beberapa kriteria:

- **Keadilan:** Algoritma harus dirancang untuk menghindari bias yang dapat merugikan kelompok tertentu.
- **Transparansi:** Mekanisme kerja algoritma harus dapat dipahami, sehingga memungkinkan evaluasi dan identifikasi potensi bias.
- **Akuntabilitas:** Pengembang dan pengguna algoritma harus bertanggung jawab atas dampak negatif yang ditimbulkan oleh algoritma tersebut.

Akuntabilitas dan Tanggung Jawab

Akuntabilitas dan tanggung jawab adalah kunci dalam memastikan bahwa data pribadi digunakan secara etis. Semua pihak yang terlibat dalam pengelolaan data, mulai dari pengumpul data hingga pengguna data, harus bertanggung jawab atas tindakan mereka.

- **Pengendali Data:** Perusahaan atau organisasi yang mengumpulkan dan mengelola data pribadi bertanggung jawab untuk memastikan bahwa data tersebut diproses sesuai dengan peraturan yang berlaku dan prinsip-prinsip etika.
- **Pemroses Data:** Pihak ketiga yang mengolah data atas nama pengendali data juga bertanggung jawab atas keamanan dan kerahasiaan data. Pengendali data sebaiknya harus bisa mengaudit dan mengakses apa yang dilakukan oleh pemroses data pihak ketiga.
- **Individu:** Individu juga memiliki tanggung jawab untuk melindungi data pribadi mereka dengan cara yang bijaksana, seperti tidak memberikan informasi pribadi secara sembarangan.

Implikasi dan Tantangan

Masalah etika dan privasi data memiliki implikasi yang luas, termasuk:

- **Perlindungan Konsumen:** Konsumen perlu dilindungi dari penyalahgunaan data pribadi yang dapat merugikan mereka secara finansial atau reputasi.

- **Keadilan Sosial:** Algoritma yang bias dapat memperkuat ketidaksetaraan sosial.
- **Inovasi:** Peraturan yang terlalu ketat dapat menghambat inovasi dalam bidang teknologi informasi.

Etika dan privasi data adalah isu yang kompleks dan terus berkembang. Untuk memastikan bahwa data pribadi digunakan secara etis, diperlukan kolaborasi antara Pemerintah, industri, akademisi, dan masyarakat sipil. Dengan demikian, kita dapat menciptakan lingkungan digital yang aman, adil, dan bermanfaat bagi semua.

BAB 15: Kesenjangan Digital dan Keterampilan

Kesenjangan digital di Indonesia, terutama dalam konteks kecerdasan buatan (AI), menjadi tantangan serius. Ketidakmerataan akses terhadap teknologi dan infrastruktur, khususnya di daerah pedesaan, memperlebar jurang antara mereka yang dapat memanfaatkan AI dan yang tidak. Keterbatasan perangkat digital dan biaya akses yang tinggi semakin memperparah situasi ini.

Pemerintah Indonesia telah melakukan berbagai upaya untuk mengatasi kesenjangan digital, seperti program USO untuk memperluas akses internet, pembangunan infrastruktur digital, dan program pelatihan digital seperti BANGKIT dan Merdeka Belajar. Tujuannya jelas: meningkatkan keterampilan digital masyarakat agar bisa bersaing di era AI.

Namun, akses terhadap teknologi saja tidak cukup. Literasi digital yang rendah menjadi hambatan besar. Banyak masyarakat belum memahami cara memanfaatkan teknologi secara efektif, sehingga menghambat pemanfaatan potensi AI secara optimal.

Untuk mengatasi masalah ini, diperlukan kolaborasi lintas sektor. Pemerintah, swasta, dan komunitas AI harus bekerja sama membangun ekosistem yang mendukung pengembangan AI. Startup AI di Indonesia juga berperan penting dalam menghadirkan solusi inovatif dan meningkatkan aksesibilitas AI.

Tantangan ke depan meliputi perluasan akses internet dan perangkat digital, peningkatan kualitas pendidikan, dan pengembangan ekosistem inovasi yang inklusif. Pemerintah perlu membuat regulasi yang mendukung perkembangan AI namun tetap memperhatikan etika dan keamanan.

Meskipun tantangan besar, upaya yang telah dilakukan menunjukkan hasil positif. Kolaborasi yang kuat antara berbagai pihak menjadi kunci untuk menciptakan masa depan yang inklusif di era AI.

Kesenjangan digital di Indonesia merupakan masalah kompleks yang membutuhkan solusi komprehensif. Dengan upaya bersama, Indonesia

dapat mengatasi tantangan ini dan meraih manfaat dari perkembangan teknologi AI.

BAB 16: Kerangka Kerja dan Regulasi

Strategi Nasional Kecerdasan Buatan (AI) untuk Koperasi dan UMKM

Pengembangan kecerdasan buatan (AI) di Indonesia tidak hanya berfokus pada perusahaan besar, namun juga memberikan perhatian khusus pada koperasi dan UMKM. Berikut beberapa poin penting dalam strategi nasional AI yang relevan bagi sektor ini:

- **Aksesibilitas Teknologi:** Pemerintah perlu memastikan bahwa teknologi AI tidak hanya dikuasai oleh perusahaan besar, namun juga dapat diakses dan dimanfaatkan oleh koperasi dan UMKM. Ini bisa dilakukan melalui penyediaan pelatihan, infrastruktur yang terjangkau, serta platform AI yang mudah digunakan.
- **Pengembangan Kapasitas:** Koperasi dan UMKM perlu diberikan pelatihan dan pendampingan untuk memahami dan memanfaatkan AI dalam bisnis mereka. Ini bisa berupa pelatihan dasar mengenai AI, hingga pelatihan spesifik untuk penerapan AI dalam berbagai sektor, seperti produksi, pemasaran, dan manajemen.
- **Solusi AI yang Terjangkau:** Pengembangan solusi AI yang disesuaikan dengan kebutuhan dan skala bisnis koperasi dan UMKM sangat penting. Solusi ini harus terjangkau, mudah diimplementasikan, dan memberikan manfaat yang nyata bagi bisnis mereka.
- **Kemitraan dengan Startup:** Koperasi dan UMKM dapat bekerja sama dengan startup AI untuk mengembangkan solusi yang inovatif dan spesifik bagi kebutuhan mereka. Pemerintah dapat memfasilitasi kemitraan ini melalui program inkubator, akselerator, dan pendanaan.
- **Data yang Tersedia:** Ketersediaan data yang berkualitas dan relevan merupakan kunci dalam pengembangan AI. Pemerintah perlu mendorong keterbukaan data yang dapat dimanfaatkan oleh koperasi dan UMKM untuk mengembangkan produk dan layanan berbasis AI.
- **Regulasi yang Mendukung:** Pemerintah perlu menciptakan lingkungan regulasi yang kondusif bagi pertumbuhan AI di sektor

koperasi dan UMKM. Regulasi ini harus sederhana, jelas, dan tidak menjadi hambatan bagi inovasi.

- **Penerapan AI dalam Layanan Publik:** Penggunaan AI dalam layanan publik, seperti perizinan, pembiayaan, dan pemasaran, dapat memberikan manfaat bagi koperasi dan UMKM.

Dengan dukungan pemerintah dan ekosistem yang kondusif, koperasi dan UMKM di Indonesia dapat memanfaatkan potensi AI untuk meningkatkan produktivitas, efisiensi, dan daya saing mereka di pasar global.

Tata kelola dan standar data

Agar koperasi dan UMKM di Indonesia dapat memanfaatkan teknologi kecerdasan buatan (AI) secara optimal, tata kelola dan standar data yang baik sangatlah penting. Beberapa langkah konkret yang dapat dilakukan antara lain:

- **Pengembangan Standar Data Sederhana:** Pemerintah perlu bekerja sama dengan koperasi dan UMKM untuk membuat standar data yang mudah dipahami dan diterapkan. Standar ini akan membantu koperasi dan UMKM dalam mengelola data mereka dengan lebih baik.
- **Pelatihan Praktis:** Memberikan pelatihan khusus kepada pengurus koperasi dan UMKM tentang cara mengelola data dengan aman dan efektif. Pelatihan ini harus fokus pada praktik-praktik yang relevan dengan kebutuhan sehari-hari mereka.
- **Akses Data Pemerintah yang Mudah:** Menyediakan data pemerintah yang relevan dengan sektor koperasi dan UMKM secara terbuka dan mudah diakses. Data ini dapat digunakan untuk mengembangkan berbagai aplikasi AI yang bermanfaat bagi koperasi dan UMKM.
- **Bantuan Infrastruktur:** Memberikan dukungan dalam bentuk infrastruktur teknologi informasi yang terjangkau bagi koperasi dan UMKM. Hal ini akan membantu mereka dalam mengelola data dan mengembangkan sistem informasi yang dibutuhkan.
- **Kemitraan dengan Startup:** Memfasilitasi kemitraan antara koperasi dan UMKM dengan startup yang memiliki keahlian di bidang AI. Kemitraan ini dapat menghasilkan solusi-solusi inovatif

yang dapat meningkatkan efisiensi dan produktivitas koperasi dan UMKM.

Dengan langkah-langkah di atas, diharapkan koperasi dan UMKM di Indonesia dapat memanfaatkan potensi AI untuk meningkatkan daya saing dan memberikan manfaat yang lebih besar bagi anggota dan masyarakat.

Keamanan siber dan ketahanan data

Keamanan Siber dan Ketahanan Data untuk Operasional AI di Indonesia
Keamanan siber dan ketahanan data menjadi aspek penting dalam operasional AI di Indonesia. Hal ini dikarenakan AI mengandalkan data dalam jumlah besar untuk menjalankan fungsinya. Data tersebut dapat berupa informasi pribadi, data keuangan, atau data sensitif lainnya. Oleh karena itu, penting untuk memastikan bahwa data tersebut aman dan terlindungi dari berbagai ancaman siber.

Berikut adalah beberapa poin penting mengenai keamanan siber dan ketahanan data untuk operasional AI di Indonesia:

Ancaman Keamanan Siber pada Operasional AI

Koperasi dan UMKM di Indonesia yang mulai memanfaatkan kecerdasan buatan (AI) dalam operasionalnya, tak luput dari ancaman siber. Beberapa ancaman yang sering mengintai antara lain:

- **Serangan yang Membahayakan:** Serangan seperti peretasan, serangan malware, dan upaya penipuan (phishing) bisa mencuri data penting koperasi atau UMKM, bahkan mengganggu kegiatan sehari-hari.
- **Kebocoran Data yang Merugikan:** Kebocoran data bisa terjadi karena kesalahan manusia, masalah sistem, atau serangan siber. Akibatnya, informasi sensitif koperasi atau UMKM bisa jatuh ke tangan yang salah.
- **Penyalahgunaan Data yang Membahayakan:** Data yang dikumpulkan oleh AI bisa disalahgunakan untuk berbagai tujuan jahat, seperti penipuan, pencurian identitas, atau menyebarkan informasi bohong.

Cara Melindungi Koperasi dan UMKM dari Ancaman Siber

Untuk melindungi diri dari ancaman siber, koperasi dan UMKM perlu melakukan beberapa langkah penting:

- **Terapkan Standar Keamanan:** Adopsi standar keamanan seperti ISO/IEC 27001 dapat membantu melindungi data-data penting koperasi atau UMKM.
- **Tingkatkan Kesadaran Karyawan:** Edukasi dan pelatihan karyawan tentang pentingnya keamanan siber sangat krusial. Semua anggota koperasi atau karyawan UMKM perlu memahami cara melindungi data perusahaan.
- **Manfaatkan Teknologi Keamanan:** Gunakan teknologi keamanan seperti firewall, antivirus, dan enkripsi data untuk melindungi sistem dan data. Sebaiknya pilih produk dalam negeri yang telah diakui oleh Kementerian Pertahanan (KEMHAN) sebagai bagian dari industri pertahanan (INHAN).
- **Siapkan Rencana Darurat:** Buatlah rencana penanganan insiden keamanan siber. Rencana ini akan sangat membantu jika terjadi serangan siber. Biasanya, tim Computer Security Incident Response Team (CSIRT) di berbagai lembaga atau institusi di Indonesia dapat membantu, dengan koordinasi bersama Badan Siber Sandi Negara (BSSN).

Intinya, ancaman siber adalah risiko nyata yang harus dihadapi oleh koperasi dan UMKM. Dengan menerapkan langkah-langkah keamanan yang tepat, koperasi dan UMKM dapat melindungi aset digital mereka dan memastikan keberlangsungan bisnis.

Peran Pemerintah dalam Memproteksi Data Koperasi dan UMKM

Keamanan data saat ini menjadi isu krusial, terutama bagi Koperasi dan UMKM yang semakin mengadopsi teknologi digital. Pemerintah Indonesia memiliki peran strategis dalam melindungi aset berharga ini. Beberapa langkah konkret yang dapat dilakukan antara lain:

- **Penyusunan Peraturan yang Jelas:** Pemerintah perlu merumuskan aturan main yang jelas dan mudah dipahami oleh

Koperasi dan UMKM terkait keamanan data. Aturan ini harus disesuaikan dengan skala dan kompleksitas operasional mereka.

- **Penguatan Infrastruktur:** Pemerintah perlu membangun infrastruktur keamanan siber yang handal dan terjangkau bagi Koperasi dan UMKM. Ini bisa berupa penyediaan layanan keamanan siber yang bersubsidi atau kemudahan akses terhadap solusi keamanan yang sudah ada.
- **Kemitraan yang Efektif:** Pemerintah perlu menjalin kerjasama yang erat dengan sektor swasta untuk mengembangkan solusi keamanan siber yang khusus dirancang untuk Koperasi dan UMKM. Selain itu, program pelatihan dan edukasi juga sangat penting untuk meningkatkan kesadaran akan pentingnya keamanan data di kalangan pelaku usaha mikro.
- **Transparansi Anggaran:** Alokasi anggaran untuk keamanan siber harus transparan dan dikelola secara profesional. Hal ini akan menjamin efektivitas penggunaan dana dan mencegah terjadinya penyelewengan.

Mengapa Keamanan Data Penting bagi Koperasi dan UMKM?

Data adalah aset berharga bagi setiap bisnis, termasuk Koperasi dan UMKM. Dengan melindungi data, mereka dapat:

- **Mencegah Kerugian Finansial:** Kebocoran data dapat mengakibatkan kerugian finansial yang besar, terutama bagi bisnis berskala kecil.
- **Menjaga Reputasi:** Serangan siber dapat merusak reputasi bisnis dan membuat pelanggan enggan menggunakan layanan mereka.
- **Meningkatkan Kepercayaan:** Dengan menunjukkan komitmen terhadap keamanan data, Koperasi dan UMKM dapat membangun kepercayaan pelanggan dan mitra bisnis.

Kesimpulan

Keamanan siber bukan lagi menjadi pilihan, melainkan keharusan bagi Koperasi dan UMKM di era digital. Pemerintah memiliki peran sentral dalam menciptakan lingkungan yang aman bagi mereka untuk beroperasi. Dengan langkah-langkah yang tepat, Koperasi dan UMKM dapat tumbuh

dan berkembang secara berkelanjutan tanpa harus khawatir akan ancaman siber.

BAGIAN 4: MASA DEPAN AI, MACHINE LEARNING DAN BIG DATA

BAB 17: Strategi Big Data Efektif dan Hemat Biaya:

Bagian ini memberikan panduan bagi koperasi dan UMKM di Indonesia untuk memanfaatkan potensi besar dari data yang tersedia. Dengan mengadopsi pendekatan yang tepat, UMKM dan koperasi dapat meningkatkan efisiensi, membuat keputusan yang lebih baik, dan membuka peluang bisnis baru.

Poin-poin penting yang disesuaikan untuk UMKM dan Koperasi:

- **Manfaatkan Data yang Mudah Didapat:**
 - **Data Pemerintah:** Akses data sensus, data pasar lokal, dan data kebijakan pemerintah untuk memahami kondisi pasar dan kebutuhan konsumen.
 - **Data Transaksi:** Analisis data penjualan dan pembelian untuk mengidentifikasi tren, produk populer, dan pelanggan setia.
 - **Data Sosial Media:** Gunakan media sosial untuk memantau opini pelanggan, pesaing, dan tren pasar.
- **Kolaborasi:**
 - **Universitas:** Jalin kerjasama dengan universitas untuk mendapatkan bantuan dalam menganalisis data dan mengembangkan solusi yang inovatif.
 - **Komunitas UMKM:** Bergabung dengan komunitas UMKM untuk berbagi pengetahuan dan pengalaman dalam memanfaatkan data.
- **Manfaatkan Teknologi:**
 - **Perangkat Sederhana:** Gunakan smartphone dan aplikasi sederhana untuk mengumpulkan data.
 - **Perangkat Lunak Gratis:** Manfaatkan perangkat lunak open source untuk menganalisis data tanpa biaya tambahan.

- **Cloud Computing:** Simpan dan proses data di cloud untuk menghemat biaya infrastruktur.
- **Fokus pada Masalah Lokal:**
 - **Data Lokal:** Kumpulkan data yang relevan dengan bisnis Anda, seperti data pelanggan, data pemasok, dan data produksi.
 - **Solusi Sederhana:** Terapkan solusi yang mudah dipahami dan digunakan oleh karyawan.
- **Tantangan dan Solusi:**
 - **Kualitas Data:** Pastikan data yang dikumpulkan akurat dan relevan.
 - **Privasi:** Jaga kerahasiaan data pelanggan dan informasi sensitif lainnya.

Contoh Penerapan untuk UMKM:

- **Toko Kelontong:** Analisis data penjualan untuk mengetahui produk yang sering dibeli bersamaan, sehingga dapat mengatur tata letak toko yang lebih efektif.
- **Pengrajin:** Gunakan data sosial media untuk mengidentifikasi tren desain yang sedang populer dan menyesuaikan produk yang ditawarkan.
- **Petani:** Memanfaatkan data cuaca dan data tanah untuk menentukan waktu tanam yang optimal dan jenis tanaman yang cocok.

Dengan memanfaatkan data secara efektif, UMKM dan koperasi di Indonesia dapat meningkatkan daya saing, membuat keputusan bisnis yang lebih tepat, dan membuka peluang pasar yang lebih luas.

BAB 18: Trend dan Perkembangan Terbaru

Dari Laboratorium ke Dunia Nyata: Penerapan AI di Indonesia

Kecerdasan Buatan (AI) telah merevolusi berbagai sektor, termasuk bisnis. Kemajuan pesat dalam AI, seperti kemunculan model bahasa besar, menawarkan peluang besar bagi Usaha Mikro, Kecil, dan Menengah (UMKM) di Indonesia. Dengan AI, UMKM dapat mengoptimalkan proses bisnis, meningkatkan efisiensi, dan bahkan menciptakan produk serta layanan baru. Misalnya, AI dapat membantu UMKM dalam analisis data pelanggan, personalisasi pemasaran, serta pengelolaan rantai pasok yang lebih efektif.

Meskipun potensi AI sangat besar, UMKM di Indonesia masih menghadapi sejumlah tantangan dalam mengadopsi teknologi ini. Salah satunya adalah keterbatasan akses terhadap infrastruktur teknologi yang memadai, seperti koneksi internet yang stabil dan perangkat keras yang mumpuni. Selain itu, kurangnya sumber daya manusia yang kompeten di bidang AI juga menjadi kendala. Namun, di sisi lain, UMKM Indonesia juga memiliki peluang besar untuk memanfaatkan AI. Dengan dukungan pemerintah dan pengembangan ekosistem inovasi yang kuat, UMKM dapat bersaing di pasar global.

Koperasi memiliki peran penting dalam mendorong adopsi AI oleh UMKM. Sebagai organisasi ekonomi yang berorientasi pada anggota, koperasi dapat memfasilitasi akses UMKM terhadap teknologi AI melalui berbagai cara. Misalnya, koperasi dapat menyediakan pelatihan dan pendampingan bagi anggotanya dalam memanfaatkan AI, serta mengembangkan platform digital yang terintegrasi dengan AI untuk mendukung kegiatan bisnis UMKM. Selain itu, koperasi juga dapat berperan sebagai agregator data UMKM, sehingga data tersebut dapat diolah menggunakan AI untuk menghasilkan insights yang bernilai.

Untuk mempercepat adopsi AI oleh UMKM, diperlukan beberapa strategi. Pertama, pemerintah perlu memberikan insentif dan dukungan finansial bagi UMKM yang ingin mengadopsi AI. Kedua, perlu dibangun ekosistem inovasi yang kuat, yang melibatkan kolaborasi antara pemerintah, sektor

swasta, akademisi, dan pelaku UMKM. Ketiga, pengembangan sumber daya manusia di bidang AI sangat penting, baik melalui pendidikan formal maupun pelatihan non-formal. Terakhir, perlu dibangun infrastruktur teknologi yang memadai dan terjangkau bagi UMKM.

Dengan potensi yang besar dan dukungan yang tepat, AI dapat menjadi katalisator pertumbuhan UMKM di Indonesia. AI dapat membantu UMKM meningkatkan produktivitas, memperluas pasar, dan menciptakan lapangan kerja baru. Namun, penting untuk diingat bahwa pengembangan AI harus dilakukan secara bertanggung jawab, dengan memperhatikan aspek etika dan keamanan. Selain itu, perlu dibangun kerangka kerja yang jelas untuk mengatur penggunaan AI, sehingga manfaatnya dapat dinikmati oleh seluruh lapisan masyarakat.

Indonesia di Era Big Data: Tantangan dan Peluang

Era Big Data telah membawa perubahan signifikan dalam dunia bisnis. Dengan adanya teknologi kecerdasan buatan, internet of things, dan cloud computing, analisis data menjadi lebih mendalam dan cepat. Hal ini membuka peluang besar bagi koperasi dan UMKM untuk mengoptimalkan kinerja mereka. Melalui analisis data yang tepat, koperasi dan UMKM dapat mengidentifikasi tren pasar, perilaku konsumen, serta efisiensi operasional. Informasi berharga ini dapat digunakan untuk membuat keputusan bisnis yang lebih baik, meningkatkan produktivitas, dan memperluas pasar.

Meskipun potensi Big Data sangat besar, koperasi dan UMKM di Indonesia masih menghadapi sejumlah tantangan dalam memanfaatkan teknologi ini. Salah satu kendala utama adalah keterbatasan sumber daya, baik itu sumber daya manusia yang memiliki keahlian di bidang data maupun infrastruktur teknologi yang memadai. Selain itu, biaya implementasi solusi Big Data juga menjadi hambatan bagi sebagian besar koperasi dan UMKM. Kurangnya kesadaran akan pentingnya data dan analisis data juga menjadi faktor penghambat.

Untuk mengatasi tantangan tersebut, diperlukan beberapa upaya strategis. Pertama, pemerintah perlu memberikan dukungan yang lebih besar kepada koperasi dan UMKM dalam mengadopsi teknologi Big Data. Hal ini dapat dilakukan melalui penyediaan pelatihan, insentif, dan akses

yang lebih mudah terhadap teknologi. Kedua, perguruan tinggi dan lembaga penelitian perlu berperan aktif dalam mengembangkan solusi Big Data yang sesuai dengan kebutuhan koperasi dan UMKM. Ketiga, perlu dibangun ekosistem yang mendukung kolaborasi antara koperasi, UMKM, perusahaan teknologi, dan pemangku kepentingan lainnya.

Dengan memanfaatkan Big Data, koperasi dan UMKM dapat memperoleh berbagai manfaat, antara lain:

- **Peningkatan efisiensi:** Mengoptimalkan proses produksi dan distribusi, mengurangi biaya operasional.
- **Pengembangan produk dan layanan:** Mengembangkan produk dan layanan baru yang sesuai dengan kebutuhan pasar.
- **Pengambilan keputusan yang lebih baik:** Membuat keputusan bisnis yang lebih tepat berdasarkan data yang akurat.
- **Penguatan posisi di pasar:** Bersaing dengan perusahaan besar dan meningkatkan daya saing.
- **Pemberdayaan anggota:** Memberikan informasi yang relevan kepada anggota koperasi untuk meningkatkan kesejahteraan mereka.

Big Data memiliki potensi yang sangat besar untuk mendorong pertumbuhan dan perkembangan koperasi dan UMKM di Indonesia. Namun, untuk dapat memanfaatkan potensi tersebut, diperlukan upaya bersama dari berbagai pihak. Dengan dukungan yang tepat, koperasi dan UMKM dapat menjadi pemain yang lebih kuat di era digital.

Indonesia Digital: Tantangan dan Solusi dalam Era Transformasi Teknologi

Big data, dengan kemampuannya menganalisis data dalam skala besar, menawarkan peluang emas bagi koperasi dan UMKM di Indonesia. Melalui analisis data, koperasi dan UMKM dapat mengidentifikasi tren pasar, perilaku konsumen, serta efisiensi operasional. Informasi berharga ini dapat digunakan untuk menyusun strategi bisnis yang lebih efektif, mengembangkan produk atau layanan baru yang sesuai dengan kebutuhan pasar, serta meningkatkan produktivitas. Dengan demikian, big data dapat menjadi tool yang ampuh bagi koperasi dan UMKM untuk bersaing di pasar yang semakin kompetitif.

Industri fintech dan IT di Indonesia telah menunjukkan pertumbuhan yang pesat. Berbagai inovasi teknologi seperti pembayaran digital, pinjaman online, dan manajemen keuangan berbasis cloud telah memberikan kemudahan bagi koperasi dan UMKM dalam mengelola keuangan, mengakses modal, serta memperluas jaringan bisnis. Selain itu, teknologi informasi juga memungkinkan koperasi dan UMKM untuk berkolaborasi dengan pihak lain, seperti pemasok dan distributor, sehingga dapat meningkatkan efisiensi rantai pasok.

Meskipun teknologi menawarkan banyak manfaat, koperasi dan UMKM di Indonesia juga menghadapi sejumlah tantangan dalam proses transformasi digital. Kesenjangan digital, kurangnya literasi digital, serta infrastruktur yang belum merata menjadi beberapa kendala utama. Untuk mengatasi tantangan tersebut, pemerintah perlu memberikan dukungan yang lebih besar, seperti penyediaan akses internet yang terjangkau, pelatihan digital bagi pelaku UMKM, serta pengembangan infrastruktur digital di daerah-daerah tertinggal.

Kolaborasi antara pemerintah, sektor swasta, dan akademisi sangat penting untuk mempercepat transformasi digital koperasi dan UMKM di Indonesia. Pemerintah dapat berperan sebagai fasilitator, regulator, dan penyedia infrastruktur. Sektor swasta, terutama perusahaan teknologi, dapat mengembangkan solusi-solusi inovatif yang sesuai dengan kebutuhan koperasi dan UMKM. Akademisi dapat berperan dalam melakukan penelitian dan pengembangan, serta memberikan pelatihan kepada pelaku UMKM. Dengan bekerja sama, semua pihak dapat menciptakan ekosistem digital yang kondusif bagi pertumbuhan koperasi dan UMKM di Indonesia.

Transformasi Digital: Integrasi Blockchain dan Process Mining

Teknologi blockchain dan process mining, dua inovasi digital yang saling melengkapi, menawarkan solusi revolusioner bagi dunia bisnis. Blockchain, dengan sistem pencatatan yang transparan dan tak terhapuskan, ideal untuk melacak setiap langkah dalam sebuah proses. Sementara itu, process mining mampu menganalisis data secara mendalam untuk mengoptimalkan alur kerja. Kombinasi keduanya dapat

meningkatkan efisiensi dan transparansi dalam berbagai sektor, termasuk koperasi dan UMKM yang seringkali terkendala oleh proses bisnis yang kompleks.

Meskipun Indonesia memiliki potensi besar dalam mengadopsi teknologi ini, beberapa tantangan perlu diatasi. Salah satunya adalah mindset yang masih berfokus pada aspek spekulatif dari blockchain, seperti cryptocurrency. Padahal, potensi sebenarnya terletak pada penerapan teknologi ini dalam meningkatkan efisiensi operasional. Untuk mengatasi hal ini, Indonesia perlu membangun ekosistem yang mendukung, mulai dari pengembangan sumber daya manusia yang kompeten hingga penyusunan regulasi yang jelas.

Koperasi dan UMKM di Indonesia dapat sangat diuntungkan dari integrasi blockchain dan process mining. Misalnya, koperasi dapat menggunakan blockchain untuk melacak asal usul produk, meningkatkan kepercayaan konsumen, dan memastikan transparansi dalam pengelolaan keuangan. UMKM juga dapat memanfaatkan teknologi ini untuk mengoptimalkan rantai pasok, mengurangi biaya operasional, dan meningkatkan akses ke pasar yang lebih luas.

Dengan dukungan pemerintah, investasi yang tepat, dan kolaborasi yang kuat antara industri dan akademisi, Indonesia memiliki potensi besar untuk menjadi pemimpin regional dalam pengembangan solusi berbasis blockchain dan process mining. Industri IT Indonesia dapat mengambil peran sentral dalam mengembangkan aplikasi-aplikasi inovatif yang menjawab kebutuhan spesifik koperasi dan UMKM. Hal ini tidak hanya akan meningkatkan daya saing bisnis di dalam negeri, tetapi juga menarik investasi asing dan menciptakan lapangan kerja baru di sektor digital.

BAB 19: Rekomendasi dan Strategi untuk Implementasi yang Efektif

Membangun kepemimpinan dan komitmen

Agar koperasi dan UMKM di Indonesia dapat memanfaatkan potensi kecerdasan buatan (AI) secara maksimal, diperlukan komitmen kuat dari berbagai pihak. Salah satu langkah penting adalah merumuskan visi yang jelas tentang bagaimana AI dapat meningkatkan efisiensi, memperluas pasar, dan mendorong inovasi dalam bisnis koperasi dan UMKM.

Kolaborasi antara koperasi, UMKM, akademisi, startup, dan pemerintah sangat krusial. Pemerintah perlu berperan aktif dalam menyediakan dukungan kebijakan, anggaran, dan infrastruktur teknologi yang memadai. Selain itu, pemerintah daerah juga dapat berperan sebagai fasilitator untuk menghubungkan berbagai pihak dan menciptakan ekosistem yang kondusif bagi pertumbuhan bisnis berbasis AI di tingkat lokal.

Implementasi Strategi yang Konkrit

Untuk mewujudkan visi tersebut, beberapa langkah strategis perlu dilakukan, antara lain:

- **Penyusunan Rencana Aksi yang Komprehensif:** Pemerintah perlu menyusun rencana aksi yang detail dengan target yang jelas dan melibatkan semua pemangku kepentingan.
- **Pengembangan Sumber Daya Manusia:** Investasi dalam pendidikan dan pelatihan AI sangat penting untuk menghasilkan tenaga kerja yang kompeten di bidang ini.
- **Peningkatan Infrastruktur Digital:** Pembangunan infrastruktur digital yang memadai, seperti jaringan internet yang cepat dan stabil, sangat penting untuk mendukung penerapan AI.
- **Pembentukan Ekosistem Startup:** Pemerintah perlu menciptakan ekosistem yang kondusif bagi pertumbuhan startup AI yang dapat memberikan solusi inovatif bagi koperasi dan UMKM.

- **Penerapan Etika AI:** Pengembangan AI harus selalu mempertimbangkan aspek etika untuk mencegah penyalahgunaan teknologi.

Potensi Besar AI untuk Koperasi dan UMKM

Dengan penerapan AI yang tepat, koperasi dan UMKM di Indonesia dapat meningkatkan produktivitas, memperluas akses pasar, dan menciptakan produk serta layanan yang lebih inovatif. Hal ini akan berdampak positif pada pertumbuhan ekonomi nasional dan kesejahteraan masyarakat.

Mengembangkan infrastruktur dan kapasitas

Pengembangan kecerdasan buatan (AI) di Indonesia membutuhkan infrastruktur yang kuat dan sumber daya manusia yang mumpuni. Namun, tantangan utama adalah keterbatasan akses terhadap teknologi canggih, terutama bagi koperasi dan UMKM. Peluang besar terbuka lebar jika kita dapat membangun ekosistem AI yang inklusif, sehingga teknologi ini dapat dimanfaatkan oleh semua pelaku ekonomi, termasuk usaha-usaha berskala kecil.

Solusi yang Dibutuhkan:

- **Infrastruktur Terjangkau:** Koperasi dan UMKM membutuhkan akses yang mudah dan terjangkau terhadap infrastruktur digital, seperti internet berkecepatan tinggi dan komputasi awan. Pemerintah dapat memberikan insentif khusus untuk penyedia layanan internet agar mereka dapat memperluas jangkauan ke daerah-daerah yang belum terjangkau.
- **Pelatihan yang Sesuai:** Program pelatihan AI harus dirancang khusus untuk memenuhi kebutuhan koperasi dan UMKM. Pelatihan ini harus bersifat praktis dan mudah dipahami, sehingga para pelaku usaha dapat langsung menerapkan AI dalam bisnis mereka.
- **Pendanaan yang Cukup:** Pemerintah dan lembaga swasta perlu menyediakan dana khusus untuk mendukung pengembangan AI di kalangan koperasi dan UMKM. Dana ini dapat digunakan untuk membiayai pelatihan, pengembangan produk berbasis AI, dan konsultasi. Pemerintah perlu menyediakan dana untuk penelitian

AI, dan swasta perlu berinvestasi di perusahaan-perusahaan AI yang potensial. Keyword Googling yang akan menarik untuk hal ini adalah “Inkubator AI Indonesia”, kita akan melihat beberapa aktivitas inkubator dari KEMENPAREKRAF, indigo.id, dll.

- **Kerjasama yang Efektif:** Perlu adanya kerjasama yang erat antara pemerintah, akademisi, pelaku industri, dan organisasi koperasi untuk membangun ekosistem AI yang inklusif. Kerjasama ini dapat dilakukan dalam bentuk program inkubator, akselerator, dan pusat inovasi.

Contoh Implementasi:

- **Koperasi Pertanian:** AI dapat digunakan untuk menganalisis data cuaca, kualitas tanah, dan harga pasar, sehingga petani dapat membuat keputusan yang lebih baik dalam produksi dan pemasaran.
- **UMKM Kerajinan:** AI dapat membantu UMKM kerajinan dalam mendesain produk baru, memprediksi tren pasar, dan menemukan pelanggan potensial melalui pemasaran digital.

Berikut adalah beberapa contoh program dan inisiatif yang telah dilakukan oleh Indonesia untuk mengembangkan AI:

- **Roadmap AI Indonesia 2020-2025:** Roadmap ini berisi strategi dan target untuk pengembangan AI di Indonesia.
- **Pusat Unggulan Artificial Intelligence:** Saat ini cukup banyak Pusat Unggulan bidang Artificial Intelligence di kampus Indonesia. Hal ini dapat di Googling dengan mudah dengan keyword “Pusat Unggulan AI Indonesia”
- **Program Digital Talent Scholarship:** Digital Talent banyak difasilitasi oleh KOMINFO & kampus2 secara gratis. Khususnya bidang AI dapat diperoleh melalui URL <https://digitalent.kominfo.go.id/cari?s=AI>.

Dengan membangun ekosistem AI yang kuat dan inklusif, koperasi dan UMKM di Indonesia dapat meningkatkan produktivitas, daya saing, dan mampu beradaptasi dengan perubahan zaman. Pemerintah dan seluruh pemangku kepentingan perlu bekerja sama untuk mewujudkan hal ini.

Memastikan etika dan privasi data

Koperasi dan UMKM di Indonesia saat ini memiliki peluang besar untuk memanfaatkan kecerdasan buatan (AI) dalam meningkatkan efisiensi dan produktivitas. Namun, penggunaan AI juga harus diiringi dengan pemahaman yang mendalam tentang etika dan privasi data.

Peraturan Privasi Data

Indonesia memiliki beberapa peraturan yang mengatur tentang privasi data, antara lain:

- Undang-Undang Nomor 11 Tahun 2008 tentang Informasi dan Transaksi Elektronik: Undang-undang ini mengatur tentang perlindungan data pribadi dalam transaksi elektronik.
- Peraturan Pemerintah Nomor 71 Tahun 2019 tentang Penyelenggaraan Sistem dan Transaksi Elektronik: Peraturan ini mengatur tentang tata cara penyelenggaraan sistem dan transaksi elektronik.
- Undang-Undang Nomor 27 Tahun 2022 tentang Perlindungan Data Pribadi (UU PDP): Merupakan tonggak sejarah baru dalam perlindungan data pribadi di Indonesia. UU PDP secara khusus mengatur berbagai aspek perlindungan data pribadi, mulai dari definisi data pribadi, prinsip-prinsip perlindungan data, hak-hak subjek data, kewajiban pengontrol dan pemroses data, hingga sanksi bagi pelanggar.
- Peraturan Menteri Komunikasi dan Informatika Nomor 20 Tahun 2016 tentang Perlindungan Data Pribadi dalam Sistem Elektronik: Peraturan menteri ini mengatur tentang tata cara perlindungan data pribadi dalam sistem elektronik.

Mengapa Etika dan Privasi Penting bagi Koperasi dan UMKM?

- **Data adalah Aset Berharga:** Baik koperasi maupun UMKM memiliki data pelanggan, keuangan, dan operasional yang sangat berharga. Data ini dapat diolah menggunakan AI untuk menghasilkan wawasan bisnis yang lebih baik.

- **Perlindungan Hukum:** Indonesia telah memiliki Undang-Undang Perlindungan Data Pribadi (UU PDP). Pelanggaran terhadap UU PDP dapat berakibat sanksi yang berat.
- **Kepercayaan Pelanggan:** Pelanggan semakin peduli dengan privasi data mereka. Penggunaan AI yang tidak etis dapat merusak kepercayaan pelanggan terhadap koperasi atau UMKM.

Prinsip-Prinsip Etika AI untuk Koperasi dan UMKM

- **Keadilan:** Gunakan AI untuk memberikan manfaat yang sama kepada semua anggota koperasi atau pelanggan UMKM.
- **Transparansi:** Pastikan anggota atau pelanggan memahami bagaimana AI digunakan dalam proses bisnis.
- **Akuntabilitas:** Tetap bertanggung jawab atas keputusan yang dihasilkan oleh AI.
- **Privasi:** Lindungi data pribadi anggota atau pelanggan sesuai dengan UU PDP.

Langkah-langkah Praktis

1. **Pelajari UU PDP:** Pahami hak dan kewajiban Anda sebagai pengelola data.
2. **Lakukan Asesmen Risiko:** Identifikasi potensi risiko pelanggaran privasi data dalam penggunaan AI.
3. **Kembangkan Kebijakan Privasi:** Buat kebijakan yang jelas tentang pengumpulan, penggunaan, dan penyimpanan data.
4. **Gunakan Teknologi yang Aman:** Pilih solusi AI yang memiliki fitur keamanan yang kuat.
5. **Edukasi Anggota/Pelanggan:** Informasikan kepada anggota atau pelanggan tentang penggunaan AI dan upaya perlindungan data yang dilakukan.

Manfaat Mengimplementasikan AI yang Bertanggung Jawab

- **Peningkatan Efisiensi:** Otomatisasi tugas-tugas rutin.
- **Pengambilan Keputusan yang Lebih Baik:** Analisis data yang lebih mendalam.
- **Peningkatan Layanan Pelanggan:** Personalisasi layanan.

- **Inovasi Produk dan Jasa:** Pengembangan produk atau jasa baru yang lebih relevan.

Dengan menerapkan prinsip-prinsip etika dan privasi data, koperasi dan UMKM di Indonesia dapat memanfaatkan potensi AI secara maksimal tanpa mengorbankan kepercayaan pelanggan dan mematuhi peraturan yang berlaku.

Saran Tambahan:

- **Kolaborasi dengan Ahli:** Bekerja sama dengan ahli data dan AI untuk memastikan implementasi yang tepat.
- **Manfaatkan Program Pemerintah:** Cari tahu program pemerintah yang mendukung adopsi teknologi AI oleh UMKM.
- **Bergabung dengan Komunitas:** Bergabung dengan komunitas UMKM atau koperasi untuk berbagi pengalaman dan pengetahuan.

Kolaborasi AI untuk Memperkuat Koperasi dan UMKM

Kecerdasan buatan (AI) menawarkan potensi besar untuk meningkatkan efisiensi dan produktivitas UMKM serta koperasi. Namun, untuk memanfaatkan potensi ini secara maksimal, dibutuhkan kolaborasi yang kuat antara berbagai pihak.

- **Membangun Jaringan Kolaborasi Inklusif:** Salah satu langkah penting adalah membentuk forum khusus yang melibatkan koperasi, UMKM, startup teknologi, pemerintah, dan akademisi. Forum ini dapat menjadi wadah untuk berbagi pengetahuan, pengalaman, dan mencari solusi bersama. Misalnya, mengadakan pertemuan rutin secara virtual atau tatap muka untuk membahas tantangan dan peluang yang dihadapi oleh koperasi dan UMKM dalam mengadopsi teknologi AI.
- **Mengembangkan Proyek Kolaboratif yang Relevan:** Proyek kolaboratif yang melibatkan koperasi dan UMKM dapat fokus pada pengembangan aplikasi AI yang spesifik untuk kebutuhan mereka. Misalnya, membangun platform e-commerce yang disesuaikan dengan kebutuhan produk UMKM, atau mengembangkan sistem

manajemen rantai pasok yang lebih efisien untuk koperasi pertanian.

- **Membangun Ekosistem Data yang Terjangkau:** Ketersediaan data yang berkualitas dan relevan sangat penting bagi pengembangan AI. Namun, banyak koperasi dan UMKM yang kesulitan mengakses data yang mereka butuhkan. Oleh karena itu, perlu dibangun ekosistem data yang lebih terbuka dan terjangkau, serta menyediakan pelatihan bagi UMKM untuk memanfaatkan data secara efektif.
- **Meningkatkan Kapasitas Sumber Daya Manusia:** Pemerintah dan lembaga pendidikan perlu menyediakan program pelatihan dan pendidikan yang khusus dirancang untuk anggota koperasi dan pelaku UMKM. Program ini dapat mencakup pelatihan dasar tentang AI, serta pelatihan yang lebih khusus untuk aplikasi AI dalam sektor tertentu.
- **Mendorong Investasi dan Inkubasi:** Untuk mendorong pertumbuhan startup AI yang fokus pada sektor koperasi dan UMKM, perlu disediakan fasilitas inkubasi dan akselerasi yang memadai. Selain itu, pemerintah dapat memberikan insentif bagi investor yang bersedia mendanai startup-startup ini.

Contoh Praktis:

- **Hackathon:** Mengadakan hackathon dengan tema khusus untuk mencari solusi AI bagi permasalahan yang dihadapi oleh koperasi dan UMKM, seperti misalnya meningkatkan kualitas produk pertanian atau mengembangkan sistem pemasaran digital yang efektif. Salah satu contoh hackaton AI adalah <https://hackathon.fekdi.co.id/> Bank Indonesia Hackaton 2024 yang dilaksanakan bulan Juli 2024.
- **Inkubator Koperasi:** Mendirikan inkubator khusus untuk membantu koperasi mengembangkan produk dan layanan berbasis AI. Googling dengan keyword “inkubator AI Indonesia” akan menemukan banyak sekali aktivitas dari berbagai lembaga hingga kementerian Parekraf dll. Salah satu yang cukup konsisten adalah <https://indigo.id/>

Dengan menerapkan strategi-strategi di atas, diharapkan kolaborasi dan kemitraan di bidang AI dapat semakin kuat, sehingga dapat mendorong inovasi, pertumbuhan ekonomi, dan kesejahteraan masyarakat.

BAGIAN 5: SIAP LEPAS LANDAS

BAB 20: Ringkasan dan Pesan Utama

Potensi AI/ML/Big Data untuk Koperasi dan UMKM

Revolusi digital yang ditandai oleh kecerdasan buatan (AI), pembelajaran mesin (ML), dan big data tidak hanya mengubah sektor pemerintahan, tetapi juga memiliki potensi besar untuk memberdayakan koperasi dan UMKM. Berikut beberapa cara teknologi ini dapat dimanfaatkan:

- **Peningkatan Efisiensi Operasional:**
 - **Otomatisasi Tugas:** AI dapat membantu mengotomatiskan tugas-tugas administratif seperti pencatatan keuangan, pengelolaan inventori, dan layanan pelanggan, sehingga koperasi dan UMKM dapat mengalokasikan waktu dan sumber daya untuk kegiatan yang lebih strategis.
 - **Optimasi Rantai Pasok:** Dengan menganalisis data besar, koperasi dan UMKM dapat mengoptimalkan rantai pasok, mulai dari pengadaan bahan baku hingga distribusi produk, sehingga mengurangi biaya dan meningkatkan efisiensi.
- **Pengambilan Keputusan yang Lebih Baik:**
 - **Analisis Pasar:** AI dapat membantu menganalisis data pasar, tren konsumen, dan perilaku pesaing untuk membantu koperasi dan UMKM dalam membuat keputusan produk, pemasaran, dan pengembangan bisnis yang lebih tepat.
 - **Prediksi Permintaan:** Dengan memprediksi permintaan pasar, koperasi dan UMKM dapat mengatur produksi dan persediaan dengan lebih baik, sehingga mengurangi risiko kelebihan atau kekurangan stok.
- **Pengembangan Produk dan Layanan Baru:**
 - **Personalisasi Produk:** AI dapat membantu dalam mengembangkan produk dan layanan yang lebih personal sesuai dengan kebutuhan dan preferensi konsumen,

sehingga meningkatkan kepuasan pelanggan dan loyalitas merek.

- **Inovasi:** Dengan menganalisis data besar, koperasi dan UMKM dapat mengidentifikasi peluang bisnis baru dan mengembangkan produk atau layanan inovatif.
- **Peningkatan Akses ke Pasar:**
 - **E-commerce:** AI dapat mendukung pengembangan platform e-commerce yang lebih cerdas, sehingga koperasi dan UMKM dapat menjangkau pasar yang lebih luas dan meningkatkan penjualan.
 - **Marketing Digital:** AI dapat digunakan untuk melakukan pemasaran digital yang lebih efektif, seperti target iklan yang lebih presisi dan analisis sentimen pelanggan.

Contoh Penerapan:

- **Koperasi Pertanian:** AI dapat membantu dalam menganalisis data cuaca, kondisi tanah, dan harga pasar untuk mengoptimalkan produksi pertanian dan memprediksi panen.
- **UMKM Kerajinan Tangan:** AI dapat digunakan untuk menganalisis tren desain dan preferensi konsumen untuk menciptakan produk kerajinan tangan yang lebih menarik dan diminati pasar.

Teknologi AI/ML/Big Data menawarkan peluang besar bagi koperasi dan UMKM untuk meningkatkan daya saing, efisiensi, dan inovasi. Dengan memanfaatkan teknologi ini secara tepat, koperasi dan UMKM dapat tumbuh dan berkembang secara berkelanjutan.

Menjinakkan AI: Strategi Mengatasi Tantangan dan Risiko Implementasi

Implementasi kecerdasan buatan (AI) di Indonesia, khususnya pada sektor koperasi dan UMKM, menghadapi sejumlah kendala.

Keterbatasan data berkualitas menjadi salah satu tantangan utama. Banyak koperasi dan UMKM yang belum memiliki sistem pencatatan data yang baik, sehingga sulit untuk memanfaatkan AI dalam pengambilan keputusan. **Kurangnya tenaga ahli** di bidang data dan AI juga menjadi hambatan besar.

Selain itu, **infrastruktur teknologi yang belum merata** di berbagai daerah, terutama di daerah pedesaan, menjadi kendala lain. Padahal, koperasi dan UMKM di daerah-daerah ini sangat berpotensi untuk memanfaatkan teknologi AI untuk meningkatkan produktivitas dan daya saing.

Risiko yang perlu diwaspadai dalam penerapan AI di sektor koperasi dan UMKM antara lain **kemungkinan terjadinya bias algoritma** yang dapat merugikan kelompok tertentu, serta **pelanggaran privasi data** anggota koperasi.

Untuk mengatasi tantangan dan meminimalisir risiko, pemerintah perlu memberikan dukungan yang lebih besar kepada koperasi dan UMKM dalam hal **aksesibilitas data, pelatihan sumber daya manusia**, serta **penyediaan infrastruktur teknologi yang memadai**. Selain itu, perlu dibuat **regulasi yang jelas** terkait penggunaan AI di sektor koperasi dan UMKM untuk melindungi kepentingan semua pihak.

Dengan dukungan yang tepat, AI dapat menjadi tool yang sangat bermanfaat bagi koperasi dan UMKM untuk meningkatkan efisiensi, inovasi, dan daya saing. Misalnya, AI dapat digunakan untuk menganalisis data penjualan, memprediksi tren pasar, serta mengoptimalkan rantai pasok.

Dari Konsep ke Kenyataan: Strategi Implementasi AI

Implementasi kecerdasan buatan (AI) di Indonesia menawarkan peluang besar bagi koperasi dan UMKM untuk meningkatkan efisiensi, produktivitas, dan daya saing. Namun, agar AI dapat memberikan manfaat nyata bagi sektor ini, diperlukan strategi yang tepat dan kolaborasi yang kuat.

Strategi yang tepat melibatkan identifikasi sektor-sektor spesifik dalam koperasi dan UMKM di mana AI dapat memberikan dampak paling signifikan, seperti dalam hal manajemen rantai pasok, pemasaran digital, atau pengembangan produk baru. Selain itu, perlu adanya upaya untuk mengembangkan sumber daya manusia yang memiliki keterampilan dalam bidang AI, serta memastikan ketersediaan infrastruktur teknologi yang terjangkau dan mudah diakses oleh koperasi dan UMKM.

Kolaborasi yang kuat antara pemerintah, lembaga penelitian, perusahaan teknologi, dan organisasi pendukung UMKM sangat penting untuk mempercepat adopsi AI di sektor ini. Kolaborasi ini dapat berupa pengembangan solusi AI yang disesuaikan dengan kebutuhan spesifik koperasi dan UMKM, penyediaan pelatihan dan pendampingan, serta pembentukan ekosistem inovasi yang mendukung pertumbuhan bisnis berbasis AI.

Dengan strategi dan kolaborasi yang tepat, koperasi dan UMKM di Indonesia dapat memanfaatkan potensi AI untuk:

- **Meningkatkan efisiensi operasional:** Automating tugas-tugas rutin, mengoptimalkan penggunaan sumber daya, dan membuat pengambilan keputusan yang lebih baik.
- **Memperluas pasar:** Menerapkan pemasaran digital yang lebih efektif, personalisasi produk dan layanan, serta membangun hubungan yang lebih kuat dengan pelanggan.
- **Mengembangkan produk dan layanan baru:** Menggunakan AI untuk menganalisis data pasar, mengidentifikasi tren, dan menciptakan inovasi yang relevan.
- **Meningkatkan kualitas produk:** Menerapkan sistem kontrol kualitas berbasis AI untuk menjamin konsistensi dan keandalan produk.

Intinya, AI dapat menjadi tool yang ampuh bagi koperasi dan UMKM di Indonesia untuk tumbuh dan berkembang. Namun, keberhasilan implementasi AI sangat bergantung pada kesiapan dan kemampuan kita untuk beradaptasi dengan perubahan teknologi.

Dilema Digital: Konflik Antara Antusiasme dan Realitas dalam Implementasi AI

Pada tahap awal pengenalan AI, antusiasme ASN, pegawai, teman-teman di koperasi dan UMKM akan tinggi karena melihat potensi peningkatan efisiensi, namun antusiasme ini sering berkurang saat beralih ke implementasi teknis yang konkret. Ketidaknyamanan muncul saat konsep abstrak AI harus diterjemahkan ke dalam pengumpulan dan analisis data nyata, ditambah dengan kurangnya pemahaman teknis dan kekhawatiran terkait keamanan data serta perubahan signifikan dalam proses kerja.

Tantangan teknis ini diperparah oleh perlunya perubahan budaya organisasi yang menuntut pergeseran dari cara kerja tradisional ke pendekatan berbasis data. Untuk mengatasi resistensi ini, diperlukan peningkatan literasi digital, pelatihan intensif, dan dukungan berkelanjutan agar ASN, koperasi dan UMKM lebih percaya diri dalam implementasi AI.

REFERENSI:

- Prihardani Ganda Tuah Purba. Mengenal Nodeflux, Start-up AI Asli Indonesia. DW.com (Accessed by 10 Agustus 2024, <https://www.dw.com/id/mengenal-nodeflux-start-up-ai-asli-indonesia/a/a-57456054>).
- "Strategi Nasional Kecerdasan Artifisial (Stranas KA)", BPPT, 2020 (Accessed 7 August 2024, <https://ai-innovation.id/images/gallery/ebook/stranas-ka.pdf>)
- Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition 2016 (pp. 2818-2826).
- Undang-undang (UU) Nomor 27 Tahun 2022 tentang Pelindungan Data Pribadi. KemenkopUKM Pusat, Indonesia, 2022(accessed by 10 August 2024, <https://peraturan.bpk.go.id/Details/229798/uu-no-27-tahun-2022>)

LAMPIRAN: MSE, RMSE, MAPE, MAR, dan R2

Mean Squared Error (MSE)

MSE adalah rata-rata kuadrat dari kesalahan prediksi. Rumusnya:

$$\text{MSE} = (1/n) * \sum (y_i - \hat{y}_i)^2$$

Di mana:

- n adalah jumlah data
- y_i adalah nilai aktual ke- i
- $\hat{y}_i$ adalah nilai prediksi ke- i

Nilai MSE yang lebih kecil menunjukkan model yang lebih akurat.

Root Mean Squared Error (RMSE)

RMSE adalah akar kuadrat dari MSE. Rumusnya:

$$\text{RMSE} = \sqrt{\text{MSE}}$$

RMSE memiliki satuan yang sama dengan variabel terprediksi. Nilai RMSE yang lebih kecil menunjukkan model yang lebih akurat.

Mean Absolute Error (MAR)

MAR adalah rata-rata dari nilai absolut kesalahan prediksi. Rumusnya:

$$\text{MAR} = (1/n) * \sum |y_i - \hat{y}_i|$$

Nilai MAR yang lebih kecil menunjukkan model yang lebih akurat.

Mean Absolute Percentage Error (MAPE)

MAPE adalah rata-rata dari nilai absolut persentase kesalahan prediksi.

Rumusnya:

$$\text{MAPE} = (1/n) * \sum(|y_i - \hat{y}_i| / y_i) * 100\%$$

MAPE berguna untuk membandingkan model pada variabel dengan skala yang berbeda. Nilai MAPE yang lebih kecil menunjukkan model yang lebih akurat.

Coefficient of Determination (R2)

R2 adalah koefisien determinasi yang menunjukkan proporsi varians total variabel terprediksi yang dapat dijelaskan oleh variabel independen.

Rumusnya:

$$R^2 = 1 - (\sum(y_i - \hat{y}_i)^2 / \sum(y_i - \hat{y}_m)^2)$$

Di mana:

- $\hat{y}_m$ adalah rata-rata dari nilai aktual

Nilai R2 yang lebih tinggi menunjukkan model yang lebih baik dalam menjelaskan varians variabel terprediksi.

Perbedaan Utama:

- MSE dan RMSE: Mengukur kesalahan kuadrat rata-rata, tetapi RMSE memiliki satuan yang sama dengan variabel terprediksi.
- MAR: Mengukur kesalahan absolut rata-rata.
- MAPE: Mengukur kesalahan absolut persentase rata-rata.
- R2: Mengukur proporsi varians total yang dijelaskan oleh model.

Pemilihan Metrik:

Metrik yang paling tepat untuk digunakan tergantung pada konteks dan tujuan analisis. Berikut adalah beberapa panduan umum:

- MSE dan RMSE: Cocok untuk model regresi di mana kesalahan kuadrat lebih penting.
- MAR: Cocok untuk model di mana kesalahan absolut lebih penting.
- MAPE: Cocok untuk model di mana variabel terprediksi memiliki skala yang berbeda.
- R2: Cocok untuk model regresi di mana tujuannya adalah untuk menjelaskan varians variabel terprediksi.

MSE, RMSE, MAR, MAPE, dan R2 adalah metrik yang digunakan untuk mengevaluasi performa model prediksi. Metrik yang paling tepat untuk digunakan tergantung pada konteks dan tujuan analisis.

Tentang Penulis

Onno W. Purbo, saat ini bertugas sebagai Rektor di Institut Teknologi Tangerang Selatan (ITTS). Onno memperoleh gelar Ph.D bidang Electrical Engineering dari University of Waterloo, Canada, adalah seorang copyleft, educator dan ICT evangelist. Dia sudah mempublikasikan 50+ buku, termasuk free ICT ebook untuk sekolah tahun 2008. Beberapa buku terakhirnya adalah "Internet-TCP/IP: Konsep Dan Implementasi", 2018; "Sistem Operasi, Konsep Dan Membuat Linux OpenWRT Dan ROM Android", 2019; "IPv6 Untuk Mendukung Operasi Jaringan Dan Domain Name System", 2019; "Kubernetes untuk Pemula", 2024 dan "Membuat Operator Seluler 5G Sendiri", 2024. Dia memimpin sambungan pertama Internet di Institut Teknologi Bandung, tahun 1993-2000, dan menggunakannya untuk membuat jaringan Internet pendidikan yang pertama di Indonesia. Dia membebaskan frekuensi WiFi, memperkenalkan RT/RW-net, antenna Wajanbolic dan jaringan selular OpenBTS. Dia memimpin jaringan telepon pertama di atas Internet, VoIP Merdeka, yang kemudian hari dikenal sebagai VoIP Rakyat berbasis SIP dan menggunakan kode area +62520 dan +62521. Dia saat ini aktif memperkenalkan e-Learning, dan menjalankan server e-Learning di <http://lms.onnocenter.or.id/moodle/> 65,000++ siswa / mahasiswa dan <https://opencourse.itts.ac.id> 25.000+ mahasiswa secara gratis.

Sinopsis

Buku ini merupakan panduan komprehensif yang mengajarkan penerapan praktis kecerdasan buatan (AI) dan pembelajaran mesin (*machine learning*) bagi koperasi dan UMKM di Indonesia. Mulai dari dasar-dasar AI, buku ini memberikan pemahaman mendalam tentang konsep, manfaat, dan tantangan dalam mengadopsi teknologi ini. Pembaca akan diajak melalui studi kasus konkret dan tutorial menggunakan tool open source seperti Orange Data Mining untuk menganalisis data, membuat prediksi, dan mengoptimalkan proses bisnis. Selain itu, buku ini juga membahas isu-isu penting seperti etika AI, privasi data, dan regulasi yang perlu diperhatikan dalam implementasi AI. Dengan gaya penulisan yang mudah dipahami dan contoh-contoh yang relevan, buku ini menjadi referensi berharga bagi siapa saja yang ingin memanfaatkan potensi AI untuk meningkatkan kinerja dan daya saing bisnisnya.