

Analisis Minat Pelanggan

Pada Penjualan Parfum

Menggunakan Metode K-Nearest Neighbor
dan Algoritma Apriori

Evi Yulia Sinta
Budianto Bangun, S.Sos., M.Kom.
Irmayanti, S.Si., M.Pd.
Angga Putra Juledi, S.Kom., M.Kom.

 Penerbit
litrus.

**Analisis Minat Pelanggan pada Penjualan Parfum
Menggunakan Metode K-Nearest Neighbor dan Algoritma
Apriori**

Ditulis oleh:

Evi Yulia Sinta

Budianto Bangun, S.Sos., M.Kom.

Irmayanti, S.Si., M.Pd.

Angga Putra Juledi, S.Kom., M.Kom.

Diterbitkan, dicetak, dan didistribusikan oleh

PT. Literasi Nusantara Abadi Grup

Perumahan Puncak Joyo Agung Residence Kav. B11 Merjosari

Kecamatan Lowokwaru Kota Malang 65144

Telp: +6285887254603, +6285841411519

Email: literasinusantaraofficial@gmail.com

Web: www.penerbitlitnus.co.id

Anggota IKAPI No. 340/JTI/2022



Hak Cipta dilindungi oleh undang-undang. Dilarang mengutip
atau memperbanyak baik sebagian ataupun keseluruhan isi buku
dengan cara apa pun tanpa izin tertulis dari penerbit.

Cetakan I, Maret 2025

Perancang sampul: Rosyiful Aqli

Penata letak: Dicky Gea Nuansa

ISBN:

viii + 62 hlm. ; 15,5x23 cm.

©Maret 2025

Prakata

Puji syukur kami panjatkan ke hadirat Tuhan Yang Maha Esa, karena berkat rahmat dan karunia-Nya, buku monograf yang berjudul “Analisis Minat Pelanggan pada Penjualan Parfum Menggunakan Metode K-Nearest Neighbor dan Algoritma Apriori” ini dapat diselesaikan dengan baik. Buku ini disusun sebagai bentuk kajian ilmiah mengenai penerapan metode data mining dalam menganalisis pola pembelian dan preferensi pelanggan terhadap produk parfum.

Dalam dunia bisnis, memahami perilaku pelanggan menjadi faktor penting dalam meningkatkan strategi pemasaran dan penjualan. Oleh karena itu, metode K-Nearest Neighbor (K-NN) dan algoritma Apriori digunakan dalam penelitian ini untuk mengidentifikasi pola minat pelanggan berdasarkan data transaksi yang tersedia. Dengan pendekatan ini, diharapkan pelaku usaha dapat mengambil keputusan yang lebih akurat dalam menentukan strategi pemasaran dan pengelolaan stok produk.

Kami menyadari bahwa buku ini masih jauh dari sempurna. Oleh karena itu, kami mengharapkan kritik dan saran yang membangun dari pembaca guna perbaikan di masa mendatang. Semoga buku ini dapat memberikan manfaat bagi akademisi, praktisi bisnis, serta siapa saja yang tertarik dalam bidang analisis data dan pengelolaan bisnis berbasis data mining.

Akhir kata, kami mengucapkan terima kasih kepada semua pihak yang telah membantu dalam proses penyusunan buku ini. Semoga ilmu yang terkandung di dalamnya dapat bermanfaat dan menjadi inspirasi bagi pengembangan penelitian di bidang yang relevan.

Kata Pengantar

Bisnis penjualan parfum terus berkembang pesat seiring dengan meningkatnya kesadaran masyarakat akan pentingnya penampilan dan aroma tubuh. Di tengah persaingan yang semakin ketat, pelaku usaha di sektor ini menghadapi tantangan untuk memahami dan memenuhi kebutuhan serta preferensi pelanggan secara efektif. Dalam konteks ini, penting bagi perusahaan untuk menganalisis minat pelanggan agar dapat menyusun strategi pemasaran yang lebih tepat sasaran. Salah satu permasalahan utama yang dihadapi adalah kurangnya pemahaman tentang perilaku dan preferensi konsumen. Selain menjaga kualitas produk parfum, pelaku usaha parfum dituntut untuk melakukan berbagai inovasi serta meningkatkan nilai tambah bisnis untuk mempertahankan daya saing dan meningkatkan pembelian konsumen dengan menyusun strategi perencanaan bisnis yang tepat.

Penulis menyadari bahwa buku ini masih memiliki kekurangan, oleh karena itu, kritik dan saran yang membangun dari pembaca sangat diharapkan demi penyempurnaan di masa mendatang. Ucapan terima kasih yang sebesar-besarnya penulis sampaikan kepada semua pihak yang telah memberikan dukungan dan kontribusi dalam penyusunan buku ini.

Semoga buku ini dapat memberikan manfaat dan menjadi inspirasi bagi pembaca dalam mengembangkan strategi bisnis berbasis data mining.

Daftar Isi

Prakata	iii
Kata Pengantar	v
Daftar Isi	vii

BAB I

Pendahuluan	1
--------------------------	----------

BAB II

Konsep Dasar Data Mining	5
Pengertian Data Mining	5
Knowledge Discovery in Database.....	6

BAB III

Konsep Minat Pelanggan pada Penjualan Parfum	9
Kerangka Kerja Kajian	9
Analisis Minat Data Penjualan Parfum	13
Transformasi Data	15

BAB IV

Pengenalan Metode K-Nearest Neighbor dan Penerapannya pada Penjualan Parfum	17
Pengertian K-Nearest Neighbor	17
Klasifikasi	18
Uji Performa	20

BAB V

Pengenalan Algoritma Apriori dan Penerapannya

pada Penjualan Parfum.....	27
Pengertian Algoritma Apriori.....	27
Association Rule	28
Uji Performa	29

BAB VI

Rapidminer sebagai Alat Bantu Pengaplikasian

Pengertian Rapidminer.....	35
Implementasi Rapidminer pada Metode K-Nearest Neighbor.....	37
Implementasi Rapidminer Pada Algoritma Apriori.....	43

BAB VII

Kesimpulan dan Rekomendasi

Kesimpulan	53
Daftar Pustaka.....	55
Tentang Penulis.....	59

BAB I



Pendahuluan

Toko Sultan Parfum adalah toko yang bergerak dibidang penjualan parfum dengan berbagai macam merek dan varian, toko ini telah berkembang cukup lama yang berada di Jl.Bima, Lohsari 1, Perlabian. Toko ini menjual parfum dari berbagai merek dan varian. Kegiatan penjualan dilakukan setiap hari maka dari itu data transaksi penjualan akan meningkat dan menumpuk.

Data penjualan pada toko parfum hanya dipakai sebagai rekap data penjualan dan arsip saja, sedangkan data tersebut bisa diproses kembali untuk mendapatkan sebuah informasi baru yang berguna untuk mengetahui jenis parfum apa yang paling banyak diminati dan tidak diminati customer agar tidak terjadi penumpukan stok barang sehingga dapat meningkatkan penjualan parfum. Terkadang hasil pemrosesan data tidak bisa memperoleh dengan hasil yang baik sebab volume data yang diproses itu besar, oleh karena itu diperlukan adanya pendekatan baru agar dapat memperoleh informasi strategis.

Untuk mengatasi masalah ini, penulis menggunakan Metode *K-Nearest Neighbor (KNN)* dan *Algoritma Apriori* adalah sebuah algoritma

yang populer dan sederhana dalam pembelajaran mesin. Metode ini digunakan baik untuk klasifikasi maupun regresi, dan beroperasi dengan konsep bahwa objek dengan fitur yang serupa atau berada dalam lingkungan yang dekat, cenderung memiliki label atau nilai yang sama. Algoritma Apriori adalah metode yang digunakan untuk menemukan pola asosiasi dalam data. Dalam konteks penjualan parfum, algoritma ini dapat membantu dalam mengidentifikasi kombinasi produk yang sering dibeli bersamaan oleh pelanggan. Informasi ini sangat berharga bagi perusahaan dalam merancang strategi bundling produk atau promosi yang lebih menarik. Dengan memanfaatkan kedua metode ini, perusahaan dapat memperoleh wawasan yang lebih mendalam mengenai preferensi dan perilaku konsumen.

Penerapan KNN dan *algoritma apriori* dalam analisis minat pelanggan juga dapat meningkatkan pengalaman pelanggan. Dengan memahami apa yang diinginkan oleh pelanggan, perusahaan dapat memberikan rekomendasi

produk yang lebih relevan, sehingga meningkatkan kepuasan pelanggan dan loyalitas terhadap merek. Hal ini juga berpotensi meningkatkan penjualan dan profitabilitas perusahaan dalam jangka panjang. Dalam persaingan pasar yang semakin ketat, perusahaan parfum harus lebih inovatif dan responsif terhadap kebutuhan pelanggan. Oleh karena itu, penelitian ini tidak hanya berfokus pada pemahaman minat pelanggan, tetapi juga memberikan solusi praktis bagi perusahaan untuk meningkatkan strategi pemasaran mereka. Tantangan lain yang dihadapi adalah perubahan tren pasar yang cepat dan fluktuasi minat konsumen. Dengan adanya media sosial dan influencer, preferensi pelanggan dapat berubah dalam waktu singkat. Oleh karena itu, perusahaan perlu mengadaptasi strategi mereka secara dinamis untuk tetap relevan dan menarik di mata pelanggan. Analisis berbasis data dapat membantu perusahaan merespons perubahan ini dengan lebih cepat dan tepat. Penelitian ini bertujuan untuk memberikan kontribusi yang berarti bagi pengembangan strategi pemasaran dalam industri parfum. Dengan

menggunakan metode *K-Nearest Neighbor* dan *Algoritma Apriori*, diharapkan perusahaan dapat memahami dan memenuhi kebutuhan pelanggan dengan lebih baik, sehingga dapat menciptakan hubungan yang lebih kuat dan berkelanjutan dengan konsumen. Hasil dari penelitian ini diharapkan dapat menjadi referensi bagi perusahaan dalam merumuskan kebijakan dan strategi penjualan yang lebih efektif di masa depan.



BAB II

Konsep Dasar Data Mining

Pengertian Data Mining

Data mining berasal dari dua kata yakni “data” dan “mining”. Dalam buku *Algoritma Data Mining dan Pengujian*, data ialah kumpulan fakta yang terekam atau sebuah entitas yang tidak memiliki arti selama ini terabaikan, Mining sendiri bermakna proses penambangan. Sehingga *data mining* dapat diartikan sebagai proses penambangan data yang menghasilkan sebuah *output* (luaran) berupa pengetahuan. Pengertian lain menyebutkan bahwa *data mining* adalah rangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data. Informasi yang diperoleh, dihasilkan dari mengekstraksi dan mengenali pola yang penting atau menarik dari data yang ada dalam basis data.

Data mining adalah suatu istilah yang digunakan untuk menemukan pengetahuan yang tersembunyi di dalam database. Proses semi otomatis yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi

pengetahuan potensial dan berguna yang bermanfaat yang tersimpan di dalam database besar.

Tahapan Dalam Proses Data Mining

Data mining merupakan melakukan proses terhadap data yang sudah dikumpulkan terlebih dahulu baik data yang dikumpulkan sendiri (data primer) ataupun data yang sudah dikumpulkan pada dataset (data sekunder). Seluruh data yang tersedia akan diproses menggunakan teknik data mining untuk mendapatkan sebuah informasi baru yang berguna juga bermanfaat dalam pengambilan keputusan. Sehingga kelengkapan data akan sangat penting pada penelitian komputer terkait data mining.

Secara umum, Data Mining dibagi menjadi dua kategori utama yaitu:

1. Prediktif

Proses untuk menemukan karakteristik penting dari data dalam satu basis data. Teknik data mining yang termasuk *descriptive mining* adalah clustering, asosiasi, dan sequential mining.

2. Deskriptif

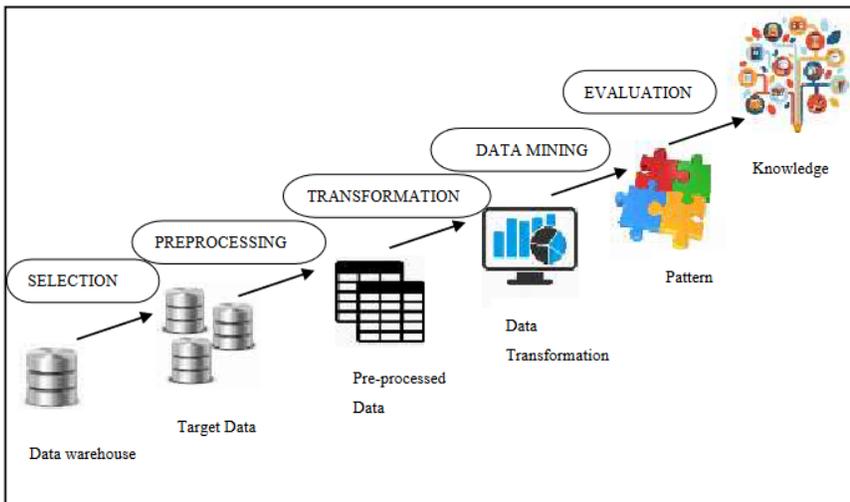
Proses untuk menemukan pola dari data dengan menggunakan beberapa variabel lain di masa depan. Salah satu teknik yang terdapat dalam predictive mining adalah klasifikasi. Secara sederhana data mining biasa dikatakan sebagai proses penyaringan atau “menambang” pengetahuan dari sejumlah data yang besar. *Data mining* sering juga disebut *Knowledge Discovery Databases* (KDD) karena digunakan untuk mencari pengetahuan yang terdapat pada basis data yang besar.

Knowledge Discovery in Database

Knowledge discovery atau *knowledge discovery in database* (KDD) adalah salah satu metode untuk memperoleh pengetahuan dari basis data yang memiliki tabel-tabel yang saling berhubungan atau berelasi. Hasil dari pengetahuan yang didapatkan dari proses tersebut menjadi dasar sebagai basis pengetahuan (*knowledge base*) dalam pengambilan keputusan. Data mining dan *Knowledge discovery* (KDD) dapat menjelaskan pencarian

informasi berguna yang ada pada suatu data yang sangat besar. Namun, kedua istilah ini tetap dua hal yang berbeda walau berkaitan antar satu sama lain. Hal ini karena *data mining* merupakan salah satu proses di dalam KDD. *Knowledge discovery in database* memiliki kaitan dengan proses menemukan pengetahuan pada suatu *database*. Pada proses ini dilakukan identifikasi untuk mencari data yang valid, sesuatu yang baru dan memiliki manfaat sehingga dapat ditemukan suatu pola yang dapat dimengerti yang dapat disebut sebagai suatu proses non-trivial.

Gambaran dari proses KDD terlihat seperti gambar berikut:



Gambar 2.1. *Knowledge Discovery in Database*

Sumber Gambar: <https://www.sciencedirect.com>

Secara keseluruhan, ada lima tahapan dalam proses KDD yang dilakukan secara terurut sebagai berikut :

1. Seleksi data (*data selection*)

Seleksi data merupakan proses menyeleksi kumpulan operasional data sebelum dilakukan penggalian informasi dalam KDD. Data hasil seleksi disimpan dalam suatu berkas yang dipisahkan dari database operasional.

2. Pembersihan data (*data preprocessing/ data cleaning*)
Pada tahapan pembersihan data, dilakukan penghapusan data yang memiliki duplikat, pemeriksaan data yang tidak konsisten, serta perbaikan kesalahan pada data seperti kesalahan cetak. Selain itu, terdapat proses enrichment, yakni mengeksplorasi data yang ada dengan informasi lain yang sesuai dan dibutuhkan KDD, misalnya informasi atau data eksternal.
3. Transformasi (*transformation*)
Pada tahapan transformasi, bentuk data diubah menjadi lebih valid sehingga siap untuk dilakukan proses data mining.
4. Data mining
Pada tahap ini, beragam algoritma atau metode diterapkan. Pemilihan algoritma atau metode yang tepat dipengaruhi oleh tujuan dan proses kdd secara keseluruhan.
5. Evaluasi (*evaluation*)
Pada tahapan terakhir adalah interpretasi/evaluasi. Dengan menggunakan informasi yang bersumber dari data mining, pada tahap ini dibentuk keluaran yang mudah dipahami.

BAB III



Konsep Minat Pelanggan pada Penjualan Parfum

Kerangka Kerja Kajian

Kerangka kerja kajian adalah pondasi intelektual suatu penelitian yang memberikan arah dan struktur terhadap perjalanan penyelidikan. Kerangka ini mencakup landasan teoritis yang mendukung pemahaman konsep-konsep kunci serta hubungan antarvariabel yang ingin dikaji. Dengan demikian, kerangka penelitian membantu memastikan bahwa setiap langkah penelitian memiliki dasar yang kuat, memudahkan interpretasi hasil, dan meningkatkan kontribusi penelitian terhadap pemahaman kita terhadap suatu fenomena.

BAB III

ANALISIS MINAT PELANGGAN PADA PENJUALAN

Kerangka Kerja Kajian

Kerangka kerja kajian adalah pondasi intelektual yang memberikan arah dan struktur terdefinisi. Kerangka ini mencakup landasan konseptual yang mencakup pemahaman konsep-konsep kunci yang relevan yang ingin dikaji. Dengan demikian membantu memastikan bahwa setiap langkah adalah bagian dari kerangka yang kuat, memudahkan interpretasi dan analisis data.

tkan kontribusi penelitian terhadap
suatu fenomena.

Gambar 3.1. Kerangka Kerja Kajian

Dari Gambar 3.1. di atas terlihat jelas bahwa penelitian ini dilakukan secara sistematis dan bertahap, berikut ini adalah penjelasan dari gambar kerangka kerja di atas:

1. Identifikasi Masalah

Langkah awal penelitian ini adalah mengidentifikasi permasalahan utama yang dihadapi Toko Sultan Parfum dalam pengelolaan persediaan barang sembako. Pengumpulan data dikumpulkan dari pelanggan untuk memahami preferensi dan minat mereka terhadap penjualan parfum. Metode KNN digunakan untuk menganalisis pola-pola spasial dalam data tersebut, memungkinkan identifikasi preferensi yang mungkin tidak terlihat secara langsung. Dengan demikian, pengumpulan data menjadi langkah awal yang kritis dalam memahami dan memprediksi minat pelanggan terhadap penjualan parfum dengan menggunakan dua metode analisis yang berbeda.

2. Pengumpulan Dan Pengolahan Data

Data yang dikumpul adalah data penjualan Parfum yang terdiri dari 91 data dan akan diolah pada pengelolaan metode *K-Nearest Neighbor* dan *Algoritma Apriori*.

3. Proses Seleksi Data

Data transaksi penjualan selama 3 bulan terakhir dikumpulkan sebagai bahan analisis utama. Data ini diperoleh dari catatan transaksi manual yang digunakan Toko Sultan Parfum kosong maupun data yang tidak jelas.

4. Analisis KNN dan Algoritma Apriori

KNN digunakan untuk memprediksi kebutuhan stok barang berdasarkan data historis. Dengan menghitung jarak antara data baru dan data historis menggunakan *Euclidean Distance*, metode ini mengklasifikasikan data baru ke dalam kelompok tertentu. Algoritma Apriori digunakan untuk menemukan pola pembelian atau hubungan antara item dalam data transaksi. Dengan menghitung nilai *support* dan *confidence*, algoritma ini menghasilkan aturan asosiasi yang menunjukkan kombinasi barang yang sering dibeli bersama. KNN digunakan untuk memprediksi minat parfum berdasarkan data historis. Dengan menghitung jarak antara data baru dan data historis menggunakan *Euclidean Distance*, metode ini mengklasifikasikan data baru ke dalam kelompok tertentu.

5. Implementasi *Rapidminer*

RapidMiner, sebagai perangkat lunak data mining yang user-friendly, digunakan untuk mempermudah analisis. Data yang telah diproses dimasukkan ke dalam *RapidMiner* untuk menjalankan KNN dan Algoritma Apriori. Implementasi ini memungkinkan analisis yang cepat, efisien, dan meminimalkan kesalahan manual. Fitur visualisasi dari *RapidMiner* juga membantu memahami pola dan prediksi yang dihasilkan.

6. Hasil Evaluasi Dan Model

Evaluasi penting dilakukan untuk melihat sejauh mana sistem dapat berjalan dengan baik dan mengkaji perhitungan algoritma untuk mendapatkan hasil yang optimal.

7. Kesimpulan dan Saran

Tahap terakhir adalah kesimpulan menarik dari hasil penelitian. Kesimpulannya mencakup keunggulan KNN dan Algoritma Apriori dalam mengelola persediaan barang serta efektivitas penggunaan *RapidMiner*. Selain itu, saran diberikan untuk penerapan teknologi serupa pada toko lain atau penelitian lebih lanjut, seperti

pengembangan sistem otomatis berbasis web atau integrasi dengan sistem manajemen toko.

Analisis Minat Data Penjualan Parfum

Analisis data merupakan suatu proses dalam menyederhanakan data agar mudah di baca dan didefinisikan. Penelitian ini menggunakan data selama 3 bulan pada bulan Juli 2024 sampai bulan September 2024 untuk diterapkan pada tahapan langkah data mining. Data tersebut diperoleh dari data manual pada saat observasi. Data transaksi produk akan di ubah ke dalam format data excel kemudian data dimasukkan kedalam rumus *KNN* untuk mendapatkan Asosiasi. Dalam menentukan persediaan barang diperlukan data sampel *item* Produk Parfum dapat dilihat pada tabel dibawah ini.

Tabel 3.1. Data Barang (Nama Produk)

No	Nama Produk
1	Black Opium
2	Taylor Swift
3	Romantic Wish
4	Vanilla Milk
5	Victoria Scandal
6	Lovely
7	Buble Gum
8	Roman Wish
9	Avril Lavigne
10	Baccarat
11	Diamor Sauvage
12	Dunhil Blue
13	Aqua Kiss
14	Annasui Dream
15	Aigner Blue Emotion
16	Kasturi Kijang

No	Nama Produk
17	Jlo Still
18	Diamor Women
19	Gues Pink
20	Selena Gomes
21	Issey Miyake
22	Jesika Parker
23	Kenzo Batang
24	Soft
25	Olla Ramlan
26	Bulgaria Aqua
27	Raffi Ahmad
28	Aqua Man
29	Switzal
30	Pink Chiffon

Transformasi Data

Untuk mempermudah pengolahan data pada data mining, maka data akan diubah kedalam bentuk tabel seperti pada tabel dibawah ini.

Tabel 3.2. Data Transaksi Penjualan (Juli 2024 –September 2024)

Kode Transaksi	Tanggal	Nama Pelanggan	Jenis Kelamin	Nama Item	Jumlah Item Terjual
1	Selasa, 02 Juli 2024	Silva	Perempuan	Dunhil Blue, Black Opium, Taylor Swift, Roman Wish, Buble Gum, Baccarat, Aqua Kiss, Vanilla Milk, Victoria Scandal, Annasui Dream, Aqua Man, Romantic Wish	12
2	Rabu, 03 Juli 2024	Chika	Perempuan	Vanilla Bodies, Victoria Scandal, Vanilla Milk, Rafi Ahmad, Lovely, Argner Black, Taylor Swift, Black Opium	8
3	Kamis, 04 Juli 2024	Tata	Perempuan	Buble Gum, Black Opium, Dunhil Blue, Taylor Swift, Aqua Kiss, Roman Wish, Vanilla Milk, Victoria Scandal, Aigner Black Emotion, Aqua Man, Hugo Orange Man, Dunhil Blue	12

Kode Transaksi	Tanggal	Nama Pelanggan	Jenis Kelamin	Nama Item	Jumlah Item Terjual
4	Jumat, 05 Juli 2024	Febri	Laki-Laki	Aigner Blue Emotion, Victoria Scandal, Romantic Wish, Taylor Wish, Olla Ramlan, Black Opium, Vanilla Milk, Kasturi Kijang	8
5	Sabtu, 06 Juli 2024	Tika	Perempuan	Jlo Still, Roman Wish, Enternity Women, Avril Lauiqne, Nagita Slavina, Black Opium, Victoria Scandal, Romantic Wish, Dunhil Blue, Avril Lavigne, Lovely, Aigner Blue Emotion, Aqua Man	14
91	Senin, 30 September 2024	Mayang	Perempuan	Vanilla Bodies, Black Opium, So Sexy, Buble Gum, Malaikat Subuh, Romantic Wish, Victoria Scandal, Taylor Swift, Vanilla Milk, Annasui Dream, Baccarat, Olla Ramlan, Pinkchiffon, Bulgaria Aqua, Buble Gum, Lovely, Avril Lavigne, Annasui Dream, Diamor Sauvage, Switzal, Raffi Ahmad, Gues Pink, Issey Miyake, Jesika Parker	24

BAB IV



Pengenalan Metode K-Nearest Neighbor dan Penerapannya pada Penjualan Parfum

Pengertian K-Nearest Neighbor

K-Nearest Neighbor (KNN) adalah salah satu algoritma dalam pembelajaran mesin yang digunakan untuk klasifikasi dan regresi. K-NN adalah algoritma yang sederhana tetapi kuat dalam mengklasifikasikan objek atau contoh data berdasarkan kelas mayoritas tetangga terdekatnya. Algoritma ini termasuk dalam kategori algoritma pembelajaran berbasis instan (*instance-based learning*). Selain itu, metode *K-Nearest Neighbor* memiliki kelebihan yakni kemudahan dalam implementasi dan kemampuan yang baik dalam menangani dataset yang besar.

Metode KNN merupakan algoritma dalam menganalisa data yang digunakan untuk pengelompokan atau perhitungan data dalam jumlah

banyak, tetapi memiliki kelemahan jika nilai k kecil maka hasil kurang akurat, namun metode KNN masih tetap menjadi algoritma yang paling akurat, dengan adanya perhitungan data penjualan yang banyak ini menjadi big data yang harus di analisa dan di manfaatkan untuk meningkatkan penjualan dan menambah pendapatan. Jadi jika di simpulkan Algoritma K-NN adalah cara yang digunakan dalam dunia machine learning untuk proses klasifikasi dan Tujuan dari algoritma ini untuk mengelompokkan objek tertentu berdasarkan data dan atribut pembelajaran. Oleh karena itu, dengan mengimplementasikan algoritma *K-Nearest Neighbor*, didapatkan analisa penjualan produk dengan lebih mudah dengan mendapatkan informasi dari perhitungan data penjualan dan menarik kesimpulan. Dengan Algoritma K-NN menerapkan perhitungan sebagai nilai dari sampel uji baru. Nilai data dekat atau jauhnya tetangga biasanya dihitung berdasarkan jarak menggunakan perhitungan aritmatika yang sering di sebut dengan rumus *Euclidean distance*.

Klasifikasi

Berikut adalah rumus umum untuk uji performa metode KNN, terutama dalam konteks klasifikasi :

$$1. \text{ Akurasi (Accuracy)} = \frac{\text{Jumlah Prediksi Benar}}{\text{Jumlah Total Data}}$$

Akurasi mengukur sejauh mana model dapat memprediksi secara tepat jumlah kelas yang benar. Ini adalah metrik yang umum digunakan, tetapi dapat kurang informatif jika kelas-kelas dalam dataset tidak seimbang.

$$2. \text{ Presisi (Precision)} = \frac{\text{True Positive}}{\text{True Positives} + \text{False Positives}}$$

Presisi mengukur seberapa banyak dari instance yang diprediksi sebagai positif yang sebenarnya positif. Metrik ini berguna ketika

kita ingin memastikan bahwa Ketika model mengklasifikasikan suatu kelas, prediksinya benar-benar akurat.

$$3. \text{ Recall (Recall atau Sensitivitas) } = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

Recall mengukur seberapa banyak dari instance positif yang berhasil diprediksi oleh model. Ini membantu kita mengidentifikasi seberapa baik model mengatasi instance positif yang sebenarnya.

$$4. \text{ F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

F1-Score mencari keseimbangan antara *presisi* dan *recall*. Metrik ini berguna ketika kita ingin mempertimbangkan kedua aspek ini secara bersamaan.

5. *Confusion Matrix* adalah tabel yang memperlihatkan *True Positives* (TP), *True Negatives* (TN), *False Positives* (FP), dan *False Negatives* (FN). Dengan melihat confusion matrix, kita dapat mendapatkan wawasan yang lebih rinci tentang kinerja model pada setiap kelas.

Tabel 3.1. *Confusion Matrix*

Classification	True	Predicted Class False
Actual : True	True Positif (TP)	False Negatif (FN)
Actual : False	False Positif (FP)	True Negatif (TN)

6. *Classification Report* adalah menyediakan *precision*, *recall*, dan *f1-score* untuk setiap kelas. Ini membantu kita memahami kinerja model pada setiap kelas secara terperinci.

Metrik-metrik ini memberikan pandangan yang holistik tentang seberapa baik model KNN Anda berperforma pada tugas klasifikasi. Dalam konteks evaluasi model klasifikasi, penting untuk memilih metrik yang sesuai dengan karakteristik khusus dari dataset dan kebutuhan bisnis.

Uji Performa

Metode *K-Nearest Neighbors* (KNN) adalah salah satu algoritma pembelajaran mesin yang termasuk dalam kategori *supervised learning*, khususnya dalam klasifikasi dan regresi. KNN bekerja dengan cara mencari sejumlah k tetangga terdekat dari suatu data point untuk menentukan kelas atau nilai target dari data tersebut.

Salah satu cara untuk menghitung jarak dekat atau jauhnya tetangga menggunakan metode *Euclidean Distance*. *Euclidean Distance* merupakan salah satu metode perhitungan jarak yang digunakan untuk mengukur jarak dari 2 buah titik dalam *Euclidean space*. Berikut ini adalah rumus Euclidean Distance :

$$d(x,y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Keterangan :

$d(x,y)$ = jarak antara data x ke data

y x_i = nilai X pada training data

y_i = nilai X pada testing data

n = batas jumlah banyaknya data

Jika hasil nilai dari rumus di atas besar maka akan semakin jauh tingkat keserupaan antara kedua objek dan sebaliknya jika hasil nilainya semakin kecil maka akan semakin dekat tingkat keserupaan antar objek tersebut. Objek yang dimaksud adalah training data dan testing data. Dalam algoritma ini, nilai K yang terbaik itu tergantung pada jumlah data. Ukuran nilai K yang besar belum tentu menjadi nilai K yang terbaik begitupun juga sebaliknya. Langkah-langkah untuk menghitung algoritma K-NN:

1. Menentukan nilai K (nilai K dipilih secara manual).
2. Menghitung jarak antara data training dan data testing dengan menggunakan rumus *Euclidian Distance*.
3. Kemudian mengurutkan objek-objek (data training) tersebut ke dalam kelompok yang berdasarkan jarak terkecil.
4. Menetapkan kelas, dimana kelas yang dipilih adalah kelas dengan jumlah nilai K terbanyak pada data testing.

Pengelolaan Data Menggunakan Metode K-Nearest Neighbor

Tabel 4.1. Pembentukan Kuantitas Produk dan Kuantitas Terjual

Nama Produk	Kuantitas Produk	Kuanitas Terjual	Kategori
Black Opium	77	60	Minat
Taylor Swift	61	55	Minat
Romantic Wish	61	65	Minat
Vanilla Milk	50	50	Minat
Victoria Scandal	67	56	Minat
Lovely	27	23	Minat
Buble Gum	35	34	Minat
Roman Wish	41	39	Minat
Avril Lavigne	42	42	Minat
Baccarat	30	29	Minat
Diamor Sauvage	12	5	Tidak Minat
Dunhil Blue	30	28	Minat
Aqua Kiss	28	20	Minat
Annasui Dream	12	10	Minat
Aigner Blue Emotion	35	25	Minat
Kasturi Kijang	13	5	Tidak Minat
Jlo Still	17	10	Minat
Diamor Women	29	22	Minat
Gues Pink	8	4	Tidak Minat
Selena Gomes	24	20	Minat
Issey Miyake	15	15	Minat
Jesika Parker	13	12	Minat

Nama Produk	Kuantitas Produk	Kuantitas Terjual	Kategori
Kenzo Batang	6	2	Tidak Minat
Soft	16	24	Minat
Olla Ramlan	19	15	Minat
Bulgaria Aqua	11	10	Minat
Raffi Ahmad	14	5	Tidak Minat
Aqua Man	15	10	Minat
Switzal	11	5	Tidak Minat
Pink Chiffon	12	4	?

Nilai Min Kuantitas Produk : 6

Nilai Max Kuantitas Produk : 77

Nilai Min Kuantitas Terjual : 2

Nilai Max Kuantitas Terjual : 65

Tabel 4.2. Data Normalisasi Min Max

Min(x)	6	2
Max(x)	77	65

No.	Min Max Kuantitas Produk	Min Max Kuantitas Terjual
1	-1	0,857142857
2	-0,774647887	0,777777778
3	-0,774647887	0,936507937
4	-0,61971831	0,698412698
5	-0,85915493	0,793650794
6	-0,295774648	0,26984127
7	-0,408450704	0,444444444
8	-0,492957746	0,523809524
9	-0,507042254	0,571428571
10	-0,338028169	0,365079365
11	-0,084507042	-0,015873016
12	-0,338028169	0,349206349
13	-0,309859155	0,222222222
14	-0,084507042	0,063492063

No.	Min Max Kuantitas Produk	Min Max Kuantitas Terjual
15	-0,408450704	0,301587302
16	-0,098591549	-0,015873016
17	-0,154929577	0,063492063
18	-0,323943662	0,253968254
19	-0,028169014	-0,031746032
20	-0,253521127	0,222222222
21	-0,126760563	0,142857143
22	-0,098591549	0,095238095
23	0	-0,063492063
24	-0,14084507	0,285714286
25	-0,183098592	0,142857143
26	-0,070422535	0,063492063
27	-0,112676056	-0,015873016
28	-0,126760563	0,063492063
29	-0,070422535	-0,015873016
30	-0,084507042	-0,031746032

Tahap selanjutnya yaitu menentukan parameter K = jumlah tetangga terdekat. Pada penelitian ini nilai K yang digunakan adalah 5. Menghitung jarak antara data testing (uji) dengan semua data training pada tahap transformasi menggunakan perhitungan jarak *Euclidean Distance*, yang dipakai untuk menghitung jarak *Euclidean Distance*.

Menggunakan rumus :

$$\sqrt{(x_2-x_1)^2 + (y_2-y_1)^2}$$

Tabel 4.3. Data Hasil dari Jarak Euclidean Distance

Jarak Euclidean	Ranking	K=1	K=3	K=5
97,61659695	29			
82,13403679	26			
89,14033879	28			
70,71067812	25			
87,32124598	27			

Jarak Euclidean	Ranking	K=1	K=3	K=5
35,4682957	17			
48,7954916	22			
56,5862174	23			
59,39696962	24			
41,72529209	20			
13	4			Minat
41,03656906	19			
34,40930107	16			
15,62049935	8			
43,01162634	21			
13,92838828	5			Minat
19,72308292	11			
36,40054945	18			
8,94427191	2		Tidak Minat	Tidak Minat
31,2409987	15			
21,21320344	12			
17,69180601	9			
6,32455532	1	Tidak Minat	Tidak Minat	Tidak Minat
28,8444102	14			
24,20743687	13			
14,86606875	6			
14,86606875	6			
18,02775638	10			
12,08304597	3		Minat	Minat

Kemudian mengurutkan hasil secara ascending (dari nilai tinggi ke rendah).

Tabel 4.4. Data Hasil Urutan Ascending

No	Jarak Euclidean	Ranking
1	97,61659695	29
2	82,13403679	26
3	89,14033879	28
4	70,71067812	25
5	87,32124598	27
6	35,4682957	17
7	48,7954916	22
8	56,5862174	23
9	59,39696962	24
10	41,72529209	20
11	13	4
12	41,03656906	19
13	34,40930107	16
14	15,62049935	8
15	43,01162634	21
16	13,92838828	5
17	19,72308292	11
18	36,40054945	18
19	8,94427191	2
20	31,2409987	15
21	21,21320344	12
22	17,69180601	9
23	6,32455532	1
24	28,8444102	14
25	24,20743687	13
26	14,86606875	6
27	14,86606875	6
28	18,02775638	10
29	12,08304597	3

Berdasarkan hasil perhitungan jarak diatas, maka didapatkan suatu hasil dengan nilai $k=5$ memiliki keputusan yaitu Minat = 4 dan Tidak Minat= 2.

BAB V



Pengenalan Algoritma Apriori dan Penerapannya pada Penjualan Parfum

Pengertian Algoritma Apriori

Algoritma Apriori adalah suatu algoritma yang sudah sangat dikenal dalam melakukan pencarian *frequent* itemset dengan menggunakan teknik association rule. Algoritma apriori menggunakan *knowledge* mengenai *frequent* itemset yang telah diketahui sebelumnya, untuk memproses informasi selanjutnya. Pada algoritma apriori untuk menentukan kandidat yang mungkin muncul dengan cara memperhatikan minimum *support*. Algoritma apriori termasuk jenis aturan asosiasi pada data mining. Selain algoritma apriori, yang termasuk pada golongan ini adalah metode *generalized rule induction* dan algoritma *hash based*. Aturan yang menyatakan asosiasi antara beberapa atribut sering disebut *affinity analysis* atau market basket analysis. Analisis asosiasi atau *association rule* mining adalah teknik data mining untuk menemukan aturan asosiatif antara

suatu kombinasi item. Salah satu tahap analisis asosiasi yang menarik perhatian banyak peneliti untuk menghasilkan algoritma yang efisien adalah analisis pola frekuensi tinggi (*frequent pattern mining*). Penting tidaknya suatu asosiasi dapat diketahui dengan dua tolok ukur, yaitu : *support* dan *confidence*. *Support* (nilai penunjang) adalah *persentase* kombinasi item tersebut dalam *database*, sedangkan *confidence* (nilai kepastian) adalah kuatnya hubungan antara-item dalam aturan asosiasi.

Algoritma apriori digunakan untuk menemukan *Frequent* itemset yang berguna untuk pembentukan *Association Rules*. Berikut langkah yang dilakukan dalam algoritma hash-based.

1. Pemrosesan Data, dataset yang didapat pada tahap preprocessing kemudian diproses menggunakan algoritma apriori untuk menghasilkan Frequent Itemset.
2. Menentukan minimum *support* sebagai batas dalam pembentukan frequent itemset dan minimum *confidence* dalam pembentukan *Association Rules*. Penentuan minimum *support* dan minimum *confidence* disesuaikan dengan kebutuhan karena tidak ada ketentuannya.

Association Rule

Association rule adalah salah satu task data mining deskriptif yang bertujuan untuk menemukan aturan asosiasi antara item-item data.

Menyebutkan mutlak atau tidaknya suatu aturan asosiasi dapat diketahui dengan dua tolok ukur, yaitu: *support* dan *confidence*. *Support* adalah nilai penunjang atau *persentase* kombinasi sebuah item dalam database. Adapun rumus *support* yang digunakan ialah rumus berdasarkan yang digunakan pada penelitian sebelumnya oleh [13](Albab & Hidayatullah, 2022), ialah sebagai berikut:

$$\text{Support } A = \frac{\text{Juml transaksi mengandung } A}{\text{Total transaksi}} \times 100\%$$

Persamaan (1) merupakan rumus menentukan support.

$$\text{Support } A, B = \frac{\sum \text{Jumlah transaksi mengandung } A \text{ dan } B}{\sum \text{Transaksi}} \times 100\%$$

Persamaan (2) merupakan rumus menentukan support 2 item.

Setelah semua pola frekuensi tinggi ditemukan, kemudian dicari aturan asosiasi yang memenuhi syarat minimum untuk *confidence* dengan menghitung *confidence aturanasosiatif* $A \rightarrow B$.

Nilai *confidence* dari aturan $A \rightarrow B$ diperoleh dari rumus berikut:

$$\text{Confidence} = P(B|A) = \frac{\sum \text{Jumlah transaksi mengandung } A \text{ dan } B}{\sum \text{Transaksi}}$$

Uji Performa

Untuk menghasilkan aturan asosiasi yang relevan, tahap selanjutnya pada tahap awal perhitungan adalah menentukan nilai minimal *support* dan minimal *confidence*. Dalam penelitian ini, nilai minimal *support* yang ditetapkan adalah 50% dan nilai minimal *confidence* adalah 55%. Setelah menetapkan nilai minimal *support* dan *confidence*, langkah berikutnya adalah mencari *support* 1 itemset dan menghitung nilai *support* nya, dengan menggunakan rumus:

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

Tabel 5.1 Support 1 Itemset

No	Item	Transaksi	Support 1 Itemset
1	Black Opium	77	84,61538462
2	Taylor Swift	61	67,03296703
3	Romantic Wish	61	67,03296703
4	Vanilla Milk	50	54,94505495
5	Victoria Scandal	67	73,62637363
6	Lovely	27	29,67032967
7	Buble Gum	35	38,46153846

No	Item	Transaksi	Support 1 Itemset
8	Roman Wish	41	45,05494505
9	Avril Lavigne	42	46,15384615
10	Baccarat	30	32,96703297
11	Diamor Sauvage	12	13,18681319
12	Dunhil Blue	30	32,96703297
13	Aqua Kiss	28	30,76923077
14	Annasui Dream	12	13,18681319
15	Aigner Blue Emotion	35	38,46153846
16	Kasturi Kijang	13	14,28571429
17	Jlo Still	17	18,68131868
18	Diamor Women	29	31,86813187
19	Gues Pink	8	8,791208791
20	Selena Gomes	24	26,37362637
21	Issey Miyake	15	16,48351648
22	Jesika Parker	13	14,28571429
23	Kenzo Batang	6	6,593406593
24	Soft	26	28,57142857
25	Olla Ramlan	19	20,87912088
26	Bulgaria Aqua	11	12,08791209
27	Raffi Ahmad	14	15,38461538
28	Aqua Man	15	16,48351648
29	Switzal	11	12,08791209
30	Pink Chiffon	12	13,18681319

Pada tahap berikutnya memilih data yang memenuhi minimal *support* yaitu 50%, data yang kurang dari 50% akan dihapus dari tabel. Dari proses diatas terdapat 5 produk yang masih memenuhi minimal *support*. Setiap kandidat 1 *itemset* yang memenuhi nilai minimal ini akan dianggap sebagai support 1 *itemset*. Data hasil support 1 *itemset* ditunjukkan pada tabel dibawah ini:

Tabel 4.2. Hasil Support 1 Itemset

No	Item	Transaksi	Support
1	Black Opium	77	87,5
2	Taylor Swift	61	69,31818182
3	Romantic Wish	61	69,31818182
4	Vanilla Milk	50	56,81818182
5	Victoria Scandal	67	76,13636364

Setelah mendapatkan support 1 *itemset*, langkah berikutnya adalah melakukan tahap kedua yaitu menggabungkan atau mengkombinasikan 2 produk yang memenuhi minimal *support*. Disini 2 *itemset* tersebut saling dipasangkan dan tidak boleh ada pasangan yang sama. *Support 2 itemset* dihasilkan melalui perhitungan kombinasi dari 2 produk yaitu perhitungan dukungannya tetap sama seperti pada langkah sebelumnya saat mencari minimal *support 1 itemset*. Untuk menghitung dukungan kombinasi dari dua *item*, dengan menggunakan rumus:

$$Support (A, \cap B) = \frac{\sum \text{Jumlah Transaksi Mengandung A dan B}}{\text{Total Transaksi}} \times 100\%$$

Tabel 5.3. Hasil Support 2 Itemset

No	Item	Transaksi	Support
1	Black Opium, Taylor Swift	56	61,53846154
2	Black Opium, Romantic Wish	53	58,24175824
3	Black Opium, Victoria Scandal	61	67,03296703

Seperti langkah sebelumnya, tahap berikutnya memilih data yang memenuhi minimal *support* yaitu 50%, data yang kurang dari 50% akan dihapus dari tabel. Dari proses diatas terdapat 3 produk yang masih memenuhi minimal *support*. Setiap kandidat 2 *itemset* yang memenuhi nilai minimal ini akan dianggap sebagai *support 2 itemset*.

Jika nilai minimal *support* telah didapatkan maka selanjutnya kita menghitung nilai minimal *confidence*. Untuk mendapatkan nilai minimal *confidence* kita dapat menghitung dengan menggunakan jumlah transaksi

yang mengandung A dan B dibagi jumlah transaksi yang mengandung A. Contoh rumusnya sebagai berikut:

Confidence $P(B|A) =$

$$\frac{\sum \text{Jumlah Transaksi Mengandung A dan B}}{\text{Jumlah Transaksi Mengandung A}} \times 100\%$$

Tabel 5.4. Data Nilai Confidence

No	Aturan	Confidence	Transaksi
1	Jika membeli black opium, maka akan membeli taylor swift	56/80	0,875
2	Jika membeli taylor swift, maka akan membeli black opium	56/64	0,7
3	Jika membeli black opium, maka akan membeli romantic wish	53/80	0,828125
4	Jika membeli romantic wish, maka akan membeli black opium	53/64	0,6625
5	Jika membeli black opium, maka akan membeli victoria scandal	61/80	0,871428571
6	Jika membeli victoria scandal, maka akan membeli black opium	61/70	0,7625

Dari nilai confidence yang dirangkum dengan penentuan minimum confidence sebesar 55% lalu sisihkan yang tidak sesuai dengan syarat minimum confidence. Kombinasi hasil nilai support dan confidence yang tertinggi dapat dilihat dari aturan asosiasi. Dari hasil tersebut maka kita mendapatkan aturan asosiasi seperti dibawah ini:

Tabel 5.5. Hasil Aturan Asosiasi

Jika Membeli	Maka Akan Membeli	Support	Confidenece
Black Opium	Taylor Swift	61,53%	87,50%
Black Opium	Tromantic Wish	58,24%	83%
Black Opium	Victoria Scandal	67,03%	87,14%

Akhir dari penentuan aturan asosiasi diatas, menunjukkan barang yang sering terjual adalah Black Opium dan Victoria Scandal dengan nilai

support 67,03%% serta *confedence* 87,14%, sehingga pemilik mengetahui parfum yang banyak yang diminati.

Tabel 5.6. Hasil 3 Itemset

No	Item			Transaksi
	Black Opium	Taylor Swift	Romantic Wish	
1	1	1	1	Y
2	1	1	0	N
3	1	1	0	N
...
30	1	1	1	Y
Jumlah Transaksi 3 Itemset				38
Support (%)				41,75824176

No	Item			Transaksi
	Taylor Swift	Romantic Wish	Vanilla Milk	
1	1	1	1	Y
2	1	0	1	N
3	1	0	1	N
...
30	1	1	1	Y
Jumlah Transaksi 3 Itemset				33
Support (%)				36,26373626

No	Item			Transaksi
	Romantic Wish	Vanilla Milk	Victoria Scandal	
1	1	1	1	Y
2	0	1	1	N
3	0	1	1	N
...
30	1	1	1	Y
Jumlah Transaksi 3 Itemset				31
Support (%)				34,06593407

Dan hasil akhir dari data 3 itemset yaitu parfum *black opium*, *taylor swift*, dan *romantic wish* dengan *jumlah transaksi* 38 dan *support* 41,75.

BAB VI



Rapidminer sebagai Alat Bantu Pengaplikasian

Pengertian Rapidminer

RapidMiner adalah sebuah perangkat lunak dengan sumber terbuka (*open source*) yang dapat digunakan untuk melakukan analisis pada data mining, text mining, dan analisis prediksi. Aplikasi ini bermanfaat untuk keperluan bisnis dan komersial, penelitian, pendidikan dan pelatihan, serta pengembangan prototipe dan aplikasi dengan cepat. *RapidMiner* juga dapat membantu di semua tahapan proses pembelajaran mesin, termasuk persiapan data, visualisasi hasil, validasi, dan pengoptimalan.

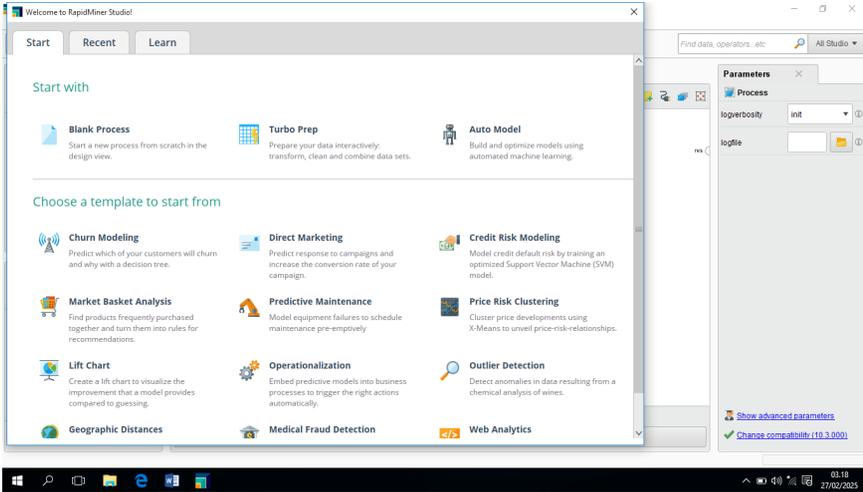
RapidMiner dalam analisis minat pelanggan terhadap penjualan parfum dengan menggunakan metode *K-Nearest Neighbors (KNN)* dan *Metode Algoritma Apriori* membuka peluang besar untuk mendapatkan pemahaman mendalam tentang preferensi konsumen. *RapidMiner* mempermudah pengguna untuk menggabungkan kekuatan *KNN* dan *Algoritma Apriori* dalam satu rangkaian analisis yang terintegrasi. Dengan *RapidMiner*, pengguna dapat mengumpulkan dan mengolah data dari

berbagai sumber, termasuk atribut penting seperti karakteristik konsumen dan preferensi terhadap penjualan parfum. Proses *preprocessing* data yang efisien dapat dilakukan untuk membersihkan, mengubah format, dan menyiapkan data untuk analisis lebih lanjut. *KNN* dapat digunakan untuk mengidentifikasi kelompok konsumen dengan preferensi serupa berdasarkan atribut yang telah ditentukan. Ini memungkinkan organisasi untuk memahami sejauh mana kesamaan atau perbedaan preferensi di antara segmen masyarakat tertentu. *RapidMiner* adalah platform perangkat lunak yang sangat berguna untuk analisis data dan penambangan data. Dengan antarmuka grafisnya yang intuitif, *RapidMiner* memungkinkan pengguna untuk menjelajahi dan memahami struktur data dengan mudah, serta melakukan praproses data untuk meningkatkan kualitasnya sebelum menerapkan teknik penambangan data.

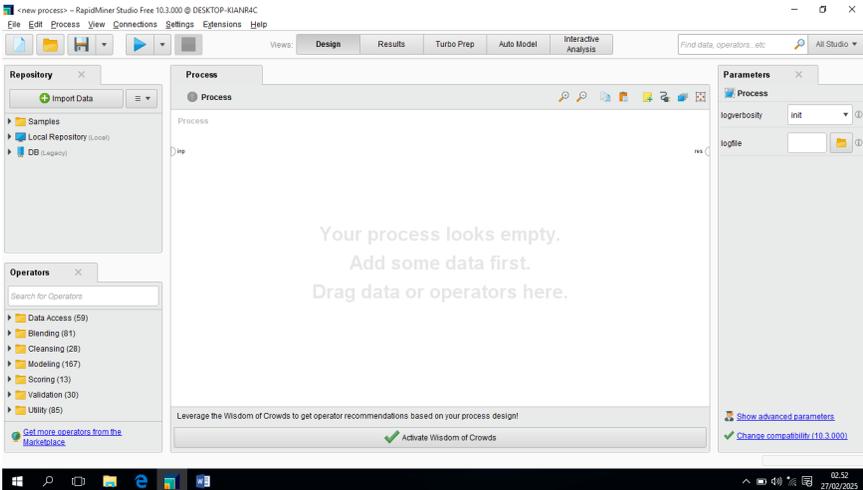
Sementara itu, *RapidMiner* juga mendukung penerapan *Algoritma Apriori*, yang dapat menangkap pola yang lebih kompleks dan hubungan non-linear dalam data. Dengan kemampuan adaptifnya, *Algoritma Apriori* dapat membantu meramalkan minat pelanggan berdasarkan pola yang teridentifikasi, membuka peluang untuk pengambilan keputusan yang lebih tepat. Dengan memanfaatkan *RapidMiner*, organisasi dapat mengoptimalkan strategi pemasaran, merancang produk yang lebih sesuai dengan preferensi pasar, dan mengidentifikasi tren minat masyarakat secara lebih akurat. Sebagai tampilan untuk lingkungan kerja, *Design Perspective* memiliki beberapa *view*. *Operator View* adalah *view* yang paling utama dan penting. Semua operator atau langkah-langkah kerja dalam *RapidMiner* ditampilkan dalam operator *view*.

Implementasi Rapidminer pada Metode K-Nearest Neighbor

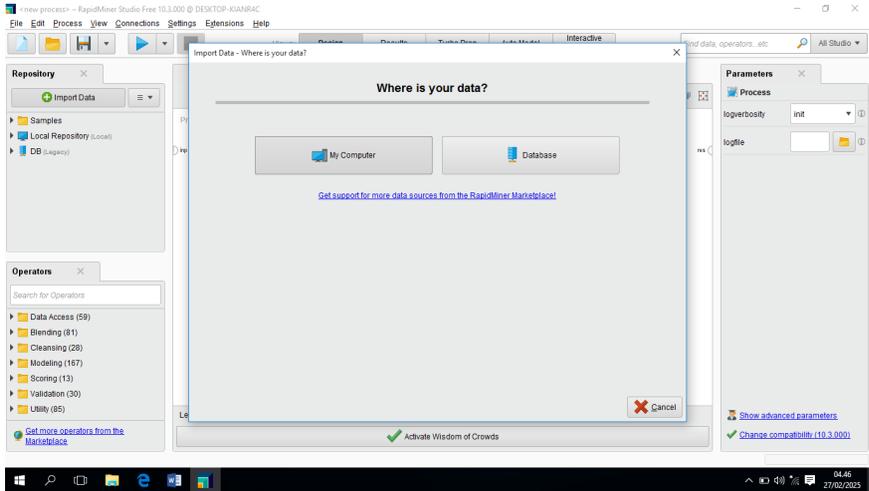
Tahap pertama yang dilakukan yaitu membuka aplikasi rapidminer dan ini merupakan tampilan *rapidminer*.



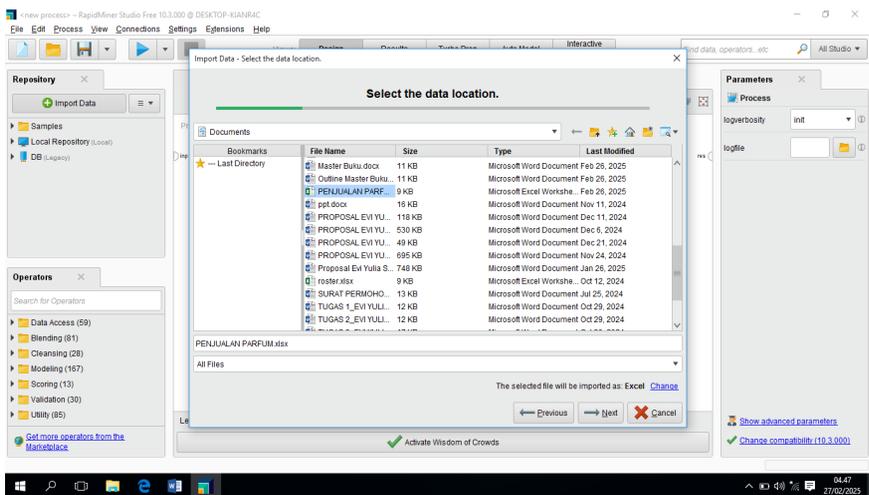
Tahap kedua disini akan membuat sebuah desain dengan metode *K-Nearest Neighbor* lalu klik Blank Process.



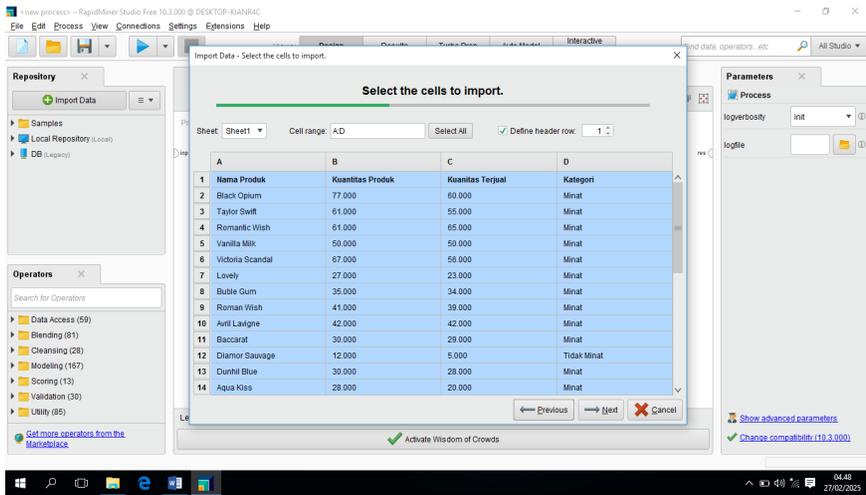
Tahap ketiga ini merupakan tampilan membuat desain nya. Selanjutnya yaitu disini akan mengimport data terlebih dahulu, klik Import Data.



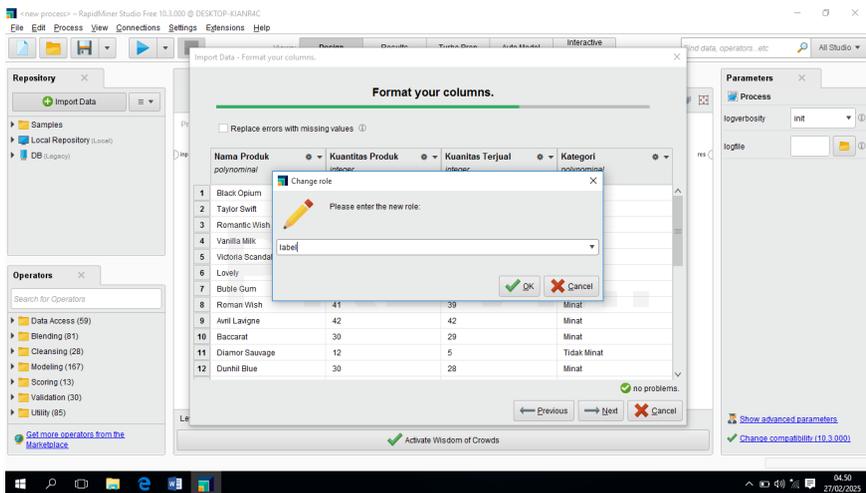
Tahap keempat jika muncul tampilan seperti ini kemudia pilih *My Computer* untuk memilih file datanya.



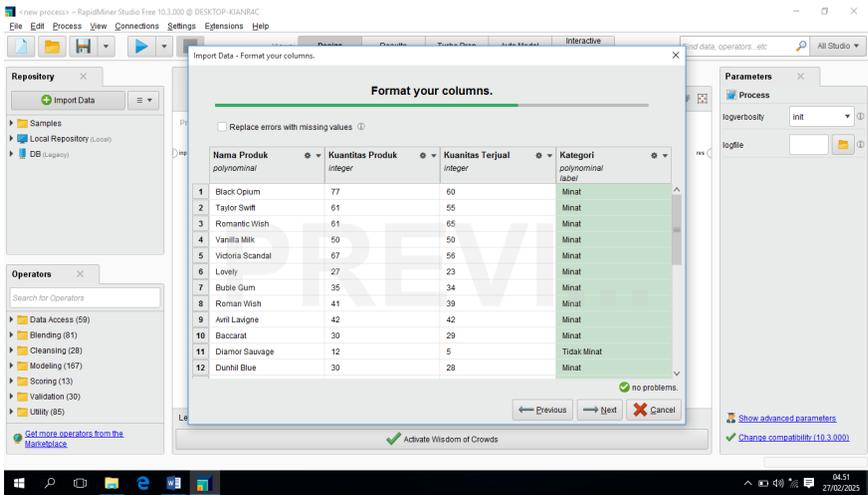
Tahap kelima kemudian pilih file data excel Penjualan Parfum lalu klik next untuk melanjutkan.



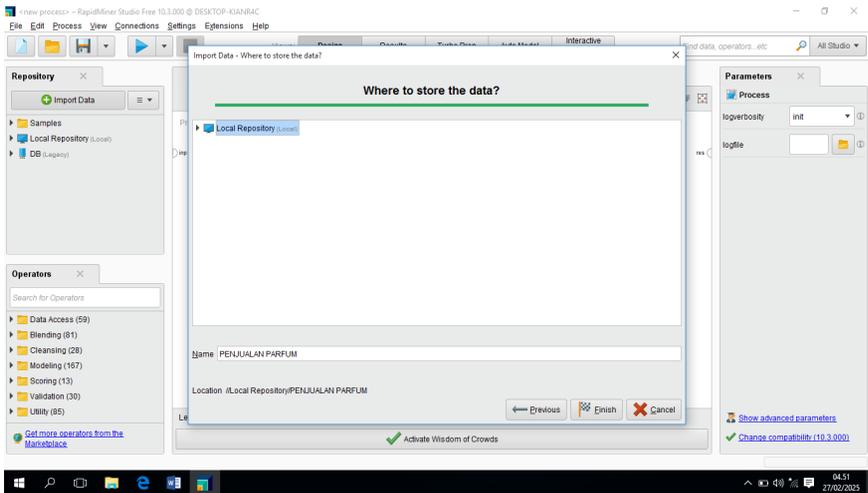
Tahap kelima selanjutnya ada tampilan Select the cells to import, nah ini adalah data excel penjualan parfum yang telah diimport.



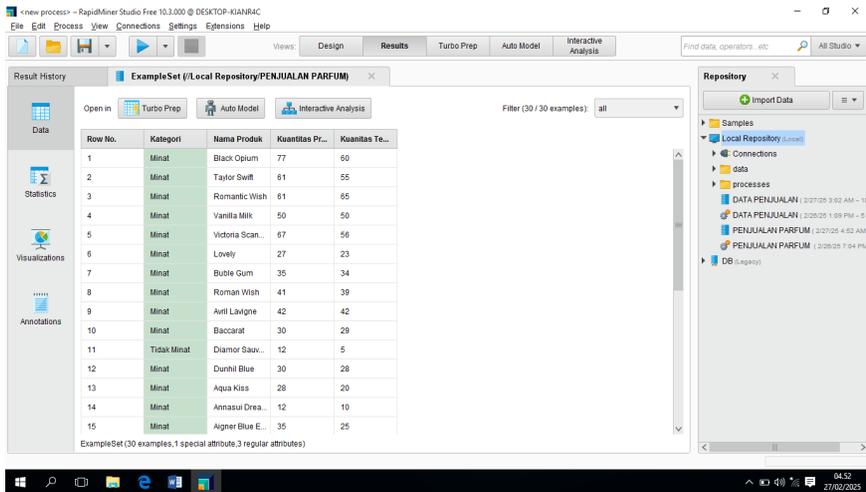
Tahap keenam selanjutnya dikolom kategori kita tambahkan *Change role* nya yaitu label lalu klik ok.



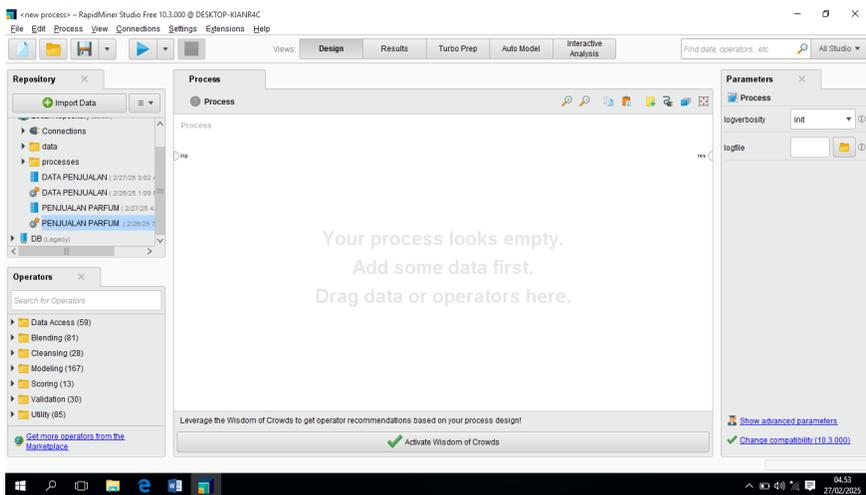
Tahap ketujuh ini merupakan tampilan Format your columns yang dimaksud yaitu data yang sudah diimport ada nama produk, kuantitas produk, kuantitas terjual dan kategori.



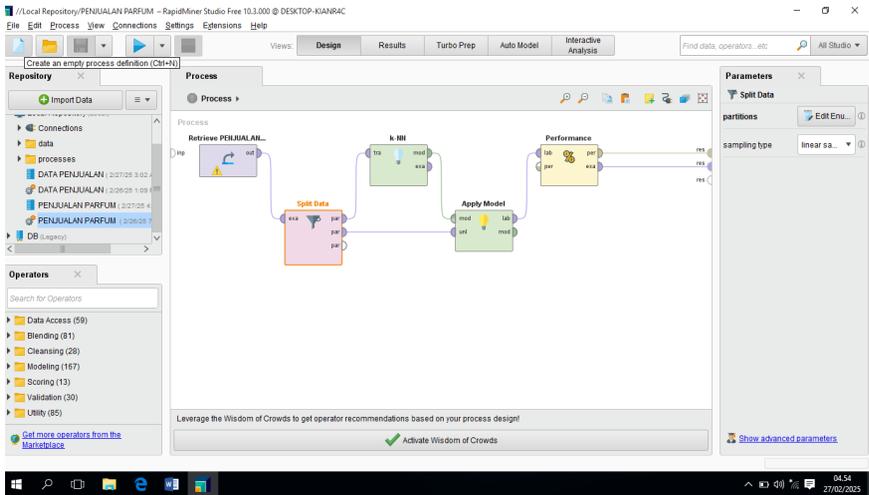
Tahap kedelapan yaitu disini kita disuru pilih tempat data penyimpanan di Local repository kemudian buat nama file datanya yaitu penjualan parfum dan klik finish.



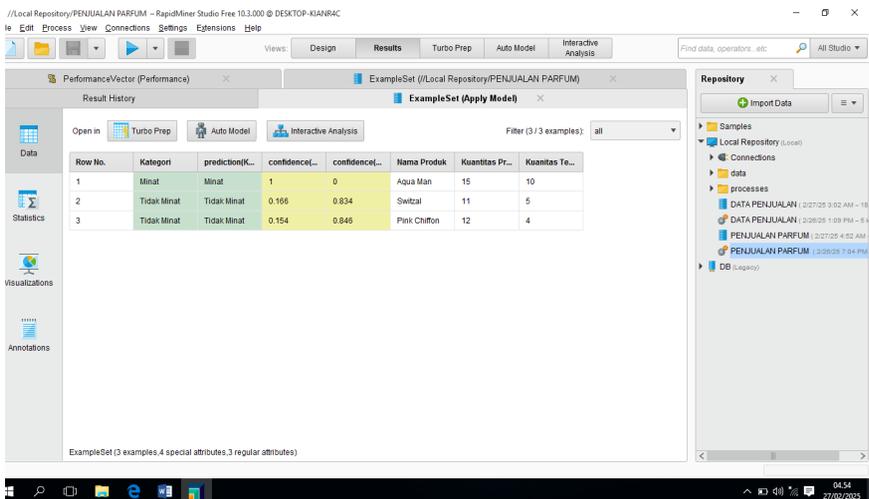
Tahap kesembilan muncullah tampilan data penjualan parfum dibagian *result*.



Tahap kesepuluh kita membuat design pilih data import penjualan parfum.



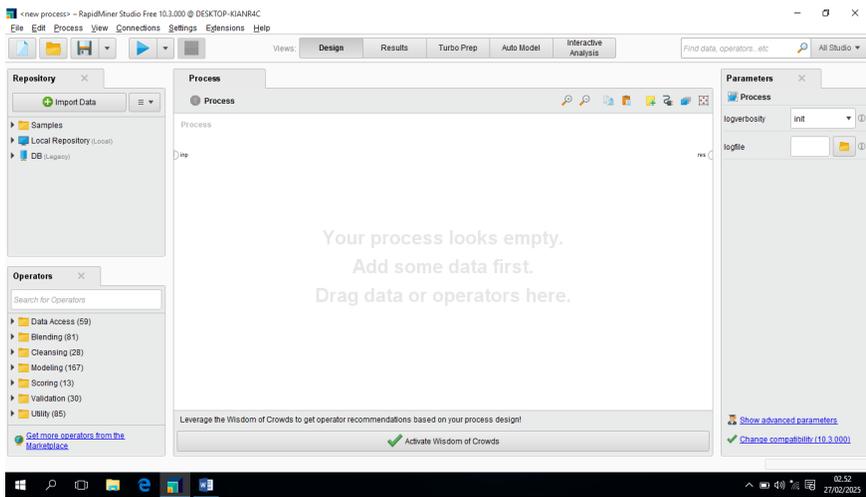
Tahap kesebelas membuat desain nya yaitu yang pertama ada penjualan parfum, kedua split data. Nah dibagian split data ini kita akan ubah dan menambahkan partisi nya menjadi 0,9 dan dibawahnya 0,1 kemudian sampling tipenya menjadi learning sampling. ketiga K-NN, keempat Apply Model, dan yang kelima Performance. Kemudian desain akan dirun kan atau dijalankan.



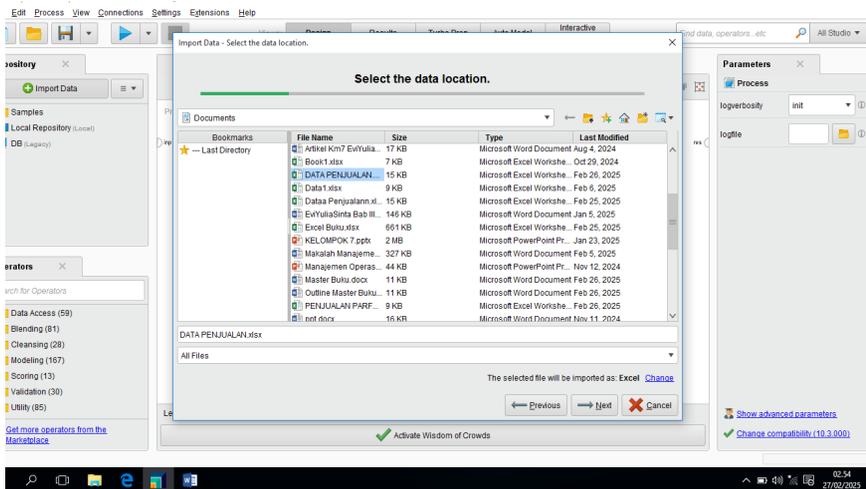
Tahap akhir, dan inilah hasil akhir penggunaan rapidminer menggunakan metode KNN. Ada nama produk yang diminati yaitu aqua man dan yang tidak diminati adalah switzal dan pink chiffon.

Implementasi Rapidminer Pada Algoritma Apriori

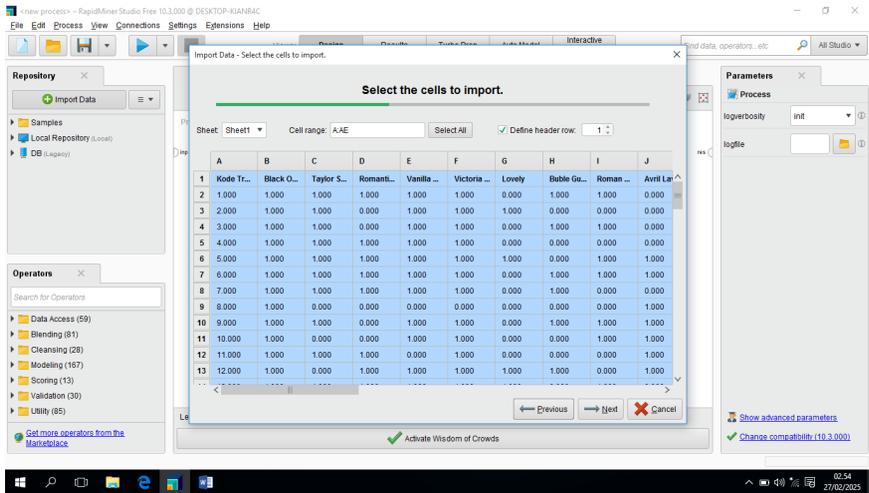
Tahap pertama yaitu import data excel data penjualan parfum yang akan ditambahkan.



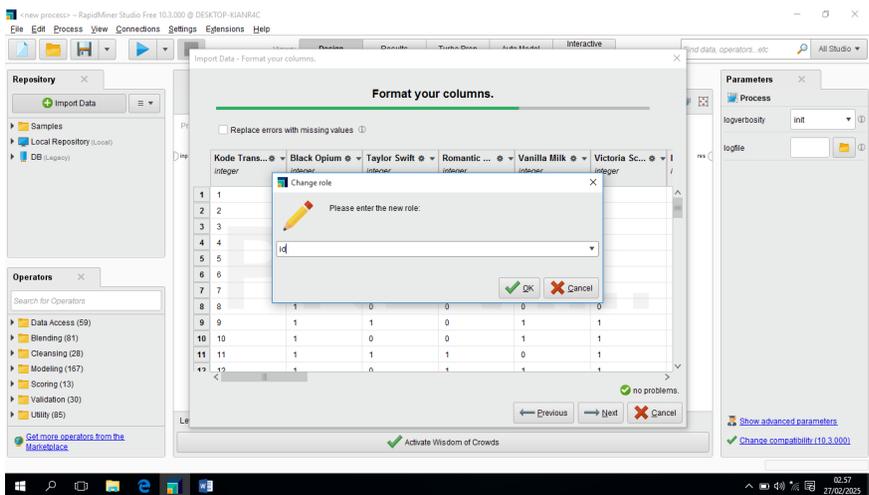
Klik import data nya dan nanti akan memilih data tersebut



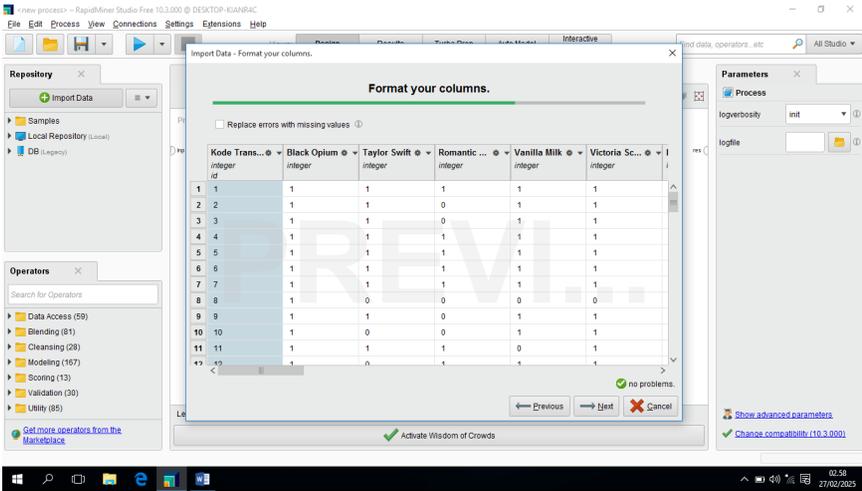
Tahap kedua yaitu pilih data penjualan parfum. Ini merupakan data excel yang berisi data tabular pada data penjualan parfum.



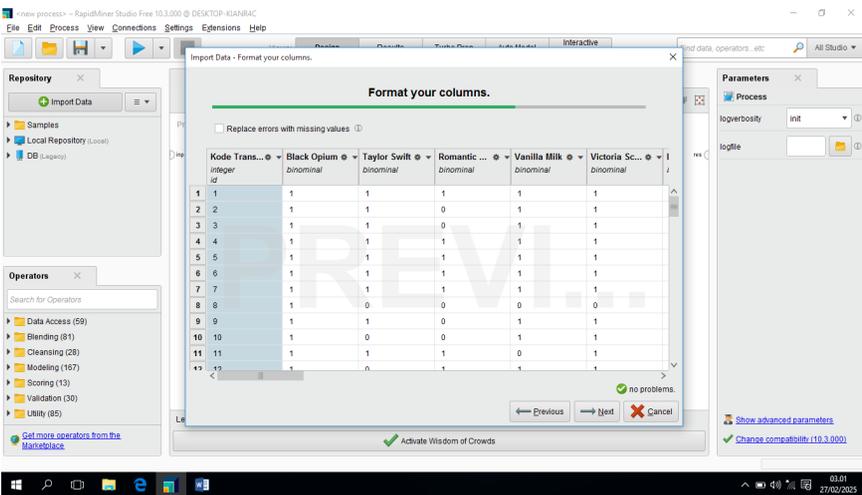
Tahap ketiga kemudian muncul tampilan data yang sudah diimport dari semua data parfum dan data tabularnya.



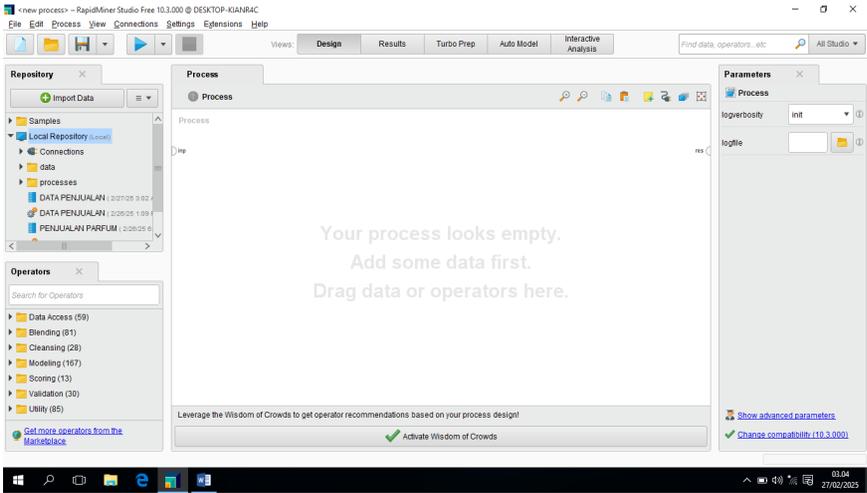
Tahap keempat selanjutnya dibagian Kode Transaksi kita tambahkan Change Role nya yaitu id, setelah itu di ok kan.



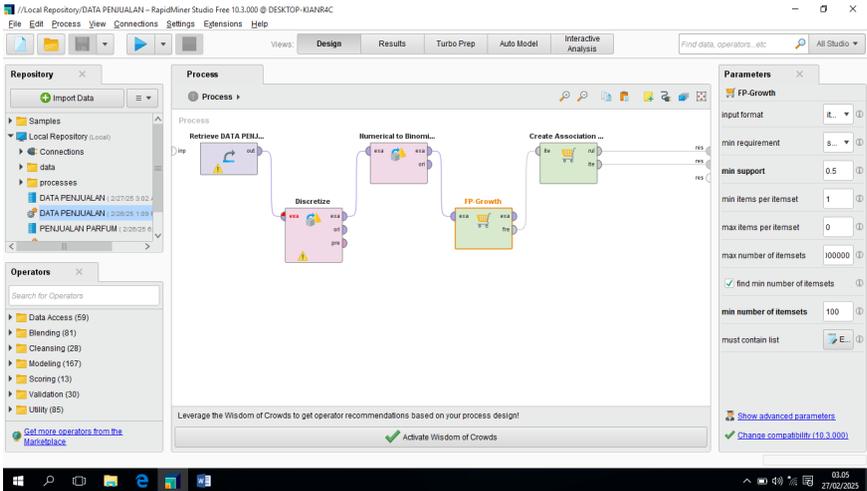
Tahap kelima nah dibagian jenis parfum masih integer dan kita akan ubah menjadi binominal.



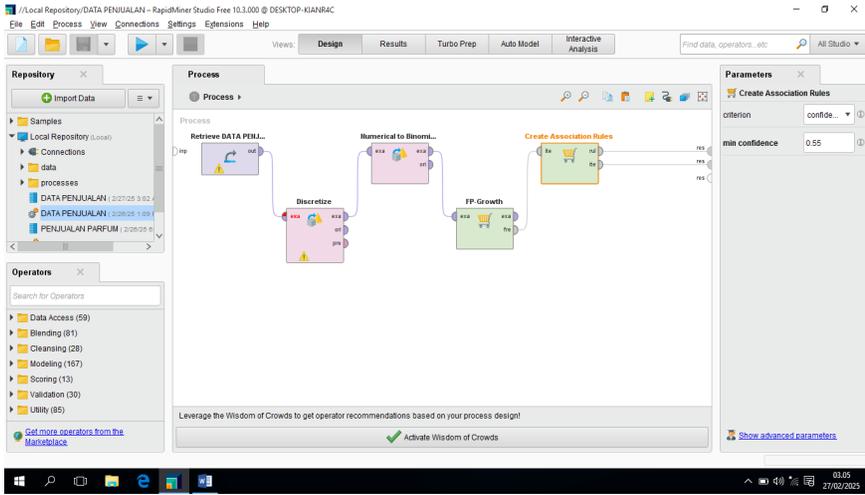
Tahap keenam sudah diubah yang tampilan awal menjadi integer dan sekarang sudah berubah menjadi binominal.



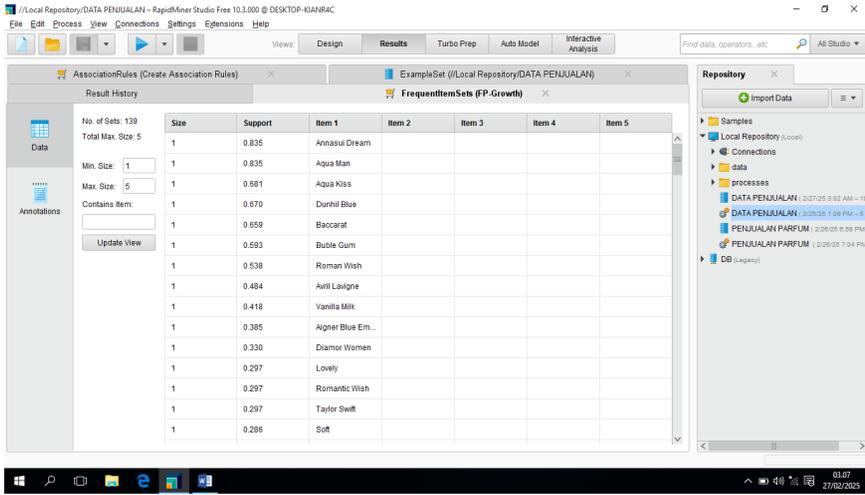
Tahap ketujuh selanjutnya adalah membuat design pada rapidminer. Untuk membuat sebuah design kita cari dahulu data serta jenis operators nya.



Tahap kedelapan setelah dicari disusun design ny, yang pertama dicari terlebih dahulu data penjualan, yang kedua cari dioperator *discretize*, yang ketiga numerical to binominal, yang keempat yaitu *fp-growth*, dan yang terakhir yaitu *create association rules*.



Tahap kesembilan dan dibagian Create Asosiation Rules min confidence nya diubah menjadi 0,55. Dिसuisuaikan dengan support data excel yang sudah dibuat.



Size	Support	Item 1	Item 2	Item 3	Item 4	Item 5
4	0.286	Annasui Dream	Aqua Man	Baccarat	Buble Gum	
4	0.253	Annasui Dream	Aqua Man	Baccarat	Roman Wish	
4	0.253	Annasui Dream	Aqua Man	Baccarat	Vanilla Milk	
4	0.297	Annasui Dream	Aqua Kiss	Dunhill Blue	Baccarat	
4	0.242	Annasui Dream	Aqua Kiss	Dunhill Blue	Buble Gum	
4	0.242	Annasui Dream	Aqua Kiss	Dunhill Blue	Roman Wish	
4	0.286	Annasui Dream	Aqua Kiss	Baccarat	Buble Gum	
4	0.253	Annasui Dream	Aqua Kiss	Baccarat	Roman Wish	
4	0.242	Annasui Dream	Dunhill Blue	Baccarat	Buble Gum	
4	0.242	Annasui Dream	Dunhill Blue	Baccarat	Roman Wish	
4	0.275	Aqua Man	Aqua Kiss	Dunhill Blue	Baccarat	
4	0.253	Aqua Man	Aqua Kiss	Dunhill Blue	Buble Gum	
4	0.253	Aqua Man	Aqua Kiss	Baccarat	Buble Gum	
4	0.242	Aqua Man	Dunhill Blue	Baccarat	Buble Gum	
5	0.242	Annasui Dream	Aqua Man	Aqua Kiss	Dunhill Blue	Baccarat

Tahap terakhir dan ini adalah hasil akhir menggunakan metode *Algoritma Apriori* dengan penggunaan aplikasi rapidminer. Dan sudah muncul hasil support dan itemnya yaitu Annasui Dream 0.855%.

Pembahasan

1. Tingkat minat terhadap pelanggan pada penjualan parfum. Tingkat minat pelanggan terhadap penjualan parfum meningkat seiring dengan berkembangnya perubahan trend dan pengaruh media sosial. Data tersebut bisa dilihat dari data yang memenuhi minimal *support* yaitu 50%, data yang kurang dari 50% akan dihapus. Dari proses tersebut terdapat 5 produk yang masih memenuhi minimal *support*. Setiap kandidat 1 *itemset* yang memenuhi nilai minimal ini akan dianggap sebagai support 1 *itemset*.

Setelah mendapatkan support 1 *itemset*, selanjutnya melakukan tahap kedua yaitu menggabungkan atau mengkombinasikan 2 produk yang memenuhi minimal *support*. Disini 2 *itemset* tersebut saling dipasangkan dan tidak boleh ada pasangan yang sama. *Support 2 itemset* dihasilkan melalui perhitungan kombinasi dari 2 produk yaitu perhitungan dukungannya tetap sama seperti pada langkah sebelumnya saat mencari minimal *support 1 itemset*. Untuk

menghitung dukungan kombinasi dari dua *item*, tahap berikutnya memilih data yang memenuhi minimal *support* yaitu 50%, data yang kurang dari 50% akan dihapus. Dari proses tersebut terdapat 3 produk yang masih memenuhi minimal *support*. Setiap kandidat 2 *itemset* yang memenuhi nilai minimal ini akan dianggap sebagai *support 2 itemset*.

Jika nilai minimal *support* telah didapatkan maka selanjutnya kita menghitung nilai minimal *confidence*. Untuk mendapatkan nilai minimal *confidence* kita dapat menghitung dengan menggunakan jumlah transaksi yang mengandung A dan B dibagi jumlah transaksi yang mengandung A.

Dari nilai *confidence* yang dirangkum dengan penentuan minimum *confidence* sebesar 55% lalu sisihkan yang tidak sesuai dengan syarat minimum *confidence*. Kombinasi hasil nilai *support* dan *confidence* yang tertinggi dapat dilihat dari aturan asosiasi. Akhir dari penentuan aturan asosiasi, menunjukkan bahwa barang yang sering terjual adalah Black Opium dan Victoria Scandal dengan nilai *support* 67,03% serta *confidence* 87,14%, sehingga pemilik mengetahui parfum yang banyak yang diminati.

2. Metode *K-Nearest Neighbor (KNN)* memberikan solusi prediktif pada minat pelanggan

Dengan membersihkan data prediktif dengan mengelompokkan pelanggan yang memiliki preferensi atau pola pembelian yang serupa serta, mendeteksi dan mengeliminasi data yang tidak relevan dan menganalisa data yang digunakan untuk pengelompokan atau perhitungan data dalam jumlah banyak, tetapi memiliki kelemahan jika nilai *k* kecil maka hasil kurang akurat, namun metode *KNN* masih tetap menjadi algoritma yang paling akurat, dengan adanya perhitungan data penjualan yang banyak ini menjadi big data yang harus di analisa dan di dimanfaatkan untuk meningkatkan penjualan dan menambah pendapatan. Jadi jika di simpulkan Algoritma *KNN* adalah cara yang digunakan dalam dunia machine learning untuk proses

klasifikasi dan Tujuan dari algoritma ini untuk mengelompokkan objek tertentu berdasarkan data dan atribut pembelajaran. Oleh karena itu, dengan mengimplementasikan algoritma *K-Nearest Neighbor*, didapatkan analisa penjualan produk dengan lebih mudah dengan mendapatkan informasi dari perhitungan data penjualan dan menarik kesimpulan.

Dengan Algoritma *K-NN* menerapkan perhitungan sebagai nilai dari sampel uji baru. Nilai data dekat atau jauhnya tetangga biasanya dihitung berdasarkan jarak menggunakan perhitungan aritmatika yang sering di sebut dengan rumus *Euclidean distance*

Metode *K-Nearest Neighbors (KNN)* adalah salah satu algoritma pembelajaran mesin yang termasuk dalam kategori *supervised learning*, khususnya dalam klasifikasi dan regresi. KNN bekerja dengan cara mencari sejumlah k tetangga terdekat dari suatu data point untuk menentukan kelas atau nilai target dari data tersebut.

3. Perhitungan hasil manual dan aplikasi dengan Rapidminer penggunaan metode *KNN* dan *Algoritma Apriori*
Berikut ini adalah hasil perhitungan manual metode *KNN* dan *Algoritma Apriori*:

Penentuan jarak *Euclidean* untuk mendapatkan nilai prediksi data uji dan tetangga terdekatnya pada *KNN*, berdasarkan hasil perhitungan jarak, maka didapatkan suatu hasil dengan nilai $k=5$ memiliki keputusan yaitu Minat = 4 dan Tidak Minat= 2.

Dari nilai confidence yang dirangkum dengan penentuan minimum confidence sebesar 55% lalu sisihkan yang tidak sesuai dengan syarat minimum confidence. Kombinasi hasil nilai support dan confidence yang tertinggi dapat dilihat dari aturan asosiasi. Dari hasil tersebut maka kita mendapatkan aturan asosiasi. Akhir dari penentuan aturan asosiasi, menunjukkan barang yang sering terjual adalah Black Opium dan Victoria Scandal dengan nilai *support* 67,03% serta *confidence* 87,14%, sehingga pemilik mengetahui

parfum yang banyak yang diminati dan hasil perhitungan manual metode *Algoritma Apriori* terhadap minat pelanggan pada penjualan parfum yaitu dari data 3 itemset yaitu parfum *black opium*, *taylor swift*, dan *romantic wish* dengan jumlah transaksi 38 dan support 41,75%. Dan diaplikasi yaitu *taylor swift* dan *romantic wish* dengan support 0.297%.

Berikut ini hasil dari aplikasi *Rapidminer* :

Hasil akhir penggunaan *rapidminer* menggunakan metode *KNN*. Ada nama produk yang diminati yaitu *aqua man* dan yang tidak diminati adalah *switzal* dan *pink chiffon*.

Hasil akhir menggunakan metode *Algoritma Apriori* dengan penggunaan aplikasi *rapidminer*. Dan sudah muncul hasil support dan itemnya yaitu *Annasui Dream* 0.855%.

BAB VII



Kesimpulan dan Rekomendasi

Kesimpulan

1. Tingkat minat pelanggan terhadap penjualan parfum cenderung tinggi dan terus berkembang, terutama seiring dengan meningkatnya kesadaran akan perawatan diri dan gaya hidup. Para pelanggan sering kali mencari produk parfum yang sesuai dengan preferensi pribadi mereka, seperti aroma yang unik atau merek yang populer. Dan yang paling banyak diminati yaitu black opium dan victoria scandal.
2. Metode *K-Nearest Neighbor (KNN)* memberikan solusi prediktif pada minat pelanggan dengan cara mengkalasidikasikan atau memprediksi preferensi pelanggan berdasarkan kedekatannya dengan pelanggan lain yang memiliki minat serupa. Untuk memberikan rekomendasi produk yang lebih tepat dan relevan, serta membantu membersihkan data yang tidak konsisten.
3. Hasil perhitungan manual pada metode *KNN* terhadap minat pada penjualan parfum yaitu maka didapatkan suatu hasil dengan nilai

$k=5$ memiliki keputusan yaitu Minat = 4 dan Tidak Minat= 2, sedangkan yang diperhitungan aplikasi yang diminati ada 1 yaitu aqua man dan yang tidak diminati ada 2 yaitu pink chiffon dan switzal. Sedangkan hasil perhitungan manual metode *Algoritma Apriori* terhadap minat pelanggan pada penjualan parfum yaitu dari data 3 itemset yaitu parfum *black opium*, *taylor swift*, dan *romantic wish* dengan jumlah transaksi 38 dan support 41,75%. Dan diaplikasi yaitu *taylor swift* dan *romantic wish* dengan support 0.297%.

Rekomendasi

Optimasi model menyesuaikan parameter dalam metode *KNN*, seperti pemilihan jumlah tetangga terdekat (K), perlu dilakukan secara berkala untuk meningkatkan akurasi prediksi minat pelanggan. Dan penggunaan *Algoritma Apriori* dapat dikembangkan lebih lanjut dengan memasukkan variasi produk atau diskon tertentu, sehingga dapat meningkatkan asosiasi produk yang relevan dan mendorong pembelian produk parfum secara bersamaan.

Daftar Pustaka

- [1] A. Frandhana and H. Ali, “Pengaruh Kualitas Produk dan Harga Pada Minat Beli terhadap Keputusan Pembelian Parfum Refill di Toko Inc . Redible,” vol. 2, no. 1, pp. 36–45, 2024.
- [2] A. Sapitri, N. Elisya, N. M. Mustafa, and M. Badrul, “Penerapan Data Mining Menggunakan Algoritma Apriori Untuk Menentukan Minat Customer Parfume Dari Riwayat Data Penjualan,” vol. 7, no. 1, pp. 11–20, 2022.
- [3] A. J. Wahidin and W. A. Syukrilla, *Data mining*.
- [4] P. Studi, M. Bisnis, F. Ekonomi, D. A. N. Bisnis, U. I. Negeri, and R. I. Lampung, “PENERAPAN MARKET BASKET ANALYSIS MENGGUNAKAN PROSES KNOWLEDGE DISCOVERY IN DATABASE (KDD) SEBAGAI STRATEGI PENJUALAN PRODUK SWALAYAN PENERAPAN MARKET BASKET ANALYSIS MENGGUNAKAN PROSES KNOWLEDGE DISCOVERY IN DATABASE (KDD) SEBAGAI,” 2022.
- [5] A. Pastika and V. Widya, “IMPLEMENTASI DATA MINING SEBAGAI PENENTU PERSEDIAAN PRODUK DENGAN ALGORITMA FP-GROWTH PADA DATA PENJUALAN SINARMART,” vol. 1, no. 2, 2022.
- [6] A. Nugraha, O. Nurdiawan, and G. Dwilestari, “Penerapan Data Mining Metode K-Means Clustering Untuk Analisa Penjualan Pada Toko Yana Sport,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 6, no. 2, pp. 849–855, 2022, doi: 10.36040/jati.v6i2.5755.
- [7] B. Bangun and A. K. Karim, “Pengembalian Data Yang Hilang Pada Dataset Dengan Menggunakan Algoritma K-Nearest Neighbor

- Imputation Data Mining,” *J. Media Inform. Budidarma*, vol. 8, no. 3, p. 1706, 2024, doi: 10.30865/mib.v8i3.8014.
- [8] S. P. Dewi, N. Nurwati, and E. Rahayu, “Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode K-Nearest Neighbor,” *Build. Informatics, Technol. Sci.*, vol. 3, no. 4, pp. 639–648, 2022, doi: 10.47065/bits.v3i4.1408.
- [9] Zulhilmi, Nahar Mardiyantoro, Dimas Prasetyo Utomo, Iman Ahmad Ihsannuddin, and Nulngafan, “Implementasi Data Mining Untuk Menentukan Pola Penjualan Di Armada Computer Menggunakan Algoritma Apriori,” *STORAGE J. Ilm. Tek. dan Ilmu Komput.*, vol. 2, no. 1, pp. 25–31, 2023, doi: 10.55123/storage.v2i1.1749.
- [10] D. Anjeli, S. T. Faulina, and A. Fakhri, “Sistem Informasi Perpustakaan Sekolah Dasar Negeri 49 OKU Menggunakan Embarcadero XE2 Berbasis Client Server,” *J. Inform. dan Komput.*, vol. 13, no. 2, pp. 57–66, 2022.
- [11] T. Hidayat, Y. Handayani, A. Syaifudin, U. S. Sri, P. Produk, and D. Mining, “IMPLEMENTASI ALGORITMA K-NEAREST NEIGHBOR,” vol. 9, no. 2, pp. 118–124, 2023.
- [12] R. Ari, T. Ardani, and R. Andika, “Prediksi penjualan parfum menggunakan algoritma apriori pada queen parfume,” vol. 4, no. 2, pp. 29–35, 2023.
- [13] M. U. Albab and D. Hidayatullah, “Penerapan Algoritma Apriori pada Sistem Informasi Inventori Toko,” vol. 6, pp. 1321–1328, 2022, doi: 10.30865/mib.v6i3.4160.
- [14] F. Firmansyah and O. Nurdiawan, “Penerapan Data Mining Menggunakan Algoritma Frequent Pattern-Growth Untuk Menentukan Pola Pembelian Produk Chemicals,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 1, pp. 547–551, 2023, doi: 10.36040/jati.v7i1.6371.

- [15] C. Cv and M. Mineral, "Rancang Bangun Aplikasi Pengolahan Data Perusahaan Mining Design and Build Data Processing Applications for Mining Consultant CV . Mitra Mineral," vol. 8, no. 2, pp. 149–159, 2020, doi: 10.26418/justin.v8i2.37998.

Tentang Penulis



Evi Yulia Sinta

Lahir di Lohsari 1 Perlabian, pada tanggal 21 Juli 2002. Dalam menempuh pendidikan dimulai dari Sekolah Dasar Mis Tarbiyah Lohsari tamat pada tahun 2015, di MTsN 1 Labusel tamat pada tahun 2018, dan di SMK PGRI 17 Lohsari tamat pada tahun 2021. Lalu melanjutkan pendidikan S-1 ke Perguruan Tinggi Swasta di Universitas Labuhanbatu, Fakultas Sain Dan Teknologi, Prodi Sistem Informasi. Penulis baru pertama kali membuat buku dengan judul “Analisis Minat Pelanggan Pada Penjualan Parfum Menggunakan Metode *K-Nearest Neighbor* Dan *Algoritma Apriori*”. Melalui karya ini mudah-mudahan menjadi berkat untuk mencapai karir penulis dan dapat menjadi amal keluarga kami dan terkhusus kedua orang tua (Bapak Karno dan Ibu Satiyem) semoga selalu mendapat rahmat dari Allah SWT.



Budianto Bangun, S.Sos., M.Kom.

Lahir di Medan pada tanggal 04 April 1970. Menempuh pendidikan (S1) di Universitas Darmawangsa Medan, dan melanjutkan ke (S2) di Universitas Putra Indonesia (YPTK) Padang. Saat ini aktif sebagai tenaga pengajar di Universitas Labuhanbatu Fakultas Sains Dan Teknologi dan aktif di kegiatan sosial lainnya.



Irmayanti, S.Si., M.Pd.

Lahir di Simonis 24 Agustus 1984. Setelah menamatkan SD Negeri 112324 Aek Kota Batu pada tahun 1996, Lalu melanjutkan sekolah Madrasah Tsanawiyah Daar Al-Ulum Asahan Kisaran lulus pada tahun 1999, SMU Al-Azhar Medan lulus pada tahun 2002, S1 Matematika UNIMED lulus pada tahun 2007, dan melanjutkan S2 Pendidikan Matematika lulus pada tahun 2013. Menikah

dengan Beni Yunanta Hasibuan pada tahun 2009, mempunyai dua orang anak satu putri (Ratu Nabila Yunanta Hasibuan) dan satu putra (Rauf Athaya Alby Yunanta Hasibuan). Memulai karir dari guru honor sampai diperguruan tinggi STKIP Labuhanbatu sebagai Dosen tetap. Melalui karya ini muda-mudahan dapat menjadi amal keluarga kami dan terkhusus kedua orang tua (Alm. H. Zulkarnaen Ritonga dan Hj. Normah Dalimunthe) semoga selalu mendapat rahmat dari Allah SWT. Lahir dari keluarga sederhana dengan didorong motivasi tinggi, Alhamdulillah dapat menerbitkan sebuah buku dengan pembahasan “Kalkulus”. Besar harapan kiranya buku ini dapat berguna bagi mahasiswa dan masyarakat umum.



Angga Putra Juledi, S.Kom., M.Kom.

Lahir di Kota Padang pada tanggal 19 Juli 1994. Dalam menempuh Pendidikan dimulai dari Sekolah Dasar 19 Padang tamat tahun 2006, SMPN 3 Padang tamat pada tahun 2009, dan di SMA Pertiwi 2 Padang tamat pada tahun 2012. Lalu melanjutkan ke Pendidikan perguruan tinggi swasta yaitu S1 (Sarjana) di Universitas Putra Indonesia “YPTK” Padang lulus pada tahun 2018 dengan jurusan Sistem Informasi,

Dan melanjutkan Program Pascasarjana (S2) di Universitas Putra Indonesia "YPTK" Padang pada tahun 2019 Program Studi Teknik Informatika. Konsentrasi Sistem Informasi. Saya mengabdikan diri sebagai salah satu Dosen di bidang Ilmu Komputer pada Fakultas Sains Dan Teknologi dengan Program Studi Sistem Informasi di Universitas LabuhanBatu dan menjadi dosen tetap pada tahun 2020 pada kampus tersebut. Saat ini menjadi bagian Struktural di Universitas Labuhanbatu sebagai Kepala Bagian Sumber Daya Manusia periode 2023 s/d 2027. Buku pertama terbit pada 31 Desember 2021 dengan judul Internetworking Dan TCP/IP. Buku kedua terbit pada tanggal 17 Oktober 2023 dengan judul Panduan Belajar HTML, CSS, dan JavaScript untuk Pemula. Buku ke 3 terbit pada 09 Mei 2024 dengan Judul DASAR-DASAR SISTEM BASIS DATA.

