

KONSEP DASAR DATA MINING

MENGGUNAKAN APLIKASI RAPIDMINER DAN ORANGE

Cindy Claudia, Dr. Iwan Purnama S. Kom., M. Kom,
Rohani S.pd.i., M. Pd, Rahmadani Pane, S.Kom., M.Kom

KONSEP DASAR DATA MINING
MENGUNAKAN APLIKASI RAPIDMINER
DAN ORANGE

UU No 28 tahun 2014 tentang Hak Cipta

Fungsi dan sifat hak cipta Pasal 4

Hak Cipta sebagaimana dimaksud dalam Pasal 3 huruf a merupakan hak eksklusif yang terdiri atas hak moral dan hak ekonomi.

Pembatasan Pelindungan Pasal 26

Ketentuan sebagaimana dimaksud dalam Pasal 23, Pasal 24, dan Pasal 25 tidak berlaku terhadap:

- i. penggunaan kutipan singkat Ciptaan dan/atau produk Hak Terkait untuk pelaporan peristiwa aktual yang ditujukan hanya untuk keperluan penyediaan informasi aktual;
- ii. Penggandaan Ciptaan dan/atau produk Hak Terkait hanya untuk kepentingan penelitian ilmu pengetahuan;
- iii. Penggandaan Ciptaan dan/atau produk Hak Terkait hanya untuk keperluan pengajaran, kecuali pertunjukan dan Fonogram yang telah dilakukan Pengumuman sebagai bahan ajar; dan
- iv. penggunaan untuk kepentingan pendidikan dan pengembangan ilmu pengetahuan yang memungkinkan suatu Ciptaan dan/atau produk Hak Terkait dapat digunakan tanpa izin Pelaku Pertunjukan, Produser Fonogram, atau Lembaga Penyiaran.

Sanksi Pelanggaran Pasal 113

1. Setiap Orang yang dengan tanpa hak melakukan pelanggaran hak ekonomi sebagaimana dimaksud dalam Pasal 9 ayat (1) huruf i untuk Penggunaan Secara Komersial dipidana dengan pidana penjara paling lama 1 (satu) tahun dan/atau pidana denda paling banyak Rp100.000.000 (seratus juta rupiah).
2. Setiap Orang yang dengan tanpa hak dan/atau tanpa izin Pencipta atau pemegang Hak Cipta melakukan pelanggaran hak ekonomi Pencipta sebagaimana dimaksud dalam Pasal 9 ayat (1) huruf c, huruf d, huruf f, dan/atau huruf h untuk Penggunaan Secara Komersial dipidana dengan pidana penjara paling lama 3 (tiga) tahun dan/atau pidana denda paling banyak Rp500.000.000,00 (lima ratus juta rupiah).

KONSEP DASAR DATA MINING MENGUNAKAN APLIKASI RAPIDMINER DAN ORANGE

Cindy Claudia

Dr. Iwan Purnama S.Kom., M.Kom

Rohani S.Pd.I., M.Pd

Rahmadani Pane, S.Kom.,M.Kom



KONSEP DASAR DATA MINING MENGGUNAKAN APLIKASI RAPIDMINER DAN ORANGE

Cindy Claudia
Dr. Iwan Purnama S.Kom., M.Kom
Rohani S.Pd.I., M.Pd
Rahmadani Pane, S.Kom.,M.Kom

Editor :
Sahat Parulian Sitorus

Desain Cover :
Tubagus Rahardi

Sumber :
Link
<https://isbn.jndi.my.id/products/konsep-dasar-data-mining-menggunakan-aplikasi-rapidminer-dan-orange>

Tata Letak :
Nurhasanah

Ukuran :
Jml hal judul 112, Jml hal isi naskah 102, Uk: 14x21 cm

ISBN :
978-634-04-2315-0

Cetakan Pertama :
19 Agustus 2025

Hak Cipta 2024, Pada Penulis

Isi diluar tanggung jawab percetakan

Copyright © 2025 by PT.JNDI
All Right Reserved

Hak cipta dilindungi undang-undang
Dilarang keras menerjemahkan, memfotokopi, atau
memperbanyak sebagian atau seluruh isi buku ini
tanpa izin tertulis dari Penerbit.

PENERBIT PT. JASA NIAGA DIGITAL INDONESIA
Jl. H. Maulana Link. VI. Labuhan Ruku, Talawi, Batu Bara, Sumatera Utara 21254
Telp/Wa : 082361274081
<https://isbn.jndi.my.id>
<https://perpusnas.jndi.my.id>
E-mail: office@jndi.my.id

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa, karena berkat rahmat dan karunia-Nya, penulis dapat menyelesaikan buku ini yang berjudul "Konsep Dasar Menggunakan Aplikasi Rapidminer dan Orange". Buku ini disusun sebagai bagian dari tugas akhir dalam rangka menyelesaikan studi di Program Studi Teknologi Informasi. Universitas Labuhanbatu.

Buku ini disusun secara sistematis, dimulai dari pengenalan dasar data mining, sejarahnya, berbagai tipe data mining seperti klasifikasi, clustering, dan association rule, hingga pengenalan dan panduan penggunaan aplikasi RapidMiner dan Orange. Penulis berharap buku ini dapat menjadi panduan yang komprehensif bagi para pembaca, baik yang memiliki latar belakang data science maupun yang baru memulai.

Penulis

Cindy Claudia

DAFTAR ISI

| | |
|---|-----------|
| KATA PENGANTAR..... | vi |
| BAB I | 1 |
| PENGANTAR DATA MINING..... | 1 |
| BAB II | 11 |
| TIPE TIPE DATA MINING | 11 |
| •Decision Trees | 12 |
| •Naive Bayes | 14 |
| •Support Vector Machines (SVM)..... | 17 |
| •K-Means 21 | |
| ❖Fuzzy C-Mean | 25 |
| •Aturan Asosiasi Umum (Generalized Association Rules) | 28 |
| •Aturan Asosiasi Multilevel (Multilevel Association Rules) | 28 |
| •Aturan Multirelasional (Multirelational Association Rules) | 29 |
| •Aturan Asosiasi Klasifikasi (Classification Association Rules) | 29 |
| BAB III | 33 |
| PENGENALAN APLIKASI RAPIDMINNER | 33 |
| 1.Apa itu RapidMinner? | 33 |
| 2.Sejarah RapidMinner | 33 |
| 3.Instalasi Aplikasi RapidMiner | 39 |
| BAB IV | 45 |

| | |
|---|-----------|
| PENGENALAN APLIKASI ORANGE | 45 |
| 2. Sejarah Orange | 48 |
| 4. Instalisasi Aplikasi Orange | 54 |
| BAB V | 61 |
| PENERAPAN METODE APRIORI PADA PENJUALAN KOPI KELILING | 61 |
| 1. Pengumpulan Data | 61 |
| 2. Pre- Processing | 61 |
| 3. Data Selection | 62 |
| BAB VI | 70 |
| IMPLEMENTASI DATA MENGGUNAKAN APLIKASI RAPIDMINER | 70 |
| BAB VII | 79 |
| IMPLEMENTASI DATA MENGGUNAKAN APLIKASI ORANGE | 79 |
| DAFTAR PUSTAKA | 85 |
| TENTANG PENULIS | 85 |

BAB I

PENGANTAR DATA MINING

1. Pengertian Data mining

Data mining adalah proses yang digunakan untuk mengekstrak informasi yang berharga dari kumpulan data yang besar. Proses ini melibatkan penggunaan teknik statistik, algoritma pembelajaran mesin, dan metode analisis data untuk menemukan pola, tren, dan hubungan yang tersembunyi dalam data. Dengan kata lain, data mining bertujuan untuk mengubah data mentah menjadi informasi yang dapat digunakan untuk pengambilan keputusan yang lebih baik.

Secara sederhana, data mining dapat didefinisikan sebagai ekstraksi informasi atau pola penting dari data yang ada di dalam database yang besar. Dalam konteks ini, data mining sering kali disebut juga sebagai Knowledge Discovery in Databases (KDD), yang mencakup serangkaian langkah untuk menemukan pengetahuan dari data.

Proses ini tidak hanya terbatas pada data yang terstruktur, tetapi juga dapat diterapkan pada data tidak terstruktur, seperti teks, gambar, dan video.

Salah satu aspek penting dari data mining adalah kemampuannya untuk mengidentifikasi pola yang kompleks dalam data. Teknik-teknik yang digunakan dalam data mining, seperti klasifikasi, regresi, clustering, dan analisis asosiasi, memungkinkan analisis untuk memahami hubungan antara variabel dan membuat prediksi berdasarkan data yang ada. Misalnya, analisis asosiasi dapat digunakan untuk menemukan hubungan antara produk yang sering dibeli bersamaan oleh pelanggan.

Data mining juga memiliki berbagai aplikasi di berbagai bidang. Dalam bisnis, perusahaan menggunakan data mining untuk menganalisis perilaku pelanggan, mengoptimalkan strategi pemasaran, dan meningkatkan efisiensi operasional. Di bidang kesehatan, data mining digunakan untuk menganalisis data pasien dan meningkatkan perawatan kesehatan melalui identifikasi pola penyakit dan efektivitas pengobatan. Dengan

demikian, data mining berkontribusi pada inovasi dan peningkatan layanan di berbagai sektor.

Proses data mining biasanya terdiri dari beberapa tahap, termasuk pemilihan data, pra-pemrosesan, transformasi, dan evaluasi hasil. Setiap tahap memiliki peran penting dalam memastikan bahwa informasi yang diperoleh akurat dan relevan. Misalnya, pada tahap pra-pemrosesan, data yang tidak lengkap atau tidak akurat dapat dibersihkan dan disiapkan untuk analisis lebih lanjut. Hal ini sangat penting untuk menghasilkan hasil yang dapat diandalkan.

Meskipun data mining menawarkan banyak manfaat, ada juga tantangan yang perlu diatasi. Salah satu tantangan utama adalah masalah privasi dan keamanan data. Dengan meningkatnya jumlah data yang dikumpulkan, penting bagi organisasi untuk memastikan bahwa data tersebut dikelola dengan cara yang etis dan sesuai dengan regulasi yang berlaku. Selain itu, kualitas data juga menjadi perhatian utama, karena data yang tidak

akurat atau tidak lengkap dapat menghasilkan analisis yang menyesatkan..

Dalam konteks teknologi yang terus berkembang, data mining juga berhubungan erat dengan konsep big data dan kecerdasan buatan. Big data merujuk pada volume data yang sangat besar dan kompleks yang tidak dapat dikelola dengan metode tradisional. Sementara itu, kecerdasan buatan dan pembelajaran mesin menyediakan alat dan teknik yang lebih canggih untuk menganalisis data dalam skala besar. Dengan memanfaatkan teknologi ini, organisasi dapat meningkatkan kemampuan mereka dalam mengekstrak informasi berharga dari data.

Data mining juga berperan penting dalam pengambilan keputusan berbasis data. Dengan informasi yang diperoleh dari proses data mining, organisasi dapat membuat keputusan yang lebih tepat dan strategis. Misalnya, analisis data dapat membantu perusahaan dalam meramalkan tren pasar, memahami kebutuhan pelanggan, dan mengidentifikasi peluang baru. Hal ini sangat penting dalam lingkungan bisnis yang kompetitif saat ini.

Secara keseluruhan, data mining adalah alat yang sangat berharga dalam dunia yang semakin didorong oleh data. Dengan kemampuan untuk mengubah data mentah menjadi informasi yang berguna, data mining membantu organisasi dalam meningkatkan efisiensi, inovasi, dan pengambilan keputusan. Oleh karena itu, pemahaman yang baik tentang data mining dan teknik-tekniknya sangat penting bagi para profesional di berbagai bidang.

Dengan demikian, buku ini bertujuan untuk memberikan pemahaman yang mendalam tentang konsep, teknik, dan aplikasi data mining. Pembaca diharapkan dapat menerapkan pengetahuan ini dalam praktik mereka, serta memahami tantangan dan peluang yang ada dalam dunia data mining yang terus berkembang.

1.2 Sejarah Data Mining

Sejarah data mining dimulai pada akhir tahun 1980-an, ketika istilah ini pertama kali diperkenalkan dalam konteks pengolahan data. Sebelum istilah "data mining" menjadi populer, proses ini lebih dikenal dengan nama KDD (Knowledge Discovery in Databases). Pada masa itu,

fokus utama adalah pada pengumpulan dan penyimpanan data, sementara kemampuan untuk menganalisis dan mengekstrak pengetahuan dari data tersebut masih sangat terbatas. Dengan kemajuan teknologi informasi, terutama pada tahun 1990-an, banyak organisasi mulai mengumpulkan data dalam jumlah besar, tetapi mereka kesulitan untuk menganalisis data tersebut secara efektif.

Pada tahun 1996, sekelompok perusahaan seperti Teradata dan NCR berkolaborasi untuk menstandarisasi teknik-teknik dalam data mining. Hasil dari kolaborasi ini adalah Proses Standar Lintas Industri untuk Penambangan Data (CRISP-DM), yang menjadi acuan penting dalam praktik data mining. CRISP-DM memberikan kerangka kerja yang sistematis untuk mengembangkan proyek data mining, mulai dari pemahaman bisnis hingga evaluasi hasil. Dengan adanya standar ini, praktik data mining menjadi lebih terstruktur dan dapat diadopsi oleh berbagai organisasi.

Memasuki tahun 2000-an, perusahaan-perusahaan web mulai menyadari potensi data mining dalam mengolah data untuk mendapatkan wawasan yang berharga. Pada saat

ini, data mining tidak hanya terbatas pada data tabular, tetapi juga mulai mencakup berbagai jenis data lainnya, seperti teks, gambar, dan video. Perkembangan ini didorong oleh kemajuan dalam teknologi komputasi yang memungkinkan analisis data dalam skala besar dan kompleksitas yang lebih tinggi..

Salah satu tonggak penting dalam sejarah data mining adalah kontribusi dari para ahli seperti Usama Fayyad, Gregory Piatetsky-Shapiro, dan Padhraic Smyth, yang merupakan pelopor dalam penggunaan teknik pembelajaran mesin untuk menemukan pola dalam dataset besar. Mereka membantu mengembangkan alat dan metodologi yang menjadi dasar bagi praktik data mining modern. Selain itu, para ahli statistik seperti Leo Breiman dan Jerome Friedman juga memberikan kontribusi signifikan dalam pengembangan teknik-teknik yang digunakan dalam data mining, seperti pohon keputusan dan algoritma klasifikasi.

Popularitas data mining terus meningkat pada tahun 1990-an, dengan banyak konferensi yang didedikasikan untuk pengembangan di bidang ini. Konferensi seperti

ACM SIGKDD dan IEEE ICDM menjadi platform penting bagi para ahli untuk berbagi temuan dan inovasi terbaru dalam data mining. Pertumbuhan ini didorong oleh kemajuan teknologi, CPU yang berlimpah, dan media untuk menyimpan data dalam jumlah besar serta memprosesnya dengan cepat. Banyak perusahaan mulai menyadari bahwa data yang mereka kumpulkan dapat "ditambang" untuk mendapatkan wawasan yang dapat meningkatkan operasi bisnis mereka.

Seiring dengan perkembangan teknologi, data mining juga mulai diterapkan dalam berbagai sektor, termasuk keuangan, kesehatan, dan pemasaran. Dalam sektor keuangan, misalnya, data mining digunakan untuk mendeteksi penipuan dan menganalisis risiko. Di bidang kesehatan, teknik ini digunakan untuk menganalisis data pasien dan meningkatkan perawatan kesehatan. Dalam pemasaran, perusahaan menggunakan data mining untuk memahami perilaku konsumen dan merancang strategi pemasaran yang lebih efektif.

Pada tahun 2000-an, munculnya big data membawa tantangan dan peluang baru bagi data mining. Dengan

volume data yang terus meningkat, teknik dan algoritma yang ada harus disesuaikan untuk menangani data dalam skala yang lebih besar dan lebih kompleks. Ini mendorong pengembangan teknik baru dalam data mining, termasuk algoritma yang lebih efisien dan metode analisis yang lebih canggih. Selain itu, munculnya teknologi seperti cloud computing juga memungkinkan organisasi untuk menyimpan dan memproses data dalam jumlah besar dengan lebih mudah.

Saat ini, data mining telah menjadi bagian integral dari strategi bisnis banyak organisasi. Dengan kemampuan untuk mengekstrak informasi berharga dari data yang tidak terstruktur, data mining telah menjadi alat yang esensial dalam dunia bisnis dan pengamatan . Organisasi di berbagai sektor mulai memanfaatkan data mining untuk meningkatkan pengambilan keputusan, mengoptimalkan operasi, dan mendapatkan keunggulan kompetitif. Dengan terus berkembangnya teknologi dan kebutuhan analisis data yang semakin kompleks, masa depan data mining tampak cerah.

Dalam beberapa tahun terakhir, perkembangan teknologi seperti kecerdasan buatan (AI) dan pembelajaran mesin (machine learning) telah semakin memperkuat posisi data mining. Teknik-teknik ini memungkinkan analisis data yang lebih mendalam dan akurat, serta memberikan wawasan yang lebih berharga bagi organisasi. Dengan memanfaatkan AI dan machine learning, perusahaan dapat mengidentifikasi pola dan tren yang sebelumnya tidak terlihat, sehingga meningkatkan kemampuan mereka dalam mengambil keputusan yang lebih baik.

Secara keseluruhan, sejarah data mining mencerminkan evolusi teknologi informasi dan kebutuhan organisasi untuk mengelola dan menganalisis data dalam jumlah besar. Dari awal yang sederhana sebagai subproses dalam pengolahan data hingga menjadi disiplin ilmu yang kompleks dan terintegrasi, data mining terus berkembang.

BAB II

TIPE TIPE DATA MINING

1. Klasifikasi

Klasifikasi dalam data mining adalah salah satu teknik yang paling penting dan banyak digunakan untuk mengelompokkan data ke dalam kategori yang telah ditentukan sebelumnya. Proses ini melibatkan penggunaan algoritma untuk membangun model yang dapat memprediksi kelas dari data baru berdasarkan pola yang ditemukan dalam data pelatihan. Klasifikasi sangat berguna dalam berbagai aplikasi, mulai dari pengenalan wajah, diagnosis medis, hingga deteksi spam dalam email. Dengan kata lain, klasifikasi membantu dalam pengambilan keputusan yang lebih baik berdasarkan data yang ada.

Salah satu aspek kunci dari klasifikasi adalah pemisahan data menjadi dua set: data pelatihan dan data pengujian. Data pelatihan digunakan untuk membangun model klasifikasi, sedangkan data pengujian digunakan untuk mengevaluasi kinerja model tersebut. Proses ini penting untuk memastikan bahwa model yang dibangun

tidak hanya mampu mengenali pola dalam data pelatihan, tetapi juga dapat menggeneralisasi dengan baik pada data yang belum pernah dilihat sebelumnya. Evaluasi kinerja model biasanya dilakukan dengan menggunakan metrik seperti akurasi, precision, recall, dan F1-score.

Ada berbagai algoritma yang digunakan dalam klasifikasi, masing-masing dengan pendekatan dan keunggulan tersendiri. Beberapa algoritma yang umum digunakan termasuk :

- **Decision Trees**

Metode ini membangun model dalam bentuk struktur pohon, di mana setiap cabang mewakili keputusan berdasarkan atribut tertentu, dan setiap daun mewakili hasil akhir atau kelas dari data. Proses pembuatan Decision Tree dimulai dengan memecah dataset menjadi subset yang lebih kecil berdasarkan kriteria tertentu, sehingga memudahkan dalam pengambilan keputusan. Dengan cara ini, Decision Trees dapat memberikan gambaran yang jelas dan mudah dipahami mengenai bagaimana keputusan diambil berdasarkan data yang ada.

Salah satu keunggulan utama dari Decision Trees adalah kemampuannya untuk menangani data dengan berbagai jenis atribut, baik itu atribut kategorikal maupun numerik. Decision Trees juga memiliki kemampuan untuk melakukan eksplorasi data, yang memungkinkan pengguna untuk menemukan hubungan antara variabel input dan variabel target.

Hal ini menjadikan Decision Trees sebagai alat yang sangat berguna dalam analisis data, terutama ketika pengguna ingin memahami faktor-faktor yang mempengaruhi hasil tertentu. Beberapa algoritma yang dapat digunakan :

1. ID3 Algoritma yang digunakan untuk membangun pohon keputusan dengan cara memilih atribut yang paling informatif untuk membagi dataset.
2. C4.5 Menggunakan konsep gain ratio untuk memilih atribut yang paling informatif, yang membantu mengurangi bias terhadap atribut dengan banyak nilai.

3. CART : Algoritma ini menggunakan Gini impurity atau mean squared error (MSE) sebagai kriteria pemisahan untuk menentukan atribut yang paling baik dalam membagi data.

- **Naive Bayes**

Naive Bayes adalah metode klasifikasi yang sangat populer dalam data mining, yang didasarkan pada Teorema Bayes. Metode ini digunakan untuk memprediksi kelas dari suatu data berdasarkan probabilitas yang dihitung dari fitur-fitur yang ada. Keunikan dari Naive Bayes terletak pada asumsi independensi antar fitur, yang berarti bahwa setiap fitur dianggap tidak saling mempengaruhi satu sama lain dalam menentukan kelas. Meskipun asumsi ini sering kali tidak realistis dalam praktik, Naive Bayes tetap menunjukkan performa yang baik dalam banyak aplikasi, terutama dalam klasifikasi teks seperti pemfilteran spam dan analisis sentimen.

Terdapat beberapa tipe algoritma Naive Bayes yaitu :

- Bernoulli Naive Bayes

Digunakan untuk klasifikasi data dengan fitur biner. Dalam model ini, setiap fitur dianggap sebagai variabel Boolean yang menunjukkan apakah suatu kata muncul atau tidak dalam dokumen. Algoritma ini sangat cocok untuk aplikasi yang melibatkan teks pendek, seperti klasifikasi email spam, di mana kehadiran atau ketidakhadiran kata-kata tertentu menjadi indikator penting untuk menentukan kelas dokumen. Bernoulli Naive Bayes secara eksplisit memodelkan ketidakhadiran kata, yang membedakannya dari tipe lain.

- Multinomial Naive Bayes

Digunakan untuk data diskrit dan sering diterapkan dalam klasifikasi dokumen. Dalam model ini, fitur merepresentasikan frekuensi kemunculan kata dalam dokumen, sehingga cocok untuk analisis teks yang melibatkan banyak kata. Multinomial Naive Bayes

menghitung probabilitas berdasarkan jumlah kemunculan kata dalam setiap kelas, menjadikannya efektif untuk tugas-tugas seperti pengelompokan dokumen dan pemfilteran spam. Algoritma ini juga dapat menggunakan teknik seperti tf-idf untuk meningkatkan akurasi klasifikasi

- Gaussian Naive Bayes

Digunakan ketika fitur bersifat kontinu dan diasumsikan mengikuti distribusi normal (Gaussian). Dalam model ini, setiap fitur yang kontinu diperlakukan dengan menghitung rata-rata dan varians dari fitur tersebut untuk setiap kelas. Gaussian Naive Bayes sering digunakan dalam aplikasi di mana data memiliki atribut numerik, seperti dalam analisis data medis atau pengenalan pola. Meskipun asumsi distribusi normal tidak selalu terpenuhi, algoritma ini tetap dapat memberikan hasil yang baik dalam banyak situasi

Salah satu keunggulan dari Naive Bayes adalah kemampuannya untuk bekerja dengan baik meskipun dengan jumlah data pelatihan yang relatif kecil. Algoritma ini juga sangat efisien dalam hal waktu komputasi dan penggunaan memori, sehingga cocok untuk aplikasi yang

memerlukan kecepatan tinggi dalam pengolahan data. Selain itu, Naive Bayes dapat menangani data yang tidak seimbang dengan baik, yang sering terjadi dalam banyak situasi dunia nyata, seperti dalam klasifikasi email spam di mana jumlah email spam jauh lebih sedikit dibandingkan dengan email yang bukan spam

- **Support Vector Machines (SVM)**

Support Vector Machines (SVM) adalah metode dalam pembelajaran mesin yang digunakan untuk klasifikasi dan regresi. SVM bekerja dengan mencari hyperplane terbaik yang dapat memisahkan data ke dalam dua kelas atau lebih. Hyperplane ini adalah batas keputusan yang memisahkan ruang fitur menjadi dua bagian, di mana setiap bagian mewakili kelas yang berbeda. SVM berusaha untuk memaksimalkan margin antara hyperplane dan titik data terdekat dari masing-masing kelas, yang dikenal sebagai support vectors. Dengan memaksimalkan margin ini, SVM bertujuan untuk meningkatkan akurasi klasifikasi dan generalisasi model terhadap data baru.

SVM dapat digunakan untuk menangani masalah klasifikasi linear dan non-linear. Untuk data yang tidak

dapat dipisahkan secara linear, SVM menggunakan teknik kernel untuk mentransformasikan data ke dalam ruang dimensi yang lebih tinggi, sehingga memungkinkan pemisahan yang lebih baik. Beberapa fungsi kernel yang umum digunakan adalah kernel linear, polynomial, dan Radial Basis Function (RBF). Dengan menggunakan kernel, SVM dapat mengatasi berbagai jenis data dan meningkatkan performa klasifikasi..

Salah satu keunggulan SVM adalah kemampuannya untuk bekerja dengan baik pada dataset dengan dimensi tinggi, serta efisiensinya dalam menangani jumlah fitur yang lebih besar dibandingkan dengan jumlah sampel. Namun, SVM juga memiliki beberapa kelemahan, seperti waktu pelatihan yang lama pada dataset yang sangat besar dan sensitivitas terhadap noise dalam data. Oleh karena itu, pemilihan parameter yang tepat, seperti nilai C dan jenis kernel, sangat penting untuk mencapai hasil yang optimal.

Secara keseluruhan, SVM adalah alat yang kuat dalam analisis data dan telah banyak diterapkan dalam berbagai bidang, termasuk pengenalan pola, klasifikasi teks,

dan analisis citra. Dengan pemahaman yang baik tentang cara kerja dan penerapan SVM, pengguna dapat memanfaatkan metode ini untuk meningkatkan akurasi dan efisiensi dalam pengolahan data. Klasifikasi juga dapat diterapkan dalam berbagai domain:

- ◆ Dalam Bidang Kesehatan : Klasifikasi dapat membantu dokter dalam mendiagnosis penyakit berdasarkan gejala yang ditunjukkan oleh pasien.
- ◆ Dalam Pemasaran : Teknik ini digunakan untuk mengidentifikasi segmen pelanggan yang berbeda berdasarkan perilaku pembelian mereka.
- ◆ Deteksi Penipuan : Di mana transaksi yang mencurigakan dapat diidentifikasi dan ditandai untuk ditinjau lebih lanjut.

Dalam praktiknya, klasifikasi sering kali melibatkan beberapa langkah preprocessing data, seperti normalisasi, penghapusan outlier, dan pengisian nilai yang hilang. Langkah-langkah ini bertujuan untuk meningkatkan kualitas data dan memastikan bahwa model klasifikasi dapat berfungsi dengan baik. Selain itu, pemilihan fitur

yang tepat juga sangat penting, karena fitur yang tidak relevan dapat mengganggu proses klasifikasi dan menurunkan akurasi model.

2. Clustering

Clustering adalah teknik dalam data mining yang digunakan untuk mengelompokkan data ke dalam kelompok-kelompok yang memiliki kesamaan. Tujuan utama dari clustering adalah untuk menemukan struktur dalam data tanpa adanya label atau kategori yang telah ditentukan sebelumnya. Dengan mengelompokkan data, kita dapat mengidentifikasi pola, hubungan, dan karakteristik yang ada dalam dataset. Clustering banyak digunakan dalam berbagai aplikasi, seperti segmentasi pasar, pengenalan pola, dan analisis citra, yang memungkinkan pengambilan keputusan yang lebih baik berdasarkan informasi yang diperoleh dari data.

Beberapa algoritma clustering yang umum digunakan dalam data mining:

- **K-Means**

K-Means adalah salah satu algoritma clustering yang paling populer dan banyak digunakan. Algoritma ini bekerja dengan cara membagi data menjadi sejumlah cluster yang telah ditentukan sebelumnya. Proses dimulai dengan memilih k titik pusat (centroid) secara acak, kemudian setiap data akan dialokasikan ke cluster yang memiliki jarak terdekat dengan centroid tersebut. Setelah semua data dialokasikan, centroid dihitung ulang berdasarkan rata-rata posisi data dalam cluster. Proses ini diulang hingga tidak ada perubahan signifikan pada posisi centroid atau alokasi data ke cluster.

Kelebihan dari K-Means adalah kesederhanaannya dan kecepatan dalam proses pengelompokan, sehingga sangat efisien untuk dataset besar. Namun, algoritma ini memiliki beberapa kelemahan, seperti ketergantungan pada jumlah cluster yang harus ditentukan sebelumnya dan sensitivitas terhadap pemilihan awal centroid. Jika centroid dipilih secara tidak tepat, hasil clustering dapat bervariasi

secara signifikan. Selain itu, K-Means juga kurang efektif dalam menangani cluster dengan bentuk yang tidak bulat atau ukuran yang sangat berbeda

- **Dbscan**

Dbscan adalah algoritma clustering berbasis kepadatan yang efektif dalam mengidentifikasi cluster dengan bentuk yang tidak teratur. Algoritma ini bekerja dengan cara mengelompokkan titik data yang memiliki kepadatan tinggi dan memisahkan titik-titik yang dianggap noise. Dbscan tidak memerlukan jumlah cluster yang ditentukan sebelumnya, melainkan menggunakan dua parameter: radius (epsilon) dan jumlah minimum titik dalam radius tersebut (minPts). Jika jumlah titik dalam radius lebih besar dari minPts, maka titik tersebut dianggap sebagai bagian dari cluster.

Salah satu keunggulan Dbscan adalah kemampuannya untuk menemukan cluster dengan bentuk yang kompleks dan ukuran yang bervariasi. Selain itu, algoritma ini juga dapat mengidentifikasi outlier atau noise dalam data. Namun, kelemahan dari Dbscan adalah sensitivitas terhadap parameter yang digunakan. Jika nilai epsilon dan minPts tidak ditentukan dengan baik, hasil clustering dapat menjadi

tidak akurat. Selain itu, Dbscan juga kurang efektif jika terdapat variasi kepadatan yang signifikan di dalam dataset

❖ *Hierarchical Clustering*

Hierarchical Clustering adalah metode yang membangun struktur hirarkis dari data, yang dapat berupa agglomerative (penggabungan) atau divisive (pemisahan). Dalam pendekatan agglomerative, setiap data dimulai sebagai cluster individu dan kemudian digabungkan secara bertahap berdasarkan kedekatan. Sebaliknya, pendekatan divisive memulai dengan satu cluster besar dan membaginya menjadi beberapa cluster yang lebih kecil. Hasil dari hierarchical clustering dapat divisualisasikan dalam bentuk dendrogram, yang menunjukkan hubungan antar cluster.

Kelebihan dari hierarchical clustering adalah fleksibilitas dalam menentukan jumlah cluster dan kemampuannya untuk memberikan informasi yang lebih mendalam tentang struktur data. Namun,

kelemahannya adalah kompleksitas komputasi yang tinggi, terutama untuk dataset besar, karena algoritma ini

memerlukan perhitungan jarak antara semua pasangan data. Selain itu, hasil clustering dapat dipengaruhi oleh metode pengukuran jarak yang digunakan, seperti single linkage atau complete linkage.

❖ **Gaussian Mixture Models (GMM)**

Gaussian Mixture Models adalah algoritma clustering berbasis model yang mengasumsikan bahwa data berasal dari beberapa distribusi Gaussian yang berbeda. Gaussian Mixture Models menggunakan pendekatan probabilistik untuk menentukan kemungkinan setiap data termasuk dalam cluster tertentu. Proses ini melibatkan estimasi parameter distribusi Gaussian, seperti rata-rata dan kovarians, untuk setiap cluster. Gaussian Mixture Models dapat menangani data yang memiliki distribusi yang lebih kompleks dibandingkan dengan metode clustering lainnya.

Salah satu keunggulan Gaussian Mixture Models adalah kemampuannya untuk menangani cluster yang memiliki bentuk elips dan ukuran yang berbeda. Selain itu, Gaussian Mixture Models juga dapat memberikan informasi tentang derajat keanggotaan setiap data dalam cluster, yang memungkinkan analisis yang lebih mendalam.

Namun, kelemahan dari Gaussian Mixture Models adalah kebutuhan untuk estimasi parameter yang tepat dan sensitivitas terhadap inisialisasi awal. Jika parameter tidak diestimasi dengan baik, hasil clustering dapat menjadi tidak akurat.

❖ **Fuzzy C-Mean**

Fuzzy C-Means adalah algoritma clustering yang memungkinkan setiap data untuk menjadi bagian dari lebih dari satu cluster dengan derajat keanggotaan yang berbeda. Dalam metode ini, setiap data memiliki nilai keanggotaan untuk setiap cluster, yang menunjukkan seberapa besar kemungkinan data tersebut termasuk dalam cluster tersebut. Proses clustering dilakukan dengan cara meminimalkan fungsi objektif yang mempertimbangkan jarak antara data dan centroid cluster serta derajat keanggotaan.

Kelebihan dari Fuzzy C-Means adalah kemampuannya untuk menangani data yang memiliki batasan cluster yang tidak jelas, sehingga memberikan hasil yang lebih realistis dalam banyak aplikasi, seperti pengenalan pola dan segmentasi citra. Namun, kelemahan

dari metode ini adalah kompleksitas komputasi yang lebih tinggi dibandingkan dengan algoritma clustering tradisional, serta sensitivitas terhadap parameter yang digunakan, seperti nilai fuzziness.

Secara keseluruhan, clustering merupakan alat yang sangat berguna dalam data mining yang memungkinkan analisis data yang lebih mendalam dan pengambilan keputusan yang lebih baik. Dengan berbagai tipe dan metode yang tersedia, pengguna dapat memilih pendekatan yang paling sesuai dengan kebutuhan analisis mereka, sehingga dapat memanfaatkan data yang ada untuk mencapai tujuan yang diinginkan. Dengan pemahaman yang baik tentang berbagai metode clustering, organisasi dapat meningkatkan efektivitas analisis data mereka dan mendapatkan wawasan yang lebih berharga dari informasi yang tersedia.

3. Association Rule

Association rules dalam data mining adalah teknik yang digunakan untuk menemukan hubungan atau pola yang menarik antara item dalam dataset besar. Aturan ini sering digunakan dalam analisis pasar, di mana tujuan utamanya adalah untuk mengidentifikasi item yang sering dibeli bersamaan oleh pelanggan. Misalnya, jika banyak pelanggan yang membeli roti juga membeli mentega, maka dapat dihasilkan aturan asosiasi yang menyatakan bahwa "Jika seseorang membeli roti, maka kemungkinan besar mereka juga akan membeli mentega." Aturan ini dapat membantu dalam pengambilan keputusan bisnis, seperti penempatan produk di toko atau strategi pemasaran. Terdapat beberapa tipe Association Rule yaitu :

- **Aturan Asosiasi Umum (Generalized Association Rules)**

Tipe ini memberikan gambaran umum tentang hubungan antara item tanpa mempertimbangkan detail spesifik. Aturan ini berguna untuk mendapatkan wawasan awal mengenai pola pembelian pelanggan dan dapat digunakan sebagai dasar untuk analisis lebih lanjut.

- **Aturan Asosiasi Multilevel (Multilevel Association Rules)**

Aturan ini memisahkan data menjadi beberapa tingkat penting, mengelompokkan item berdasarkan kategori yang lebih umum dan lebih spesifik. Misalnya, dalam analisis penjualan, kategori produk seperti "minuman" dapat mencakup subkategori seperti "jus" dan "soda". Ini memungkinkan analisis yang lebih mendalam dengan mempertimbangkan konteks yang lebih luas.

- **Aturan Asosiasi Kuantitatif (Quantitative Association Rules)**

Tipe ini berfokus pada hubungan antara data numerik. Aturan ini digunakan untuk menggambarkan asosiasi antara item yang memiliki nilai numerik, seperti harga atau jumlah pembelian. Contohnya, aturan ini dapat menunjukkan

bahwa "Jika pelanggan membeli lebih dari 10 unit produk X, maka mereka cenderung membeli produk Y dengan harga di atas \$20."

- **Aturan Asosiasi Multirelasional (Multirelational Association Rules)**

Tipe ini lebih kompleks dan mengeksplorasi hubungan antara item di berbagai basis data atau dimensi. Ini memungkinkan analisis yang lebih mendalam dengan mempertimbangkan interaksi antara berbagai entitas dalam dataset, seperti dalam analisis data kesehatan untuk menemukan hubungan antara gejala penyakit dan faktor risiko.

- **Aturan Asosiasi Klasifikasi (Classification Association Rules)**

Tipe ini digunakan untuk mengklasifikasikan data berdasarkan pola yang ditemukan dalam dataset. Aturan ini membantu dalam mengidentifikasi kategori atau kelas dari item berdasarkan atribut yang ada, sehingga dapat digunakan untuk prediksi dan pengambilan keputusan.

beberapa algoritma yang umum digunakan dalam Associatin Rule.

❖ **AprioriAlgorithm:**

Algoritma ini adalah salah satu yang paling terkenal dalam menemukan itemset yang sering muncul dalam dataset. Apriori bekerja dengan prinsip bahwa jika suatu itemset adalah sering, maka semua subset dari itemset tersebut juga harus sering. Algoritma ini menggunakan dua langkah utama: pertama, menemukan semua itemset yang sering dengan menggunakan threshold support, dan kedua, menghasilkan aturan asosiasi dari itemset yang ditemukan dengan menggunakan threshold confidence.

❖ **EclatAlgorithm:**

Eclat adalah algoritma lain yang digunakan untuk menemukan itemset yang sering, tetapi berbeda dengan Apriori, Eclat menggunakan pendekatan depth-first search dan struktur data yang lebih efisien. Algoritma ini menghitung support dari itemset dengan menggunakan intersection dari transaksi yang mengandung itemset tersebut, sehingga dapat lebih cepat dalam menemukan itemset yang sering dibandingkan dengan metode Apriori.

❖ FP-GrowthAlgorithm:

FP-Growth (Frequent Pattern Growth) adalah metode yang lebih efisien dibandingkan dengan Apriori dan Eclat. FP-Growth tidak memerlukan generasi kandidat itemset, melainkan membangun struktur data yang disebut FP-tree. Dengan menggunakan FP-tree, algoritma ini dapat menemukan itemset yang sering dengan lebih cepat dan efisien, terutama pada dataset yang besar.

❖ RAP(RapidAssociationPattern):

RAP adalah algoritma yang dirancang untuk meningkatkan kecepatan dalam menemukan pola asosiasi. Algoritma ini menggunakan teknik pengurangan data untuk mempercepat proses pencarian pola asosiasi, sehingga dapat digunakan pada dataset yang sangat besar dengan lebih efisien.

❖ ClosetAlgorithm:

Algorithm adalah algoritma yang digunakan untuk menemukan itemset yang sering dan menghasilkan aturan asosiasi.Closet Algorithm

menggunakan pendekatan yang menggabungkan teknik pruning untuk mengurangi jumlah itemset yang perlu diproses, sehingga meningkatkan efisiensi dalam menemukan pola asosiasi.

Secara keseluruhan, Association Rule merupakan alat yang sangat berguna dalam data mining yang memungkinkan analisis data yang lebih mendalam dan pengambilan keputusan yang lebih baik. Dengan berbagai tipe aturan asosiasi yang tersedia, pengguna dapat memilih pendekatan yang paling sesuai dengan kebutuhan analisis mereka, sehingga dapat memanfaatkan data yang ada untuk mencapai tujuan yang diinginkan. Pemahaman yang baik tentang berbagai tipe aturan asosiasi dapat membantu organisasi dalam meningkatkan efektivitas strategi pemasaran dan pengelolaan produk mereka.

BAB III

PENGENALAN APLIKASI RAPIDMINNER

1. Apa itu RapidMiner?

RapidMiner adalah platform perangkat lunak yang dirancang untuk analisis data, data mining, dan pembelajaran mesin (machine learning). Dikenal karena antarmuka pengguna yang intuitif dan fitur drag-and-drop, RapidMiner memudahkan pengguna dalam melakukan analisis data tanpa perlu menulis kode secara manual. Platform ini mendukung berbagai tugas analisis data, mulai dari persiapan data hingga penerapan model prediktif, dan cocok untuk pengguna dari berbagai latar belakang, baik pemula maupun ahli di bidang data science

2. Sejarah RapidMiner

RapidMiner, alat yang kini menjadi rujukan dalam dunia ilmu data, memiliki perjalanan yang menarik sejak awal kemunculannya. Pengembangan RapidMiner bermula pada awal abad ke-21, tepatnya tahun 2001. Kala itu, tiga para ahli dari Unit Kecerdasan Buatan Universitas Teknik Dortmund, Ralf Klinkenberg, Ingo Mierswa,

dan Simon Fischer, merintis proyek ambisius dengan nama YALE (Yet Another Learning Environment).

Tujuan utama di balik pengembangan YALE adalah menciptakan lingkungan pembelajaran yang fleksibel dan user-friendly untuk berbagai tugas dalam bidang data mining. Dengan semangat untuk menyederhanakan proses analisis data yang kompleks, para ahli ini merancang sebuah platform yang memungkinkan pengguna, baik pemula maupun ahli, untuk melakukan berbagai macam analisis data dengan mudah.

Seiring berjalannya waktu, YALE terus berkembang dan mendapatkan perhatian yang semakin besar dari komunitas ilmu data. Pada tahun 2006, Ingo Mierswa dan Ralf Klinkenberg mendirikan perusahaan bernama Rapid-I dengan fokus utama pada pengembangan YALE. Langkah ini menjadi titik balik penting dalam sejarah RapidMiner.

Perubahan nama menjadi RapidMiner pada tahun 2007 menandai babak baru dalam perjalanan platform ini. Nama baru tersebut lebih mencerminkan tujuan utama RapidMiner yaitu menyediakan solusi analisis data yang cepat dan efisien. Perubahan nama ini juga sejalan dengan

visi perusahaan untuk menjadi pemimpin dalam industri perangkat lunak ilmu data.

RapidMiner terus mengalami pertumbuhan yang pesat dan mendapatkan pengakuan luas di kalangan para praktisi data. Platform ini berhasil memikat hati pengguna dengan antarmuka yang intuitif, fleksibilitas dalam menangani berbagai jenis data dan algoritma, serta komunitas pengguna yang aktif. Komunitas ini berperan penting dalam pengembangan RapidMiner dengan saling berbagi pengetahuan, pengalaman, dan kontribusi dalam bentuk ekstensi yang memperkaya fitur-fitur platform.

Pada tahun 2013, perusahaan Rapid-I melakukan rebranding dan mengganti namanya menjadi RapidMiner. Langkah ini semakin memperkuat posisi RapidMiner sebagai merek yang identik dengan platform analisis data terkemuka. Rebranding ini juga menunjukkan komitmen perusahaan untuk terus berinovasi dan memenuhi kebutuhan yang terus berkembang dalam dunia ilmu data.

Sejak saat itu, RapidMiner telah menjadi alat yang sangat populer di berbagai industri. Platform ini digunakan oleh para ahli, akademisi, perusahaan besar, hingga startup untuk memecahkan berbagai masalah bisnis yang

kompleks. RapidMiner telah berkontribusi dalam demokratisasi ilmu data, memungkinkan lebih banyak orang untuk mengakses dan memanfaatkan kekuatan data. Dengan fitur-fitur yang terus berkembang dan komunitas yang solid, RapidMiner diyakini akan terus menjadi pemain utama dalam lanskap perangkat lunak ilmu data di masa depan.

3. **Antar Muka Pengguna RapidMiner**

RapidMiner dikenal dengan antarmuka yang user-friendly, membuatnya dapat diakses oleh berbagai kalangan, mulai dari data scientist berpengalaman hingga mereka yang baru memulai petualangan di dunia data. Berikut adalah beberapa fitur visual yang membuat proses analisis data di RapidMiner menjadi lebih intuitif:

- **Drag-and-Drop Interface:**

Salah satu fitur paling menonjol dari RapidMiner adalah antarmuka drag-and-drop. Pengguna dapat dengan mudah membangun alur kerja (workflow) dengan menyeret dan melepaskan operator-operator yang mewakili berbagai tugas analisis data, seperti membaca data, membersihkan data, membangun

model, dan mengevaluasi hasil. Ini membuat proses membangun model menjadi sangat visual dan mudah dipahami.

- Operator yang Beragam dan Intuitif:

RapidMiner menyediakan beragam operator yang mewakili berbagai teknik analisis data. Tampilan setiap operator dirancang dengan jelas, sehingga pengguna dapat dengan mudah memahami fungsinya tanpa perlu membaca dokumentasi yang terlalu teknis.

- Visualisasi Alur Kerja:

Setelah membangun alur kerja, RapidMiner akan menampilkannya secara visual dalam bentuk diagram. Diagram ini memberikan gambaran keseluruhan tentang proses analisis data yang sedang dilakukan, sehingga pengguna dapat dengan mudah melacak setiap langkah dan mengidentifikasi potensi masalah.

- Color Coding:

RapidMiner menggunakan skema warna yang konsisten untuk membedakan jenis operator dan elemen lainnya dalam alur kerja. Hal ini membantu

pengguna dengan cepat mengidentifikasi bagian-bagian yang berbeda dalam alur kerja.

- **Tooltip dan Informasi**

Bantuan: Saat pengguna mengarahkan kursor ke suatu operator atau elemen lainnya, akan muncul tooltip yang memberikan informasi singkat tentang fungsi dan parameter yang tersedia. Selain itu, RapidMiner juga menyediakan dokumentasi yang lengkap dan mudah diakses untuk setiap operator.

- **Debugger:**

Fitur debugger memungkinkan pengguna untuk melacak eksekusi alur kerja secara langkah demi langkah. Dengan demikian, pengguna dapat dengan mudah mengidentifikasi kesalahan dan melakukan debugging.

- **Metrik Evaluasi yang Visual:**

Hasil analisis data sering kali disajikan dalam bentuk visual, seperti grafik, chart, atau tabel. Hal ini memudahkan pengguna untuk memahami kinerja model dan membuat interpretasi yang tepat.

3. Instalasi Aplikasi RapidMiner

1. Persyaratan Sistem

Sebelum memulai instalasi, pastikan sistem Anda memenuhi persyaratan minimum berikut:

- Sistem Operasi:

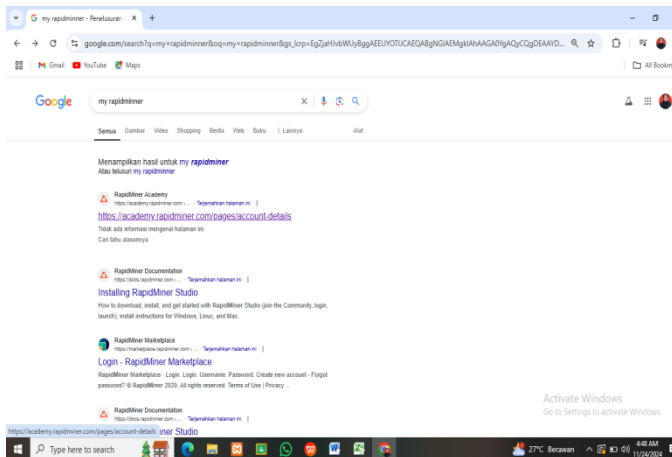
Windows 10 atau lebih baru

macOS 10.12 (Sierra) atau lebih baru

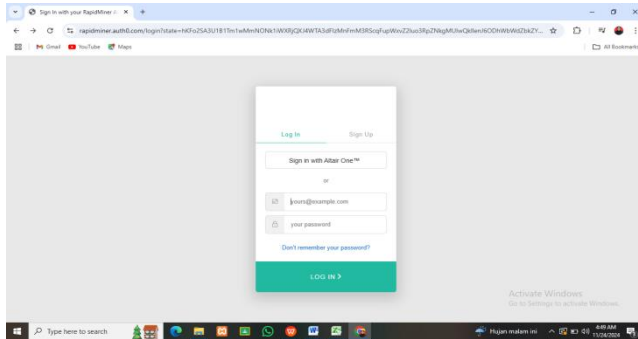
- RAM: Minimal 4 GB (disarankan 8 GB atau lebih)
- Ruang Penyimpanan: Minimal 1 GB ruang kosong
- Java: RapidMiner biasanya menyertakan Java Runtime Environment (JRE) dalam paket instalasinya, tetapi pastikan Anda memiliki versi terbaru jika diperlukan.

2. Unduh Installer :

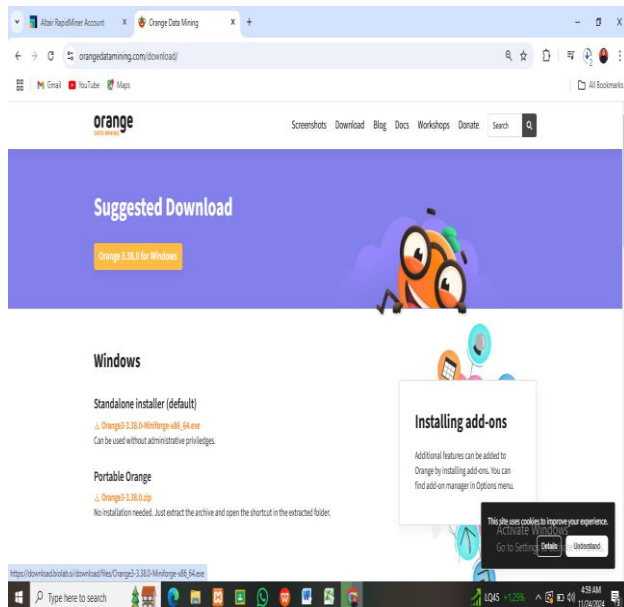
- Kunjungi Situs Resmi: Buka browser dan kunjungi situs resmi RapidMiner di rapidminer.com.



- Buat akun atau login jika sudah memiliki akkn



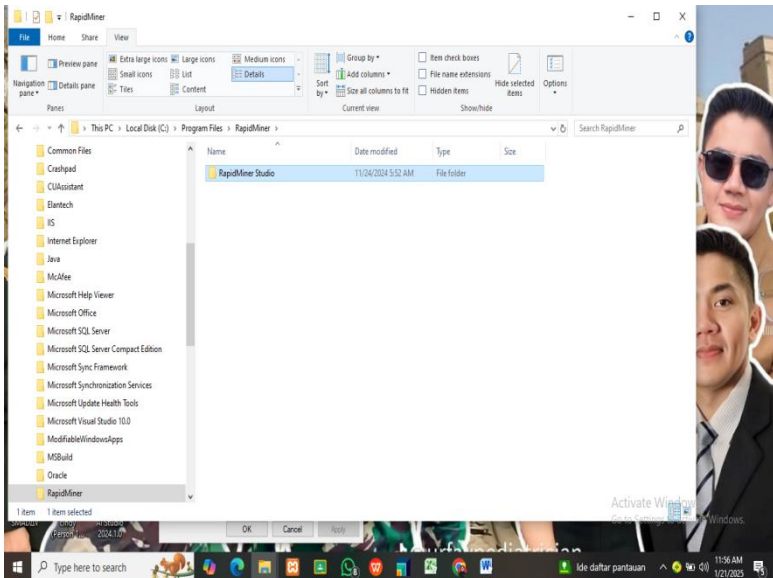
- Cari bagian unduhan dan pilih versi RapidMiner yang sesuai sistem operasi Anda (Windows atau macOS)



- Klik tombol unduh dan simpan file installer ke lokasi yang mudah di akses

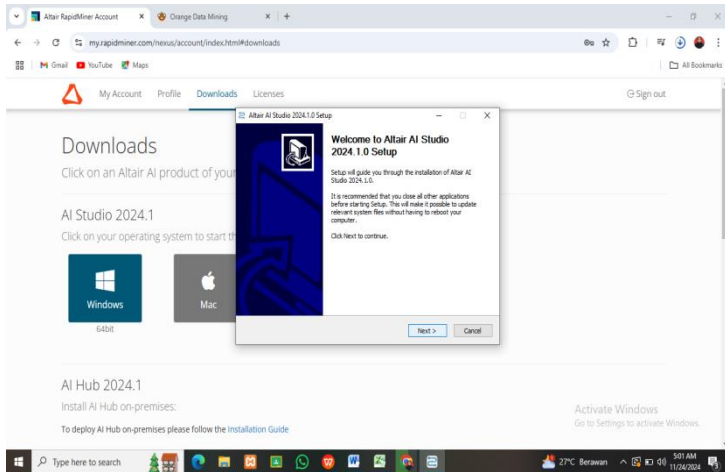
3. Jalankan Installer:

- Cari file installer yang telah diunduh (biasanya berekstensi exe)
- Klik dua kali pada file tersebut untuk memulai proses instalisasi.

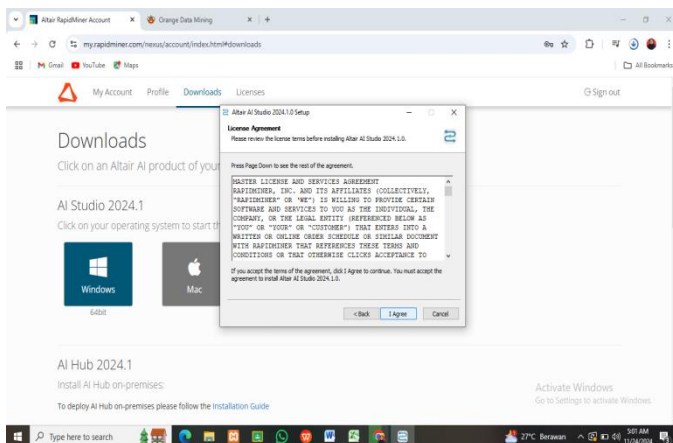


4. Ikuti Petunjuk Layar:

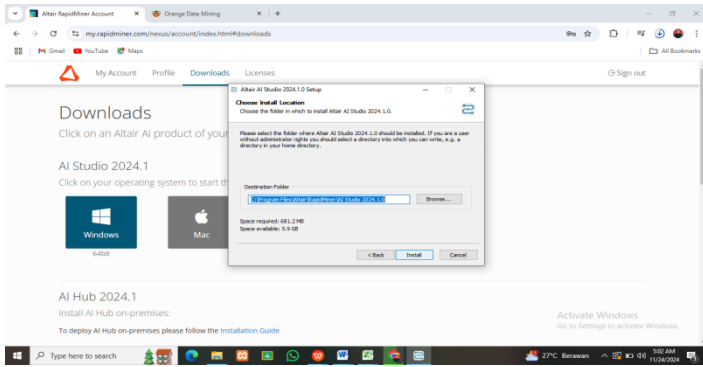
- Jendela selamat datang : klik tombol “next” untuk melanjutkan.



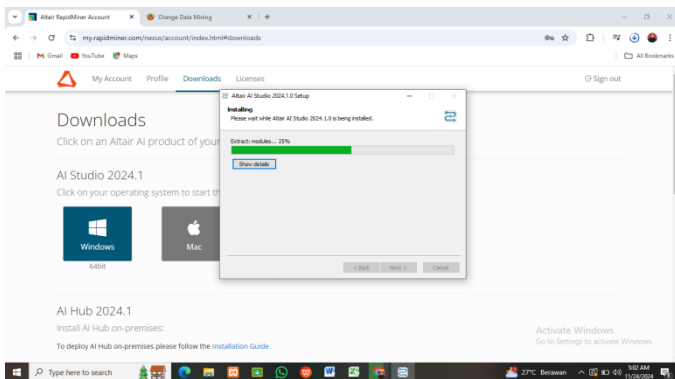
- Perjanjian lisensi : bacalah perjanjian lisensi dengan seksama , jika anda setuju centang kotak dan klik “Next”



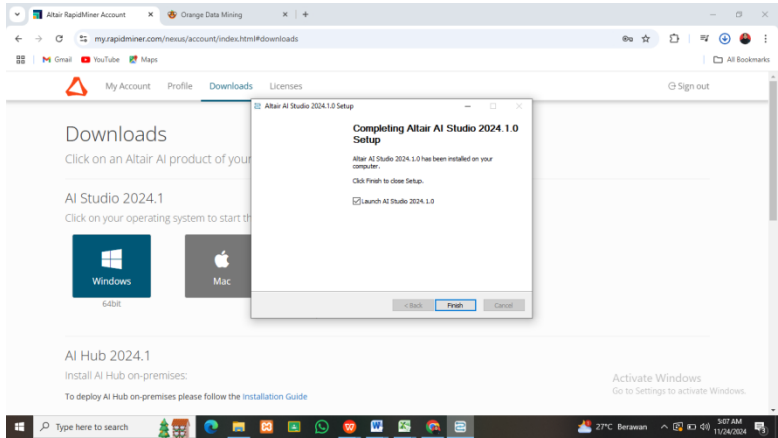
- Lokasi Instalasi: anda dapat memilih lokasi direktori untuk menginstal aplikasi rapidminer. Biarkan default atau pilih lokasi yang diinginkan, lalu klik “next”



- Instalasi : Klik tombol “Install” untuk memulai proses instalasi.



- **Selesai** : setelah instalisasi selesai, klik “Finish” untuk menutup jendela installer.



BAB IV

PENGENALAN APLIKASI ORANGE

1. Apa Itu Orange?

Aplikasi Orange adalah sebuah platform perangkat lunak yang dirancang untuk analisis data dan data mining. Aplikasi ini menyediakan berbagai alat dan teknik yang memungkinkan pengguna untuk melakukan eksplorasi data, analisis statistik, dan visualisasi data dengan cara yang intuitif dan interaktif. Salah satu fitur utama dari Orange adalah antarmuka grafisnya yang memungkinkan pengguna untuk membangun alur kerja analisis data tanpa perlu menulis kode secara langsung. Hal ini membuat Orange sangat cocok untuk pengguna yang tidak memiliki latar belakang pemrograman, seperti pengembang , akademisi, dan profesional di berbagai bidang

Orange mendukung berbagai jenis analisis, termasuk klasifikasi, regresi, clustering, dan analisis asosiasi. Dengan menggunakan widget yang dapat diseret dan dijatuhkan, pengguna dapat dengan mudah mengatur dan menghubungkan berbagai komponen analisis untuk

membangun model yang kompleks. Selain itu, Orange juga menyediakan berbagai pustaka untuk integrasi dengan bahasa pemrograman Python, sehingga pengguna yang lebih berpengalaman dapat memanfaatkan kemampuan pemrograman untuk memperluas fungsionalitas aplikasi.

Salah satu keunggulan Orange adalah kemampuannya untuk menangani data besar dan kompleks. Aplikasi ini dapat digunakan untuk menganalisis dataset yang besar, yang sering kali menjadi tantangan dalam analisis data tradisional. Dengan menggunakan teknik seperti sampling dan pemrosesan paralel, Orange dapat mempercepat proses analisis dan memberikan hasil yang lebih cepat dan efisien.

Orange juga memiliki komunitas pengguna yang aktif dan menyediakan berbagai sumber daya, termasuk tutorial, dokumentasi, dan forum diskusi. Hal ini memungkinkan pengguna baru untuk belajar dan mendapatkan dukungan dari pengguna lain yang lebih berpengalaman. Selain itu, Orange sering diperbarui dengan fitur-fitur baru dan perbaikan, sehingga tetap relevan

dengan perkembangan terbaru dalam bidang data science dan machine learning.

Dalam konteks pendidikan, Orange sering digunakan sebagai alat pengajaran di universitas dan institusi pendidikan lainnya. Dengan antarmuka yang ramah pengguna dan kemampuan untuk melakukan analisis yang kompleks, Orange membantu mahasiswa dan pengajar memahami konsep-konsep dasar dalam analisis data dan machine learning. Ini menjadikannya alat yang sangat berguna dalam kurikulum yang berfokus pada data science.

Aplikasi ini juga digunakan dalam berbagai industri, termasuk kesehatan, pemasaran, dan keuangan. Misalnya, dalam industri kesehatan, Orange dapat digunakan untuk menganalisis data pasien dan membantu dalam pengambilan keputusan klinis. Di bidang pemasaran, aplikasi ini dapat digunakan untuk menganalisis perilaku konsumen dan mengidentifikasi tren pasar.

Secara keseluruhan, Orange adalah alat yang kuat dan fleksibel untuk analisis data yang dapat digunakan oleh berbagai kalangan, dari pemula hingga profesional

berpengalaman. Dengan kemampuannya untuk menyederhanakan proses analisis data dan menyediakan alat yang diperlukan untuk eksplorasi dan visualisasi, Orange menjadi pilihan populer di kalangan pengguna yang ingin memanfaatkan data untuk pengambilan keputusan yang lebih baik.

2. Sejarah Orange

Sejarah Orange, sebagai salah satu perusahaan telekomunikasi terkemuka di dunia, dimulai pada tahun 1994 ketika perusahaan ini didirikan di Inggris. Awalnya, Orange adalah penyedia layanan telepon seluler yang berfokus pada inovasi dan layanan pelanggan yang unggul. Pada tahun 2000, Orange diakuisisi oleh France Telecom, yang kemudian berganti nama menjadi Orange S.A. pada tahun 2013. Akuisisi ini memungkinkan Orange untuk memperluas jangkauannya ke pasar Eropa dan global, serta meningkatkan portofolio layanan yang ditawarkan.

Sejak saat itu, Orange telah berkomitmen untuk menjadi pemimpin dalam inovasi teknologi dan layanan digital. Pada tahun 2019, perusahaan ini meluncurkan

strategi "Lead the Future" yang bertujuan untuk memperkuat posisinya di pasar telekomunikasi dan memperluas layanan digitalnya. Strategi ini mencakup investasi dalam infrastruktur jaringan, pengembangan layanan baru, dan peningkatan pengalaman pelanggan. CEO Orange, Christel Heydemann, menekankan pentingnya inovasi dalam strategi perusahaan untuk menghadapi tantangan di industri telekomunikasi yang terus berkembang .

Pada tahun 2020, Orange melanjutkan ekspansi internasionalnya dengan memperkuat kehadirannya di Afrika dan Timur Tengah. Perusahaan ini melaporkan pertumbuhan yang signifikan di wilayah tersebut, dengan peningkatan pendapatan yang didorong oleh layanan mobile dan digital. Orange juga meluncurkan berbagai inisiatif untuk mendukung transformasi digital di negara-negara berkembang, termasuk program pendidikan dan pelatihan untuk meningkatkan keterampilan digital masyarakat.

Selama pandemi COVID-19 pada tahun 2020 dan 2021, Orange beradaptasi dengan cepat terhadap perubahan

kebutuhan pelanggan. Perusahaan ini meningkatkan layanan digitalnya, termasuk solusi kerja jarak jauh dan platform komunikasi untuk mendukung bisnis dan individu. Inisiatif ini tidak hanya membantu pelanggan tetap terhubung, tetapi juga memperkuat posisi Orange sebagai penyedia layanan telekomunikasi yang responsif dan inovatif .

Pada tahun 2022, Orange meluncurkan beberapa proyek baru yang berfokus pada keberlanjutan dan tanggung jawab sosial. Perusahaan ini berkomitmen untuk mengurangi jejak karbonnya dan meningkatkan efisiensi energi dalam operasionalnya. Selain itu, Orange juga aktif dalam mendukung inisiatif sosial, seperti program akses internet untuk komunitas yang kurang terlayani dan proyek-proyek yang mendukung pendidikan digital .

Memasuki tahun 2023, Orange terus berinvestasi dalam teknologi 5G dan infrastruktur jaringan. Perusahaan ini berupaya untuk memperluas cakupan jaringan 5G di Eropa dan Afrika, serta meningkatkan kualitas layanan bagi pelanggan. Dengan peluncuran berbagai layanan baru yang memanfaatkan teknologi 5G, Orange bertujuan untuk

memberikan pengalaman pengguna yang lebih baik dan mendukung inovasi di berbagai sektor .

Hingga tahun 2024, Orange tetap fokus pada strategi pertumbuhan yang berkelanjutan dan inovasi. Perusahaan ini terus mengeksplorasi peluang baru di bidang teknologi digital, termasuk kecerdasan buatan dan Internet of Things (IoT). Dengan komitmen untuk memberikan layanan yang lebih baik dan mendukung transformasi digital di seluruh dunia, Orange berusaha untuk mempertahankan posisinya sebagai salah satu pemimpin di industri telekomunikasi global.

3. Antar Muka Pengguna Orange

Antarmuka pengguna (UI) aplikasi Orange merupakan elemen penting dalam pengalaman pengguna, yang dirancang untuk memudahkan interaksi dengan fitur-fitur yang ada. Dalam konteks aplikasi ini, desain antarmuka harus mempertimbangkan kebutuhan pengguna yang beragam, mulai dari pemula hingga pengguna berpengalaman. Berikut adalah beberapa aspek penting dari

antarmuka pengguna aplikasi Orange beserta referensi dan kutipannya dari tahun 2019 hingga 2024.

1. Desain Visual yang Menarik

- Antarmuka pengguna Orange dirancang dengan estetika yang menarik, menggunakan warna-warna cerah dan elemen visual yang intuitif.
- Penggunaan ikon dan grafik yang jelas membantu pengguna memahami fungsi setiap alat dan widget.

2. Fleksibilitas dan Kustomisasi

- Pengguna dapat menyesuaikan antarmuka sesuai dengan preferensi mereka, termasuk pengaturan tata letak dan pemilihan widget.
- Kemampuan untuk menambahkan dan menghapus widget dengan mudah memungkinkan pengguna untuk membangun alur kerja yang sesuai dengan kebutuhan analisis mereka.

3. Interaktivitas dan Responsif

- Antarmuka Orange menawarkan interaksi yang responsif, memungkinkan pengguna untuk melihat hasil analisis secara real-time.

- Pengguna dapat mengklik dan menggeser elemen untuk mengeksplorasi data dengan lebih mendalam, seperti menggunakan scatter plots dan heatmaps.
4. Dukungan untuk Pembelajaran dan Pengajaran
- Orange digunakan secara luas dalam konteks pendidikan, membantu siswa dan pengajar memahami konsep-konsep data science melalui visualisasi.
 - Antarmuka yang intuitif memfasilitasi pembelajaran praktis, di mana pengguna dapat langsung berinteraksi dengan data.
5. Integrasi dengan Alat Lain
- Antarmuka pengguna Orange mendukung integrasi dengan berbagai alat dan add-ons, seperti pemrosesan bahasa alami dan analisis jaringan.
 - Ini memungkinkan pengguna untuk memperluas fungsionalitas aplikasi sesuai dengan kebutuhan spesifik mereka.

6. Aksesibilitas dan Dukungan Multibahasa

- Antarmuka Orange dirancang untuk dapat diakses oleh pengguna dari berbagai latar belakang, dengan dukungan untuk beberapa bahasa.
- Ini meningkatkan inklusivitas dan memungkinkan lebih banyak pengguna untuk memanfaatkan alat ini dalam analisis data.

Dengan mempertimbangkan semua aspek ini, antarmuka pengguna aplikasi Orange tidak hanya fungsional tetapi juga dirancang untuk memberikan pengalaman pengguna yang menyenangkan dan mendidik.

4. Instalasi Aplikasi Orange

1. Persyaratan Sistem

Sebelum memulai instalasi, pastikan system

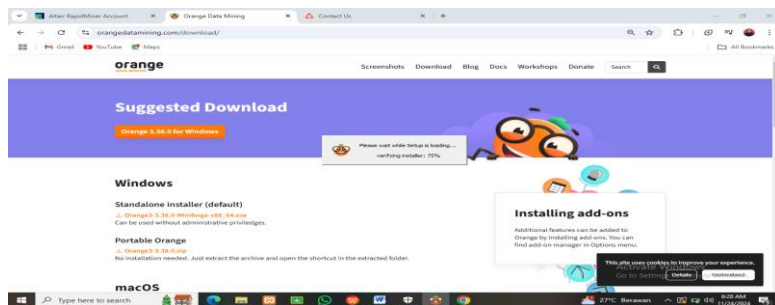
Anda memenuhi persyaratan berikut:

- Sistem Operasi: Windows 7 atau lebih baru, macOS 10.12 atau lebih baru, atau distribusi Linux yang didukung Python (seperti Ubuntu, Fedora).
- Prosesor: Prosesor dual-core atau lebih baik.
- Memori: Minimal 4 GB RAM.
- Ruang Disk: Minimal 2 GB ruang kosong.

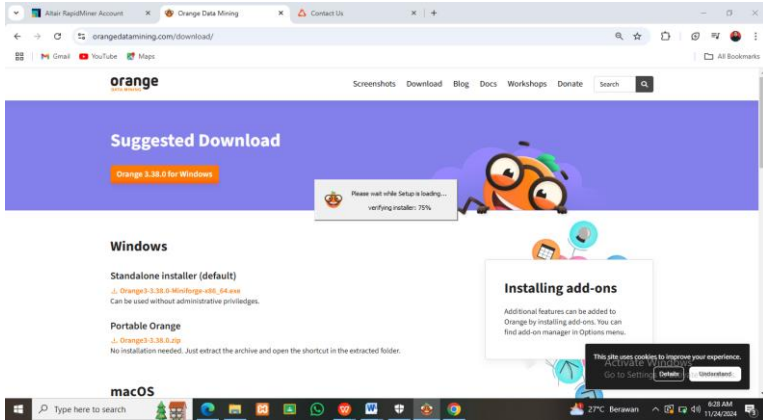
2. Unduh Installer:

- Kunjungi situs web resmi Orange:

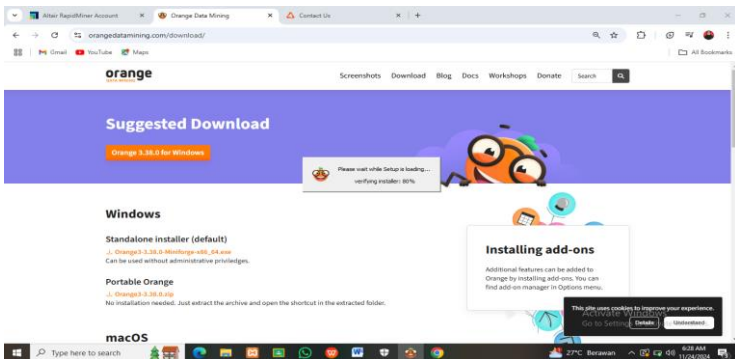
<https://orangedatamining.com/download/>



- Pilih versi Orange yang sesuai dengan sistem operasi Anda. Tersedia versi untuk Windows, macOS, dan beberapa distribusi Linux.

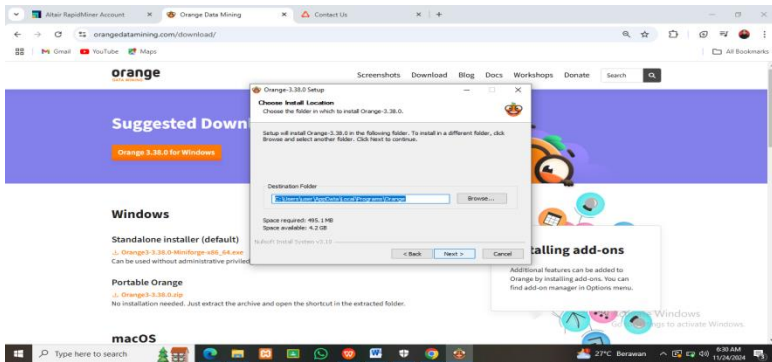


- Klik tombol unduh dan simpan file installer ke lokasi yang mudah diakses.



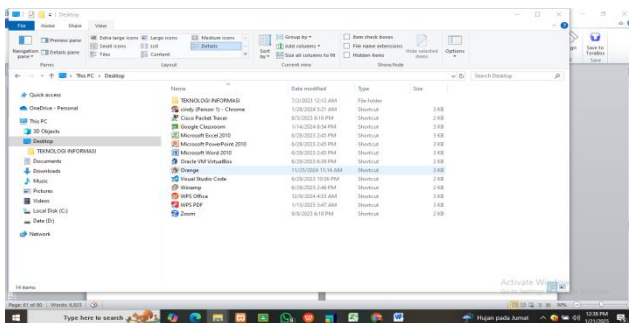
2. Jalankan Installer:

- Windows: Cari file installer yang telah diunduh (biasanya berekstensi .exe), klik dua kali untuk menjalankan. Ikuti petunjuk pada wizard instalasi.

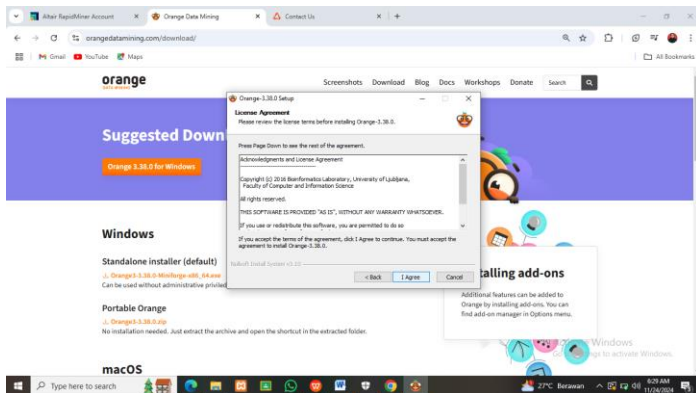


3. Verifikasi Instalasi:

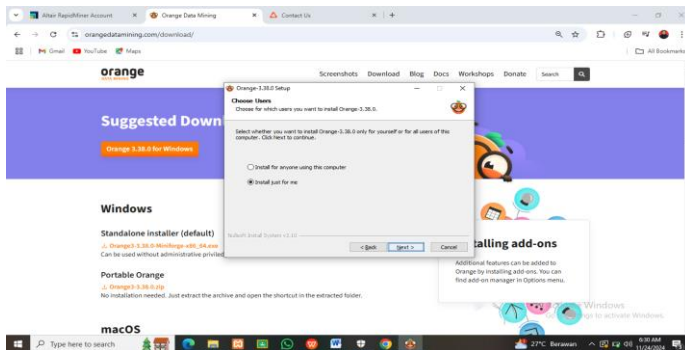
- Setelah instalasi selesai, cari aplikasi Orange di menu Start (Windows).



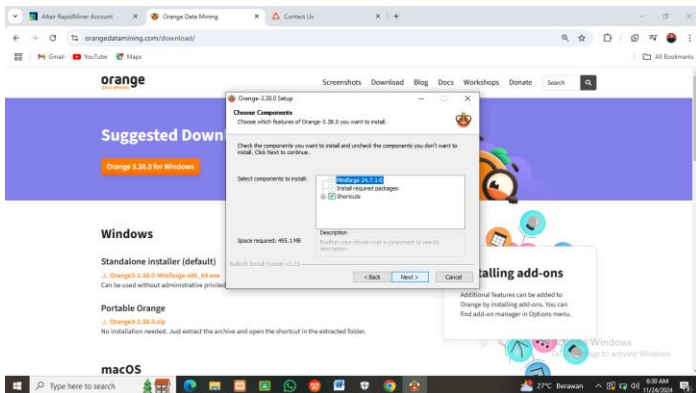
4. Penginstalan setup



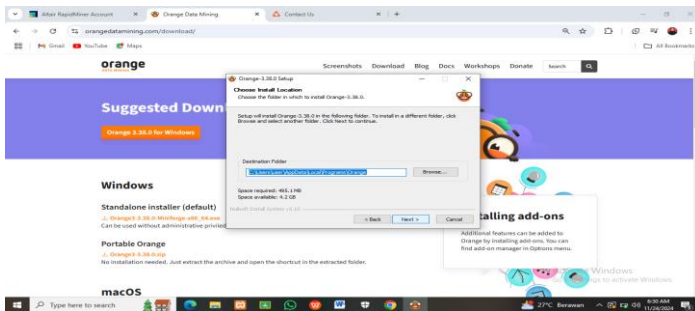
- Klik i agree untuk melanjutkan penginstalan



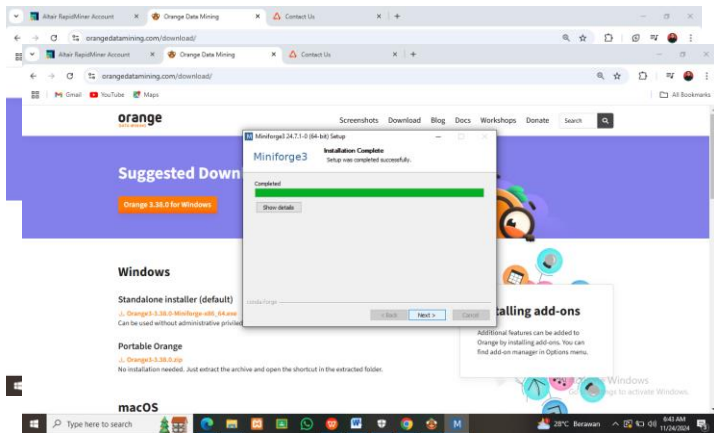
- Selanjutnya klik next



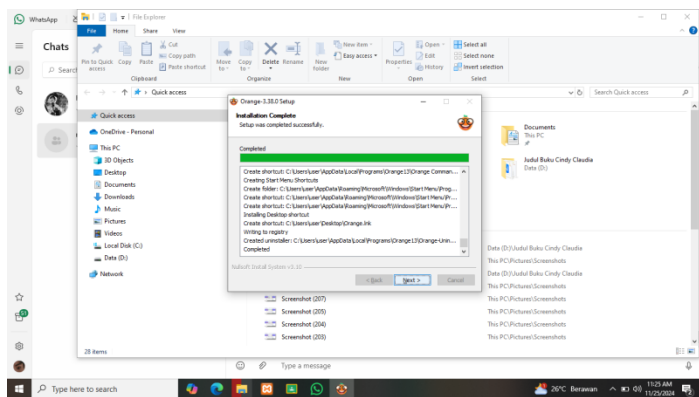
- Lalu klik shortcut lalu klik next



- Selanjutnya pilih lokasi penyimpanan



- Lalu buat nama aplikasinya



- Selanjutnya klik finish

BAB V
PENERAPAN METODE APRIORI PADA
PENJUALAN KOPI KELILING

1. Pengumpulan Data

Proses pengumpulan data dilakukan dengan wawancara kepada penjual kopi keliling dan melakukan observasi langsung, data ini diambil hanya beberapa saja, sehingga menghasilkan pola transaksi.

2. Pre-Processing

Sebelum melakukan pengujian dengan rapidminer dilakukan proses cleaning data di mana proses menghapus data item duplikat dalam setiap transaksi. Data yang sudah di bersihkan dan di transformasi ke dalam bentuk table format tabular.selanjutnya yaitu membuat tabular data penjualan dan mengubah format numerik menjadi format binomial yang berisi 1 untuk transaksi yang berisi penjualan dan 0 untuk penjualan tanpa penjualan.

3. Data Selection

Pada tahap ini yang dilakukan adalah mempersiapkan data yang akan di olah yaitu data penjualankopi keliling. Untuk pembuatan data format tabular menggunakan Microsoft excel kedalam rapidminer dengan menggunakan read excel.

- Data Penjualan Kopi Keliling Per 3 Bulan

KOPING DATA CENDI (LAIN) - Microsoft Excel - Produk Aplikasi Paket

| | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | | | |
|----|--------------------------|----------------|----|--------|---------------------------|----|------------------|----------------|---------------------------|--------|----------------|----|---------------------------|----------------|----|--------|--|--|
| 4 | NO JENIS MINUMAN | | | JUMLAH | | | NO JENIS MINUMAN | | | JUMLAH | | | NO JENIS MINUMAN | | | JUMLAH | | |
| 5 | 1 | KOPI SUSU | 32 | 1 | KOPI SUSU | 47 | 1 | KOPI SUSU | 41 | 1 | KOPI SUSU | 34 | 1 | KOPI SUSU | 35 | | | |
| 6 | 2 | KOPI PANDAN | 40 | 2 | KOPI PANDAN | 42 | 2 | KOPI PANDAN | 54 | 2 | KOPI PANDAN | 37 | 2 | KOPI PANDAN | 54 | | | |
| 7 | 3 | KOPI AREN | 55 | 3 | KOPI AREN | 40 | 3 | KOPI AREN | 43 | 3 | KOPI AREN | 43 | 3 | KOPI AREN | 42 | | | |
| 8 | 4 | KOPI CARAMEL | 49 | 4 | KOPI CARAMEL | 55 | 4 | KOPI CARAMEL | 48 | 4 | KOPI CARAMEL | 48 | 4 | KOPI CARAMEL | 38 | | | |
| 9 | 5 | KOPI AMERICANO | 36 | 5 | KOPI AMERICANO | 38 | 5 | KOPI AMERICANO | 38 | 5 | KOPI AMERICANO | 38 | 5 | KOPI AMERICANO | 49 | | | |
| 11 | DATA PENJUALAN MINGGU -5 | | | | DATA PENJUALAN MINGGU -6 | | | | DATA PENJUALAN MINGGU -7 | | | | DATA PENJUALAN MINGGU -8 | | | | | |
| 14 | NO JENIS MINUMAN | | | JUMLAH | | | NO JENIS MINUMAN | | | JUMLAH | | | NO JENIS MINUMAN | | | JUMLAH | | |
| 15 | 1 | KOPI SUSU | 42 | 1 | KOPI SUSU | 37 | 1 | KOPI SUSU | 52 | 1 | KOPI SUSU | 45 | 1 | KOPI SUSU | 45 | | | |
| 16 | 2 | KOPI PANDAN | 38 | 2 | KOPI PANDAN | 45 | 2 | KOPI PANDAN | 37 | 2 | KOPI PANDAN | 53 | 2 | KOPI PANDAN | 53 | | | |
| 17 | 3 | KOPI AREN | 52 | 3 | KOPI AREN | 44 | 3 | KOPI AREN | 45 | 3 | KOPI AREN | 40 | 3 | KOPI AREN | 40 | | | |
| 18 | 4 | KOPI CARAMEL | 40 | 4 | KOPI CARAMEL | 51 | 4 | KOPI CARAMEL | 52 | 4 | KOPI CARAMEL | 49 | 4 | KOPI CARAMEL | 49 | | | |
| 19 | 5 | KOPI AMERICANO | 35 | 5 | KOPI AMERICANO | 33 | 5 | KOPI AMERICANO | 31 | 5 | KOPI AMERICANO | 45 | 5 | KOPI AMERICANO | 45 | | | |
| 22 | DATA PENJUALAN MINGGU -9 | | | | DATA PENJUALAN MINGGU -10 | | | | DATA PENJUALAN MINGGU -11 | | | | DATA PENJUALAN MINGGU -12 | | | | | |

KOPING DATA CENDI (LAIN) - Microsoft Excel - Produk Aplikasi Paket

| | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | | | |
|----|--------------------------|----------------|----|--------|---------------------------|----|------------------|----------------|---------------------------|--------|----------------|----|---------------------------|----------------|----|--------|--|--|
| 9 | 5 | KOPI AMERICANO | 36 | 5 | KOPI AMERICANO | 38 | 5 | KOPI AMERICANO | 38 | 5 | KOPI AMERICANO | 38 | 5 | KOPI AMERICANO | 49 | | | |
| 11 | DATA PENJUALAN MINGGU -5 | | | | DATA PENJUALAN MINGGU -6 | | | | DATA PENJUALAN MINGGU -7 | | | | DATA PENJUALAN MINGGU -8 | | | | | |
| 14 | NO JENIS MINUMAN | | | JUMLAH | | | NO JENIS MINUMAN | | | JUMLAH | | | NO JENIS MINUMAN | | | JUMLAH | | |
| 15 | 1 | KOPI SUSU | 42 | 1 | KOPI SUSU | 37 | 1 | KOPI SUSU | 52 | 1 | KOPI SUSU | 45 | 1 | KOPI SUSU | 45 | | | |
| 16 | 2 | KOPI PANDAN | 38 | 2 | KOPI PANDAN | 45 | 2 | KOPI PANDAN | 37 | 2 | KOPI PANDAN | 53 | 2 | KOPI PANDAN | 53 | | | |
| 17 | 3 | KOPI AREN | 52 | 3 | KOPI AREN | 44 | 3 | KOPI AREN | 45 | 3 | KOPI AREN | 40 | 3 | KOPI AREN | 40 | | | |
| 18 | 4 | KOPI CARAMEL | 40 | 4 | KOPI CARAMEL | 51 | 4 | KOPI CARAMEL | 52 | 4 | KOPI CARAMEL | 49 | 4 | KOPI CARAMEL | 49 | | | |
| 19 | 5 | KOPI AMERICANO | 35 | 5 | KOPI AMERICANO | 33 | 5 | KOPI AMERICANO | 31 | 5 | KOPI AMERICANO | 45 | 5 | KOPI AMERICANO | 45 | | | |
| 22 | DATA PENJUALAN MINGGU -9 | | | | DATA PENJUALAN MINGGU -10 | | | | DATA PENJUALAN MINGGU -11 | | | | DATA PENJUALAN MINGGU -12 | | | | | |
| 24 | NO JENIS MINUMAN | | | JUMLAH | | | NO JENIS MINUMAN | | | JUMLAH | | | NO JENIS MINUMAN | | | JUMLAH | | |
| 25 | 1 | KOPI SUSU | 35 | 1 | KOPI SUSU | 47 | 1 | KOPI SUSU | 48 | 1 | KOPI SUSU | 37 | 1 | KOPI SUSU | 37 | | | |
| 26 | 2 | KOPI PANDAN | 40 | 2 | KOPI PANDAN | 55 | 2 | KOPI PANDAN | 39 | 2 | KOPI PANDAN | 45 | 2 | KOPI PANDAN | 45 | | | |
| 27 | 3 | KOPI AREN | 51 | 3 | KOPI AREN | 54 | 3 | KOPI AREN | 56 | 3 | KOPI AREN | 42 | 3 | KOPI AREN | 42 | | | |
| 28 | 4 | KOPI CARAMEL | 46 | 4 | KOPI CARAMEL | 43 | 4 | KOPI CARAMEL | 45 | 4 | KOPI CARAMEL | 54 | 4 | KOPI CARAMEL | 54 | | | |
| 29 | 5 | KOPI AMERICANO | 32 | 5 | KOPI AMERICANO | 31 | 5 | KOPI AMERICANO | 30 | 5 | KOPI AMERICANO | 34 | 5 | KOPI AMERICANO | 34 | | | |

- Pembuatan Format Tabular
Ubah data penjualan per 3 bulan menjadi data tabular seperti berikut:

The screenshot shows a Microsoft Excel spreadsheet with the following data table:

| KOMBINASI 1 ITEMSET | | | | | | |
|---------------------|-------------------|-------------|-----------|--------------|----------------|--|
| MINGGU | ITEM YANG DI BELI | | | | | |
| | KOPI SUSU | KOPI PANDAN | KOPI AREN | KOPI CARAMEL | KOPI AMERICANO | |
| 1 | 0 | 1 | 1 | 1 | 0 | |
| 2 | 1 | 1 | 0 | 1 | 0 | |
| 3 | 0 | 1 | 1 | 1 | 0 | |
| 4 | 1 | 0 | 1 | 0 | 1 | |
| 5 | 1 | 0 | 1 | 1 | 0 | |
| 6 | 0 | 1 | 1 | 1 | 0 | |
| 7 | 1 | 0 | 1 | 1 | 0 | |
| 8 | 1 | 1 | 0 | 0 | 1 | |
| 9 | 0 | 1 | 1 | 1 | 0 | |
| 10 | 1 | 1 | 1 | 0 | 0 | |
| 11 | 1 | 0 | 1 | 1 | 0 | |
| 12 | 0 | 1 | 1 | 1 | 0 | |

The spreadsheet also shows a taskbar at the bottom with the Windows logo, search bar, and various application icons. The system tray shows the date and time as 10:09 AM on 4/17/2025.

- Pembuatan 1 Itemset

| NO | ITEM | Σ TRANSAKSI | SUPPORT(%) |
|----|----------------|-------------|----------------|
| 1 | KOPI SUSU | 7 | 7/12x100= 58% |
| 2 | KOPI PANDAN | 8 | 8/12x100= 67% |
| 3 | KOPI AREN | 10 | 10/12x100= 83% |
| 4 | KOPI CAMEL | 9 | 9/12x100= 75% |
| 5 | KOPI AMERICANO | 2 | 2/12x100= 17% |

Support =20%

$$\text{Support}(A) = \frac{\Sigma \text{transaksi Mengandung } A}{\Sigma \text{Transaksi}} \times 100 \%$$

• Pembuatan 2 Itemset

| KOMBINASI 2 ITEMSET | | | | | | | | | | | | |
|--------------------------|-------------------|-------------|------------------|-------------------------|-------------------|-----------|------------------|-------------------------|-------------------|------------|------------------|--|
| 1. KOPI SUSU-KOPI PANDAN | | | | 2. KOPI SUSU-KOPI AREN | | | | 2. KOPI SUSU-KOPI CAMEL | | | | |
| MINGGU | ITEM YANG DI BELI | | TRANSAKSI 2 ITEM | MINGGU | ITEM YANG DI BELI | | TRANSAKSI 2 ITEM | MINGGU | ITEM YANG DI BELI | | TRANSAKSI 2 ITEM | |
| | KOPI SUSU | KOPI PANDAN | | | KOPI SUSU | KOPI AREN | | | KOPI SUSU | KOPI CAMEL | | |
| 1 | 0 | 1 | NO | 1 | 0 | 1 | NO | 1 | 0 | 1 | NO | |
| 2 | 1 | 1 | YES | 2 | 1 | 0 | NO | 2 | 1 | 1 | YES | |
| 3 | 0 | 1 | NO | 3 | 0 | 1 | NO | 3 | 0 | 1 | NO | |
| 4 | 1 | 0 | NO | 4 | 1 | 1 | YES | 4 | 1 | 0 | NO | |
| 5 | 1 | 0 | NO | 5 | 1 | 1 | YES | 5 | 1 | 1 | YES | |
| 6 | 0 | 1 | NO | 6 | 0 | 1 | NO | 6 | 0 | 1 | NO | |
| 7 | 1 | 0 | NO | 7 | 1 | 1 | NO | 7 | 1 | 1 | YES | |
| 8 | 1 | 1 | YES | 8 | 1 | 0 | NO | 8 | 1 | 0 | NO | |
| 9 | 0 | 1 | NO | 9 | 0 | 1 | NO | 9 | 0 | 1 | NO | |
| 10 | 1 | 1 | YES | 10 | 1 | 1 | YES | 10 | 1 | 0 | NO | |
| 11 | 1 | 0 | NO | 11 | 1 | 1 | YES | 11 | 1 | 1 | YES | |
| 12 | 0 | 1 | NO | 12 | 0 | 1 | NO | 12 | 0 | 1 | NO | |
| JUMLAH TRANSAKSI 2 ITEM | | | 3 | JUMLAH TRANSAKSI 2 ITEM | | | 4 | JUMLAH TRANSAKSI 2 ITEM | | | 4 | |
| SUPPORT | | | 25% | SUPPORT | | | 33% | SUPPORT | | | 33% | |

| 1. KOPI PANDAN - KOPI AREN | | | | 2. KOPI PANDAN - KOPI CAMEL | | | | 2. KOPI AREN - KOPI CAMEL | | | |
|----------------------------|-------------------|-----------|------------------|-----------------------------|-------------------|------------|------------------|---------------------------|-------------------|------------|------------------|
| MINGGU | ITEM YANG DI BELI | | TRANSAKSI 2 ITEM | MINGGU | ITEM YANG DI BELI | | TRANSAKSI 2 ITEM | MINGGU | ITEM YANG DI BELI | | TRANSAKSI 2 ITEM |
| | KOPI PANDAN | KOPI AREN | | | KOPI PANDAN | KOPI CAMEL | | | KOPI AREN | KOPI CAMEL | |
| 1 | 1 | 1 | YES | 1 | 1 | 1 | YES | 1 | 1 | 1 | YES |
| 2 | 1 | 0 | NO | 2 | 1 | 1 | YES | 2 | 0 | 1 | NO |
| 3 | 1 | 1 | YES | 3 | 1 | 1 | YES | 3 | 1 | 1 | YES |
| 4 | 0 | 1 | NO | 4 | 0 | 0 | NO | 4 | 1 | 0 | NO |
| 5 | 0 | 1 | NO | 5 | 0 | 1 | NO | 5 | 1 | 1 | YES |
| 6 | 1 | 1 | YES | 6 | 1 | 1 | YES | 6 | 1 | 1 | YES |
| 7 | 0 | 1 | NO | 7 | 0 | 1 | NO | 7 | 1 | 1 | YES |
| 8 | 1 | 0 | NO | 8 | 1 | 0 | NO | 8 | 0 | 0 | NO |
| 9 | 1 | 1 | YES | 9 | 1 | 1 | YES | 9 | 1 | 1 | YES |
| 10 | 1 | 1 | YES | 10 | 1 | 0 | NO | 10 | 1 | 0 | NO |
| 11 | 0 | 1 | NO | 11 | 0 | 1 | NO | 11 | 1 | 1 | YES |
| 12 | 1 | 1 | YES | 12 | 1 | 1 | YES | 12 | 1 | 1 | YES |
| JUMLAH TRANSAKSI 2 ITEM | | | 6 | JUMLAH TRANSAKSI 2 ITEM | | | 6 | JUMLAH TRANSAKSI 2 ITEM | | | 8 |
| SUPPORT | | | 50% | SUPPORT | | | 50% | SUPPORT | | | 67% |

| NO | ITEM | 3 TRANSAKSI | SUPPORT |
|----|----------------------------|-------------|--------------------------|
| 1 | KOPI PANDAN + SUSU | 3 | $3/12 \times 100 = 25\%$ |
| 2 | KOPI SUSU+KOPI AREN | 4 | $4/12 \times 100 = 33\%$ |
| 3 | KOPI SUSU-KOPI CAMEL | 4 | $4/12 \times 100 = 33\%$ |
| 4 | KOPI PANDAN- KOPI AREN | 6 | $6/12 \times 100 = 50\%$ |
| 5 | KOPI PANDAN- KOPI CAMEL | 6 | $6/12 \times 100 = 50\%$ |
| 6 | KOPI AREN- CAMEL | 8 | $8/12 \times 100 = 67\%$ |

• Pembuatan 3 Itemset

| 1. KOPI SUSU-KOPI PANDAN-KOPI AREN | | | | | 1. KOPI SUSU-KOPI PANDAN-KOPI AREN | | | | |
|------------------------------------|-------------------|-------------|-----------|-------------|------------------------------------|-------------------|-----------|------------|------------------|
| MINGGU | ITEM YANG DI BELI | | | TRANSAKSI 2 | MINGGU | ITEM YANG DI BELI | | | TRANSAKSI 2 ITEM |
| | KOPI SUSU | KOPI PANDAN | KOPI AREN | | | KOPI PANDAN | KOPI AREN | KOPI CAMEL | |
| 1 | 0 | 1 | 1 | NO | 1 | 1 | 1 | 1 | YES |
| 2 | 1 | 1 | 0 | NO | 2 | 1 | 0 | 1 | NO |
| 3 | 0 | 1 | 1 | NO | 3 | 1 | 1 | 1 | YES |
| 4 | 1 | 0 | 1 | NO | 4 | 0 | 1 | 0 | NO |
| 5 | 1 | 0 | 1 | NO | 5 | 0 | 1 | 1 | NO |
| 6 | 0 | 1 | 1 | NO | 6 | 1 | 1 | 1 | YES |
| 7 | 1 | 0 | 1 | NO | 7 | 0 | 1 | 1 | NO |
| 8 | 1 | 1 | 0 | NO | 8 | 1 | 0 | 0 | NO |
| 9 | 0 | 1 | 1 | NO | 9 | 1 | 1 | 1 | YES |
| 10 | 1 | 1 | 1 | YES | 10 | 1 | 1 | 0 | NO |
| 11 | 1 | 0 | 1 | NO | 11 | 0 | 1 | 1 | NO |
| 12 | 0 | 1 | 1 | NO | 12 | 1 | 1 | 1 | YES |
| JUMLAH TRANSAKSI 2 ITEM | | | | 1 | JUMLAH TRANSAKSI 2 ITEM | | | | 5 |
| SUPPORT | | | | 8% | SUPPORT | | | | 42% |

| NO | ITEM | Σ TRANSAKSI | SUPPORT |
|----|---------------------------------|--------------------|---------|
| 1 | KOPI SUSU-KOPI PANDAN-KOPI AREN | 1 | 8% |
| 2 | KOPI PANDAN-KOPI AREN-CAMEL | 5 | 42% |

- Pembentukan Tabel Confidence

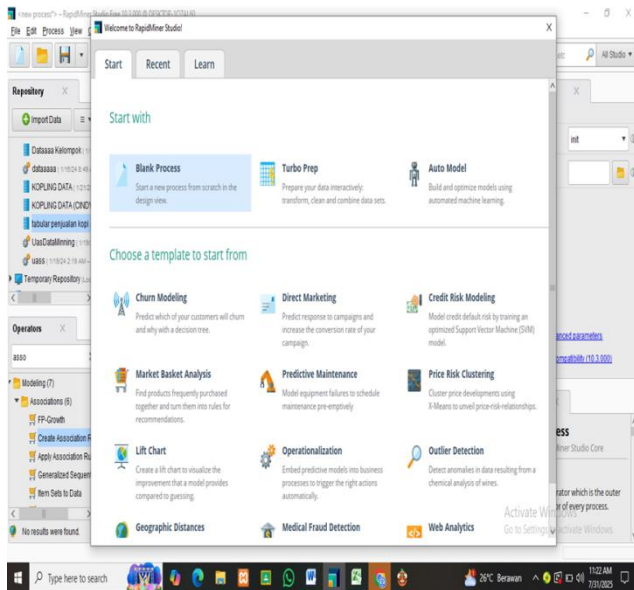
| ATURAN ASOSIASI 2 ITEMSET | | | |
|---|---------------|----------|------------|
| RULE | $\sum A \& B$ | $\sum A$ | CONFIDENCE |
| JIKA MEMBELI KOPI PANDAN MAKA MEMBELI KOPI SUSU | 3 | 8 | 38% |
| JIKA MEMBELI KOPI SUSU MAKA MEMBELI KOPI AREN | 4 | 7 | 57% |
| JIKA MEMBELI KOPI SUSU MAKA MEMBELI KOPI CAMEL | 4 | 7 | 57% |
| JIKA MEMBELI KOPI PANDAN MAKA MEMBELI KOPI AREN | 6 | 8 | 75% |
| JIKA MEMBELI KOPI PANDAN MAKA MEMBELI KOPI CAMEL | 6 | 8 | 75% |
| JIKA MEMBELI KOPI AREN MAKA MEMBELI KOPI CAMEL | 8 | 10 | 80% |
| ATURAN ASOSIASI 3 ITEMSET | | | |
| RULE | $\sum A \& B$ | $\sum A$ | CONFIDENCE |
| JIKA MEMBELI KOPI PANDAN DAN KOPI AREN MAKA MEMBELI KOPI CAMEL | 5 | 8 | 63% |
| JIKA MEMBELI KOPI PANDAN DAN KOPI CAMEL MAKA AKAN MEMBELI KOPI AREN | 5 | 8 | 63% |
| JIKA MEMBELI KOPI CAMEL DAN KOPI AREN MAKA MEMBELI KOPI PANDAN | 5 | 9 | 56% |

| 20 | ASOSIASI FINAL | | | |
|----|--|---------|------------|--------------------|
| 21 | ATURAN YANG DIGUNAKAN ADALAH ATURAN YANG MEMILIKI NILAI CONFIDENCE $\geq 60\%$ | | | |
| 22 | RULE | SUPPORT | CONFIDENCE | SUPPORT*CONFIDENCE |
| 23 | JIKA MEMBELI KOPI PANDAN MAKA MEMBELI KOPI AREN | 0.50 | 0.75 | 38% |
| 24 | JIKA MEMBELI KOPI PANDAN MAKA MEMBELI KOPI CAMEL | 0.50 | 0.75 | 38% |
| 25 | JIKA MEMBELI KOPI AREN MAKA MEMBELI KOPI CAMEL | 0.67 | 0.80 | 54% |
| 26 | JIKA MEMBELI KOPI PANDAN DAN KOPI AREN MAKA MEMBELI KOPI CAMEL | 0.42 | 0.63 | 26% |
| 27 | JIKA MEMBELI KOPI PANDAN DAN KOPI CAMEL MAKA AKAN MEMBELI KOPI AREN | 0.42 | 0.63 | 26% |
| 28 | | | | |
| 29 | | | | |
| 30 | ATURAN ASOSIASI | | | |
| 31 | ATURAN YANG DIGUNAKAN ADALAH ATURAN YANG MEMILIKI HASIL PALING BESAR DARI PERKALIAN NILAI SUPPORT DENGAN NILAI CONFIDENCE. | | | |
| 32 | 1. JIKA MEMBELI KOPI AREN MAKA MEMBELI KOPI CAMEL | | | |
| 33 | | | | |

BAB VI

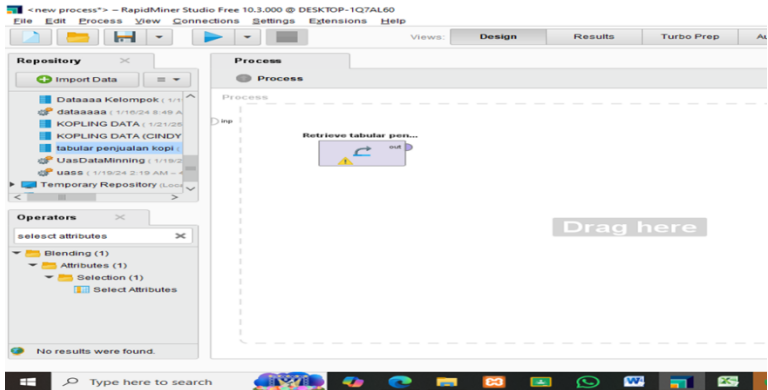
IMPLEMENTASI DATA MENGGUNAKAN APLIKASI RAPIDMINER

1. Buka File baru di RapidMiner, ini adalah langkah awal untuk memulai analisis baru.



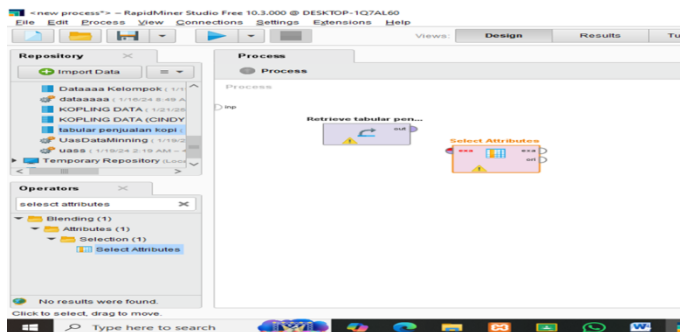
Gambar 6. 1

2. Impor data penjualan kopi yang bernama *tabular penjualan kopi* ke dalam RapidMiner.



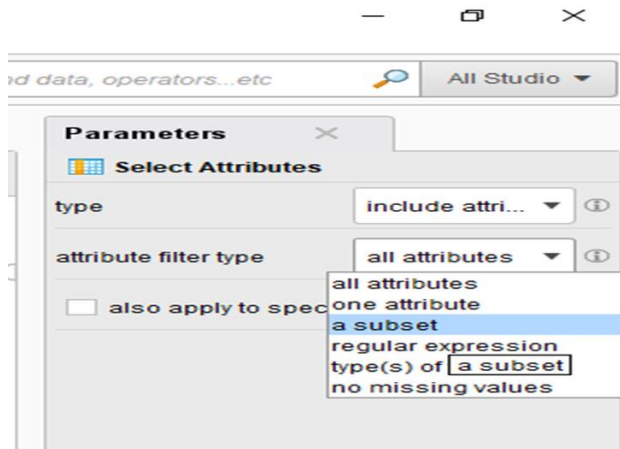
Gambar 6. 2

3. Hubungkan data ke jendela utama di sebelah kanan. Pilih operator "Select Attributes" dan seret ke panel proses. Operator ini digunakan untuk memilih kolom atau atribut yang relevan dari data Anda untuk analisis.



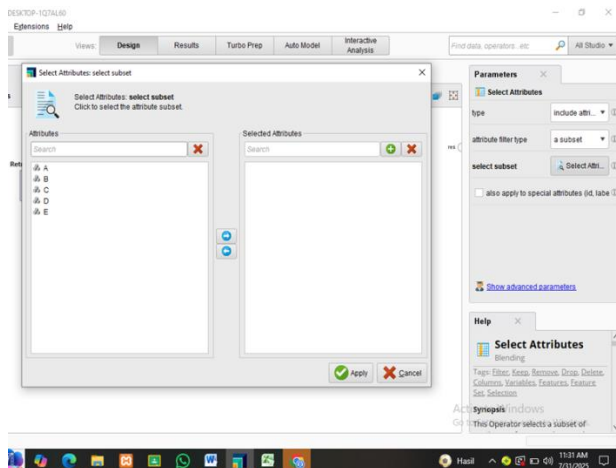
Gambar 6. 3

4. Ubah parameter di "Select Attributes" ke type a subset.



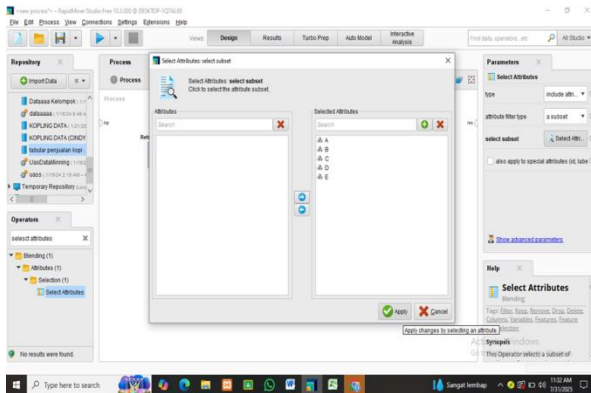
Gambar 6. 4

5. Seret atribut yang kita butuhkan, dari kolom kiri ke kanan.



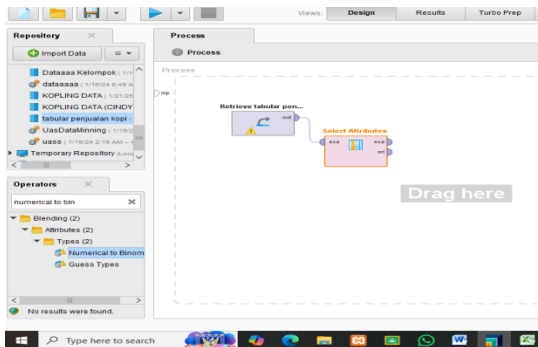
Gambar 6. 5

6. Setelah di pindai atribut yang dibutuhkan maka kita klik “apply



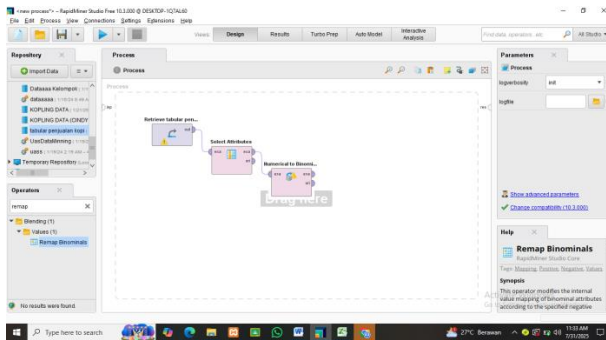
Gambar 6. 6

7. Pilih operator "Numerical to Binominal" dan seret ke panel proses. Operator ini berfungsi untuk mengubah nilai numerik



Gambar 6. 7

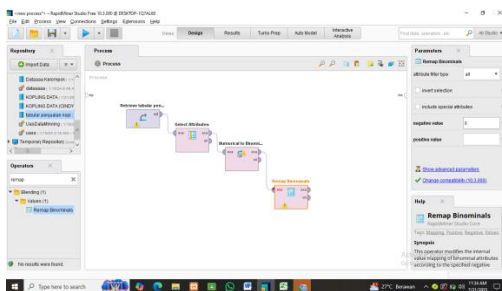
8. Hubungkan "Numerical to Binominal" dengan operator sebelumnya. Atur parameter "Numerical to Binominal" untuk mengubah nilai negatif menjadi 0 dan nilai positif menjadi 1, yang menandakan apakah suatu item dibeli atau tidak.



Gambar

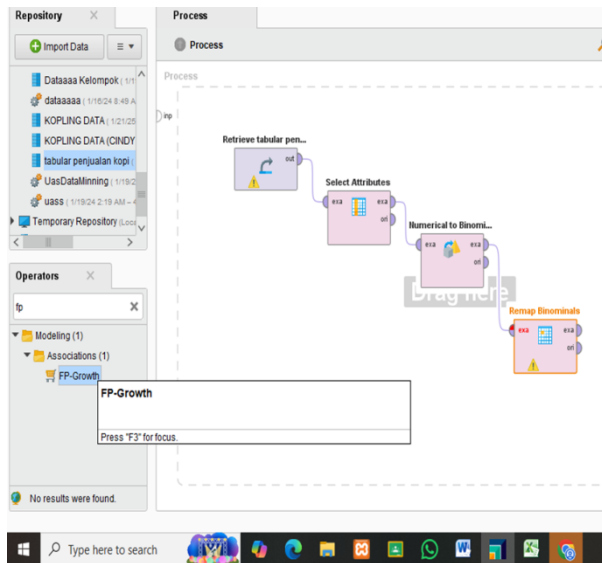
6.8

9. Pilih operator "Remap Binominal" dan seret ke panel proses. Operator ini berfungsi untuk mengubah nilai biner ke nilai yang baru.



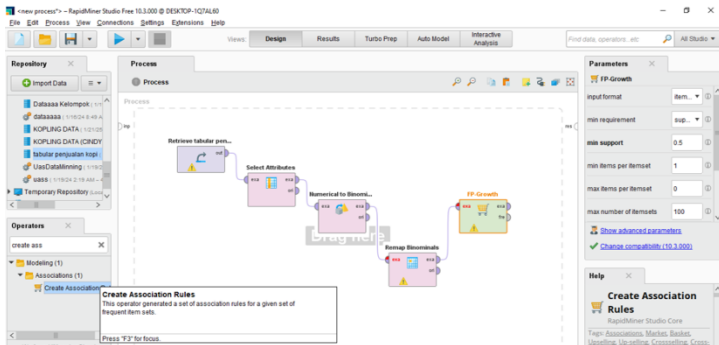
Gambar 6. 9

10. Pilih operator "FP-Growth" dan seret ke panel proses. Ini adalah algoritma yang digunakan untuk menemukan item set yang sering muncul dalam data. Hubungkan "FP-Growth" dengan "Numerical to Binominal". Atur parameter FP-Growth" seperti Minimum number of item sets diatur ke 100 dan Minimum support diatur ke 0.5.



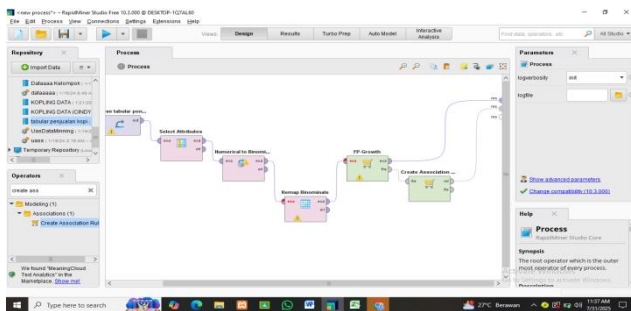
Gambar 6. 10

11. Pilih operator "Create Association Rules" dan seret ke panel proses. Operator ini digunakan untuk membuat aturan asosiasi dari hasil FP-Growth. Hubungkan "Create Association Rules" dengan "FP-Growth"



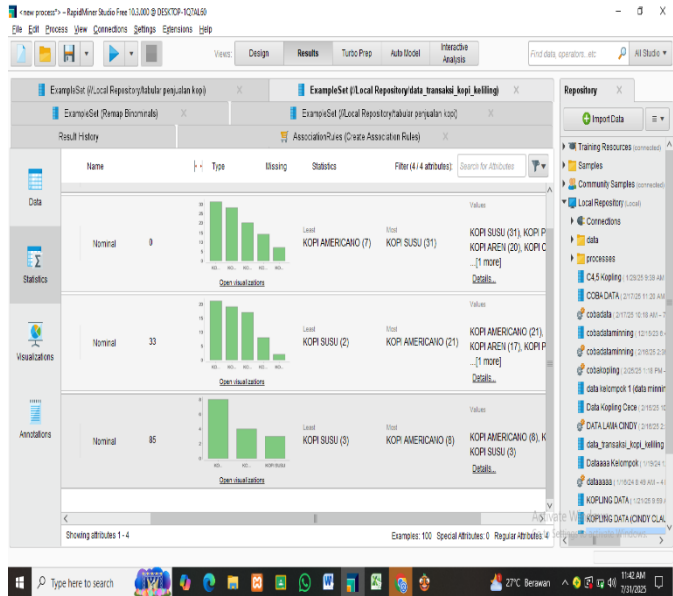
Gambar 6. 11

12. Hubungkan Fp Growth dan Create Association ke Res, lalu klik run atau segitiga biru di atas.



Gambar 6. 12

13. Grafik hasil dari Run yang menjelaskan perbandingan penjualan kopi. Dilihat dari grafik tersebut kopi susu selalu lebih unggul dari kopi Americano dan kopi lainnya.



Gambar 6. 13

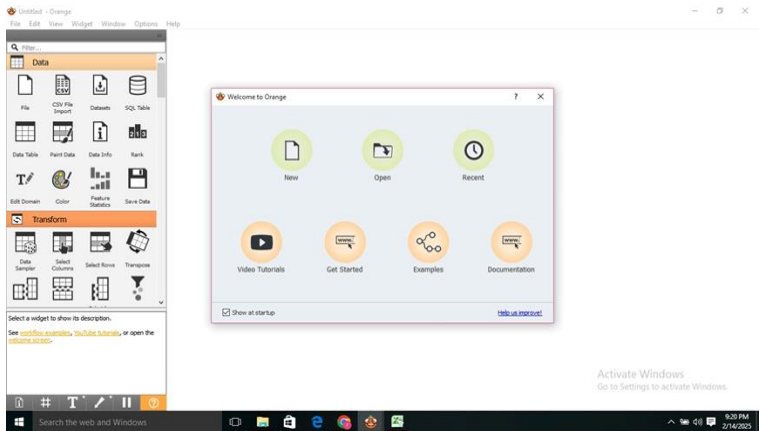
Hasil Pembahasan

Hasil analisis dengan RapidMiner memperlihatkan bahwa pola pembelian kopi cenderung terjadi pada varian *Kopi Aren* dan *Kopi Caramel*. Kombinasi ini menunjukkan preferensi konsumen terhadap cita rasa manis. Sementara itu, rendahnya pembelian *Kopi Americano* menunjukkan minat konsumen yang kurang terhadap kopi dengan rasa pahit tanpa campuran.

BAB VII

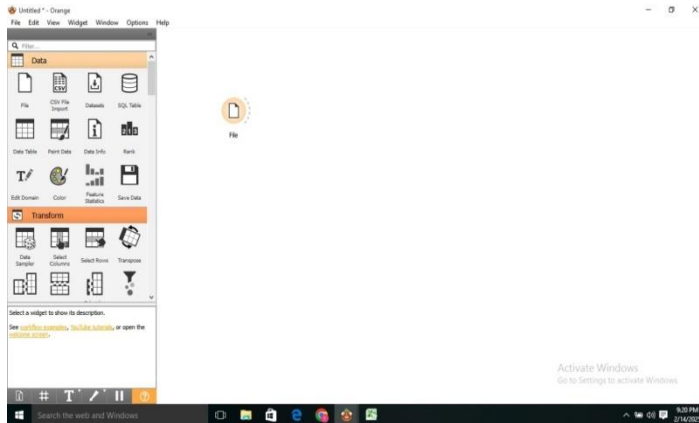
IMPLEMENTASI DATA MENGGUNAKAN APLIKASI ORANGE

1. Buka aplikasi Orange. Anda akan melihat jendela sambutan ("Welcome to Orange")



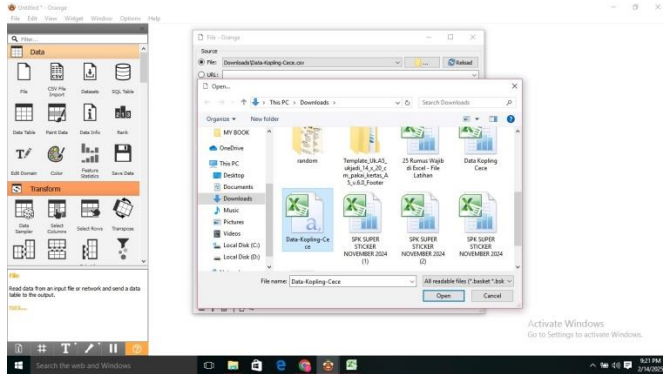
Gambar 7. 1

2. Pilih New untuk membuat lembar kerja baru. Jendela utama Orange akan terbuka, menampilkan area kerja kosong di sebelah kanan dan daftar widget di sebelah kiri.



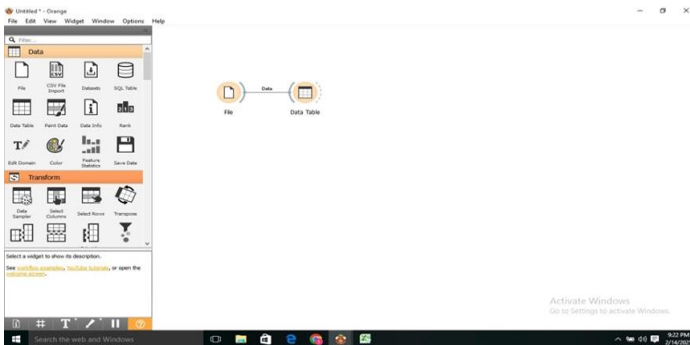
Gambar 7. 2

3. Di panel widget sebelah kiri, cari kategori Data. Klik dan seret widget File ke area kerja kosong di sebelah kanan. Klik dua kali pada widget File yang baru Anda letakkan di area kerja. Jendela konfigurasi widget akan muncul. Klik tombol Browse atau ikon folder di samping kolom "File". Ini akan membuka jendela penjelajah file komputer Anda. Cari dan pilih file data Anda. Dalam contoh Anda, file yang digunakan adalah "Data Kopling Cece.csv". Klik Open setelah memilih file. File akan dimuat ke dalam widget. Anda bisa menutup jendela konfigurasi.



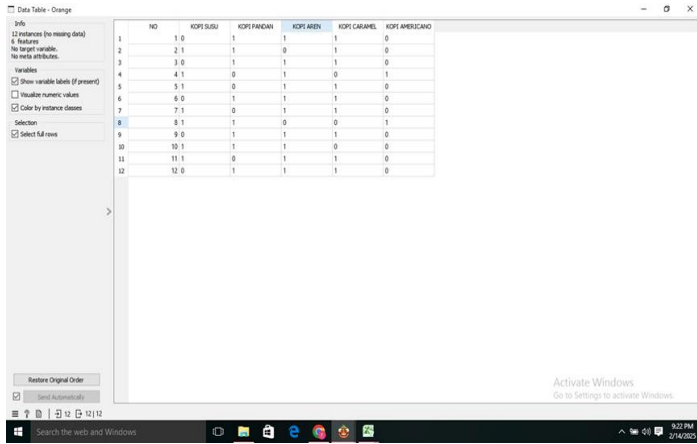
Gambar 7. 3

4. Sekarang, kita akan melihat isi dari file data tersebut. Di panel widget sebelah kiri, di kategori Data, cari dan seret widget Data Table ke area kerja. Hubungkan widget File dengan widget Data Table. Caranya, klik pada lingkaran kecil di sisi kanan widget File dan seret garis yang muncul ke lingkaran kecil di sisi kiri widget Data Table.



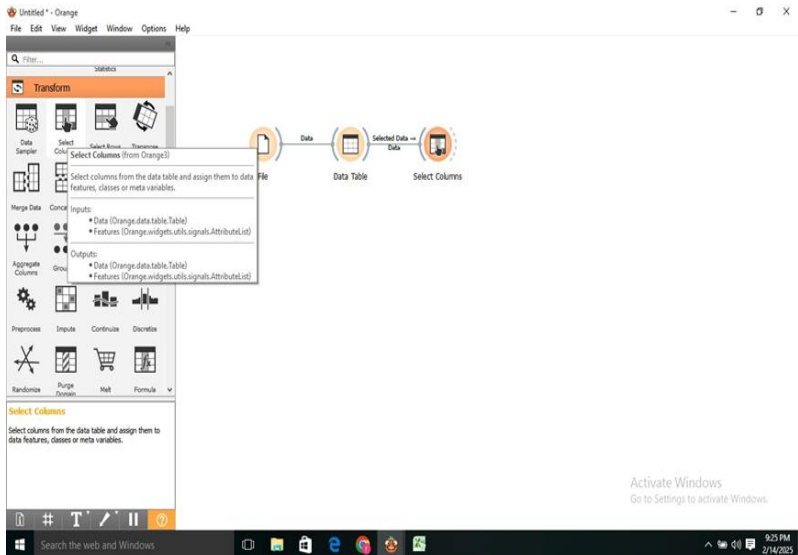
Gambar 7. 4

- Klik dua kali pada widget Data Table. Jendela baru akan muncul yang menampilkan isi dari file CSV Anda dalam bentuk tabel, data Anda berisi kolom-kolom seperti No, Kopi Susu, Kopi Pandan, Kopi Aren, dan lainnya.



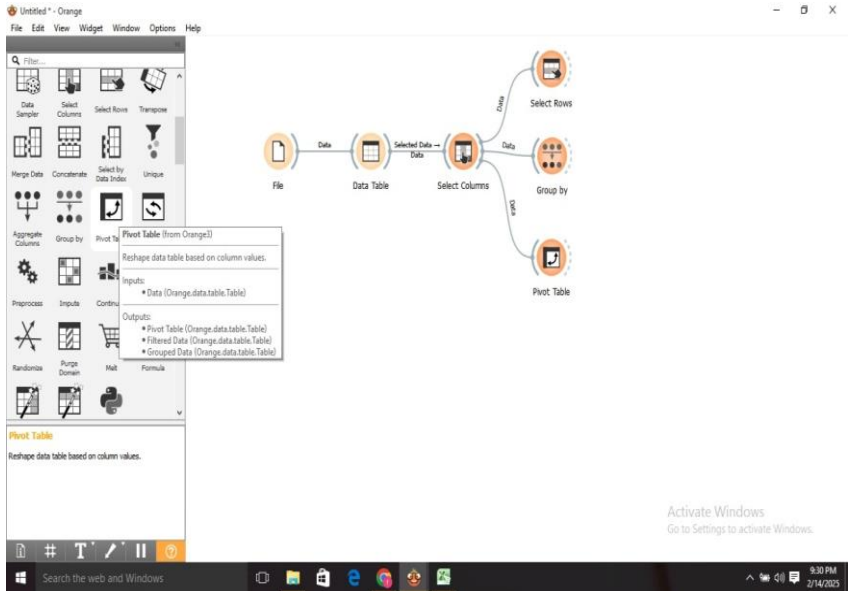
Gambar 7. 5

- Seret widget *Select Rows* dari kategori *Transform* ke area kerja. Hubungkan widget *Select Columns* (atau langsung dari Data Table jika Anda tidak ingin memilih kolom terlebih dahulu) ke widget *Select Rows*. Anda bisa melihat contoh alur kerja yang sudah tersambung pada gambar Klik dua kali pada widget *Select Rows*. Jendela ini memungkinkan Anda untuk memfilter data berdasarkan kondisi tertentu, misalnya memilih baris di mana nilai di kolom "Kopi Aren" adalah 1.



Gambar 7. 6

7. Seret widget *Pivot Table* dari kategori *Transform* ke area kerja. Hubungkan widget sebelumnya ke widget *Pivot Table*. Klik dua kali pada widget *Pivot Table*. Jendela ini akan memungkinkan Anda untuk mengubah bentuk tabel data, biasanya untuk merangkum data berdasarkan kategori, mirip dengan fitur PivotTable di Microsoft Excel. Gambar menunjukkan alur kerja yang sudah mencakup *Pivot Table*.



Gambar 7. 7

Hasil Pembahasan

Hasil implementasi Orange memperkuat temuan dari RapidMiner bahwa pola pembelian kopi konsumen lebih dominan pada varian yang memiliki rasa manis atau campuran. Visualisasi berbasis grafik pada Orange juga mempermudah interpretasi hasil, sehingga memvalidasi bahwa pola kopi favorit pelanggan adalah kombinasi Kopi Aren – Kopi Caramel – Kopi Pandan.

DAFTAR PUSTAKA

<https://www.scribd.com/document/731943817/DATA-MINING-studi-kasus-AMAZON>

<https://www.kompasiana.com/nurhadijah1630/63509a44470fe a6d2d3910f3/sejarah-dan-penerapan-data-mining>

<https://www.djkn.kemenkeu.go.id/artikel/baca/16124/Text-Mining-Sebuah-Proses-untuk-Memanfaatkan-Lautan-Data-di-Era-Industri-40.html>

https://books.google.co.id/books/about/DATA_MINING.html?hl=id&id=rTImDwAAQBAJ&redir_esc=y

<https://www.scribd.com/document/358427176/Makalah-Data-Mining>

https://books.google.co.id/books/about/DATA_MINING.html?hl=id&id=rTImDwAAQBAJ&redir_esc=y

<https://elibrary.bsi.ac.id/readbook/201016/data-mining>

https://books.google.co.id/books/about/DATA_MINING.html?hl=id&id=rTImDwAAQBAJ&redir_esc=y Daftar Pustaka

Inmon, W. H. (2005). Building the Data Warehouse. Wiley.

Santoso, B. (2007). Data Mining: Konsep dan Aplikasi. Andi Offset.

Turban, E., Sharda, R., & Delen, D. (2005). Decision Support and Business Intelligence Systems. Pearson.

Pramudiono, A. (2007). Data Mining: Konsep dan Aplikasi dalam Bisnis. Graha Ilmu.

DATA MINING Studi Kasus AMAZON | PDF

Sejarah dan Penerapan Data Mining Halaman 1 - Kompasiana.com

Text Mining : Sebuah Proses untuk Memanfaatkan Lautan Data di Era Industri 4.0

MDATA MINING Studi Kasus AMAZON. (2022). [PDF]. Diakses dari [URL jika tersedia].

Sejarah dan Penerapan Data Mining. (n.d.). Kompasiana.com. Diakses dari [URL jika tersedia].

<https://www.kompasiana.com/nurhadijah1630/63509a44470fea6cd2d3910f3/sejarah-dan-penerapan-data-mining>

<https://www.djkn.kemenkeu.go.id/artikel/baca/16124/Text-Mining-Sebuah-Proses-untuk-Memanfaatkan-Lautan-Data-di-Era-Industri-40.html>

Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.

Pritchard, J. (2014). *Data Mining: A Comprehensive Guide to Data Mining Techniques and Applications*. CreateSpace Independent Publishing Platform.

Witten, I. H., Frank, E., & Hall, M. A. (2016). *Data Mining: Practical Machine Learning Tools and Techniques* (4th ed.). Morgan Kaufmann.

Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.

Pritchard, J. (2014). *Data Mining: A Comprehensive Guide to Data Mining Techniques and Applications*. CreateSpace Independent Publishing Platform.

Witten, I. H., Frank, E., & Hall, M. A. (2016). *Data Mining: Practical Machine Learning Tools and Techniques* (4th ed.). Morgan Kaufmann. Demšar, J., Curk, T., Erjavec, A., et al. (2013). Orange: Data Mining Toolbox in Python. *Journal of Machine Learning Research*, 14, 2349-2353.

TENTANG PENULIS



Cindy Claudia lahir di Purbatua, Sumatera Utara pada 18 Oktober 2003. Pada saat ini penulis berdomisili di Jl. Manaf Lubis Perumahan Griya Anugrah, Kecamatan Rantau Utara, Kabupaten Labuhanbatu. Memulai pendidikan dasar di SD Negeri 114616 Aek Tayas, lalu melanjutkan ke jenjang pendidikan menengah pertama di SMP Swasta Pemda Rantauprapat. Setelah itu, penulis memilih untuk menempuh pendidikan menengah atas di SMA Negeri 3 Rantau Utara, Penulis telah menyelesaikan pendidikan tinggi di Universitas Labuhanbatu pada tahun 2025 dan memperoleh gelar Sarjana Komputer (S.Kom) dari Fakultas Sains dan Teknologi. Saat ini, penulis berprofesi sebagai Administrator Faktur Penjualan di salah satu perusahaan di Rantauprapat. Selain itu, penulis juga pernah mengikuti beberapa lomba Business Proposal dan mendapatkan Medali emas dan perak, penulis juga pernah menduduki juara 3 di kompetisi Krenova (Kreasi dan Inovasi) Yang diselenggarakan oleh Pemerintah kabupaten Labuhanbatu pada tahun 2024, Buku berjudul **“Konsep Dasar Data Mining Menggunakan Aplikasi Rapidminer dan Orange”** ini merupakan karya pertama penulis. Melalui buku ini, penulis berharap dapat memberikan manfaat bagi pembaca sekaligus pemahaman yang komprehensif. Penulis dapat dihubungi melalui email : claudi.cindy123@gmail.com



Dr. Iwan Purnama, S.Kom., M.Kom, Lahir di Ajamu tanggal 12 Februari 1992, menempuh pendidikan SDN 117844 Ajamu, MTS Al-Ikhlas Ajamu, MAS Al-Ikhlas Ajamu, AMIK Stiekom Sumatera Utara (03), STMIK Triguna Dharma Medan (SI),

Universitas Putra Indonesia "YPTK" Padang (S2) dan Universitas Negeri Padang (S3). Saat ini Dosen dan Ketua Program Studi Teknologi Informasi Fakultas Sains dan Teknologi Universitas Labuhanbatu. Pendiri Lembaga Kursus dan Pelatihan (LKP) IBAY KOMPUTER, Pendiri WebBagusmurah.com. Aktif di kegiatan organisasi APTIKOM, Himpunan Dosen Gemilang Indonesia, Yayasan Labuhanbatu Berbagi Gemilang, Ketua DWP Forum Komunikasi Dosen dan lainnya. Hubungi saya di iwanpurnama2014@gmail.com, Youtube @IbayKomputer dan <https://s.id/iwanpumama>.



ROHANI, dilahirkan di TD. Bolon, 30 Oktober 1987 sebagai anak pertama dari empat bersaudara atas pasangan Bapak Daud pasaribu dengan Ibu Saanah. Pendidikan formal yang ditempuh penulis dimulai dari Pendidikan Dasar di SD N 115476 Panjang Bidang pada tahun 1995 dan lulus tahun 2000. Kemudian dilanjutkan pendidikan menengah

pertama di MTS Swasta Gunting Saga dan lulus tahun 2003. Pada tahun yang sama melanjutkan pendidikan menengah di MAN Damuli Kebun. Melalui ujian masuk perguruan tinggi pada tahun 2006 penulis meneruskan pendidikan pada jurusan Pendidikan Matematika di IAIN SUMATERA UTARA, dan menyelesaikannya pada tahun 2010. Pada tahun 2011 penulis melanjutkan lagi ke Program Pascasarjana Universitas Negeri Medan. Jurusan Pendidikan Matematika dengan judul tesis “Pengembangan Perangkat Pembelajaran Menggunakan Pendekatan Matematika Realistik Untuk Meningkatkan Kemampuan Pecahan Masalah Dan Disposisi Matematika Siswa SMP Muhammadiyah-24 Aekkanopan, dan telah dipertanggung jawabkan di depan penguji pada bulan Oktober 2013. Pekerjaan yang pernah dilalui adalah Guru Matematika di SMP Swasta Muhammadiyah-24 Akkanopan, SMK Zauhari sampai than 2015 dan tahun2014 sampai sekarang dosen di Fakultas Sains dan Teknologi, Universitas Labuhanbatu, dan dapat dihubungi di email: pasariburohani@gmail.com.



Rahmadani Pane, S.Kom., M.Kom

Penulis dilahirkan di Marindal pada tanggal 10 Mei 1986, memperoleh gelar Utama pada tahun 2008, kemudian melanjutkan ke Magister Komputer di Universitas Putra Indonesia Yptk Padang pada tahun 2011, saat ini menjabat sebagai

Ketua Program Studi Teknologi Informasi, Fakultas Sains dan Teknologi, Universitas Labuhanbatu, Labuhanbatu, Indonesia, fokus penelitiannya adalah pada bidang Basis Data, Pemrograman WEB, IoT, Sistem Pendukung Keputusan, dan Kecerdasan Buatan, dapat dihubungi di email: rahmadanipane@gmail.com.

Memahami data mining kini menjadi kunci di era informasi. Buku ini menyajikan panduan komprehensif, dimulai dari teori dasar hingga implementasi praktis menggunakan RapidMiner dan Orange. Ditulis untuk mahasiswa dan praktisi, buku ini dirancang agar setiap pembaca, tanpa memandang tingkat keahlian, dapat menguasai teknik-teknik data mining. Dengan membaca buku ini, Anda akan siap menghadapi tantangan analisis data dan memanfaatkan potensi data secara maksimal.



ISBN 978-634-04-2315-0 (PDF)



9

786340

423150