



Ega Kurnia
Angga Putra Juledi, S.Kom., M.Kom.
Ibnu Rasyid Munthe, S.T., M.Kom.
Marnis Nasution, S.Kom., M.Kom.



**PENERAPAN DATA MINING MENGGUNAKAN ALGORITMA
APRIORI DAN METODE K-NEAREST NEIGHBOR DALAM
MENENTUKAN PERSEDIAAN BARANG SEMBAKO
UD. AMORA JAYA**

Ditulis oleh:

EGA KURNIA

ANGGA PUTRA JULEDI, S.KOM., M.KOM.

IBNU RASYID MUNTHE, S.T., M.KOM.

MARNIS NASUTION, 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, Februari 2025

Perancang sampul: Muhammad Ridho Naufal

Penata letak: Muhammad Ridho Naufal

ISBN :

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

©Februari 2025

Prakata

Alhamdulillahirabbil'alamin, saya mengucapkan syukur yang terdalem kepada Allah SWT atas segala rahmat dan karunia-Nya sehingga buku monograf ini dapat diselesaikan dengan baik dan tepat waktu. Shalawat dan salam kepada Rasulullah SAW yang telah berjasa besar dengan membukakan jalan dalam perkembangan ilmu pengetahuan seperti sekarang ini. Buku ini adalah hasil dari penelitian saya di Ud. Amora Jaya, Kotapinang Labuhanbatu Selatan, dengan judul "**Penerapan Data Mining Menggunakan Algoritma Apriori dan Metode K-Nearest Neighbor Dalam Menentukan Persediaan Barang Sembako Ud. Amora Jaya**". Pada kesempatan ini, saya ingin mengucapkan terima kasih yang sedalam-dalamnya kepada semua pihak yang telah memberikan dukungan moral maupun material sehingga buku ini dapat diselesaikan. Saya menyadari bahwa hasil dari buku ini masih belum sempurna. Oleh karena itu, saya sangat mengharapkan saran-saran dan kritikan dari pembaca demi untuk kesempurnaan buku ini. Semoga buku ini dapat menjadi ilmu pengetahuan yang bermanfaat bagi pembaca semua. Aamiin yaa rabba'lalamin.

Rantauprapat, Februari 2024

Ega Kurnia

Pengantar

Perkembangan teknologi informasi telah membawa perubahan signifikan dalam berbagai aspek bisnis, termasuk dalam manajemen persediaan barang. Dalam lingkungan bisnis yang semakin kompetitif, pengelolaan stok yang efektif menjadi faktor krusial untuk menjaga keseimbangan antara permintaan pelanggan dan ketersediaan barang. Salah satu tantangan utama yang dihadapi oleh pelaku usaha adalah bagaimana memastikan stok barang selalu tersedia tanpa mengalami kelebihan atau kekurangan, terutama dalam sektor perdagangan ritel yang memiliki volume transaksi tinggi dan pola pembelian yang dinamis.

Buku ini “Penerapan Data Mining Menggunakan Algoritma Apriori dan Metode K-Nearest Neighbor Dalam Menentukan Persediaan Barang Sembako Ud. Amora Jaya” disusun dengan tujuan untuk memberikan pemahaman yang mendalam mengenai penerapan teknologi data mining dalam mengoptimalkan manajemen persediaan barang. Metode analisis data berbasis algoritma memungkinkan pelaku usaha untuk memperoleh wawasan yang lebih akurat dalam menentukan strategi stok berdasarkan pola pembelian pelanggan. Dalam buku ini, dibahas secara sistematis penerapan algoritma Apriori dalam menemukan keterkaitan antarproduk yang sering dibeli bersamaan serta metode K-Nearest Neighbor (KNN) dalam memprediksi kebutuhan stok berdasarkan data historis.

Buku ini disusun dengan pendekatan akademik, namun tetap mempertahankan keterbacaan bagi pembaca umum yang tertarik untuk memahami konsep dasar serta penerapan praktis dari teknologi data mining dalam dunia bisnis. Selain menjadi referensi bagi mahasiswa dan akademisi di bidang ilmu komputer dan analisis data, buku ini juga diharapkan dapat memberikan wawasan yang bermanfaat bagi pelaku usaha ritel yang ingin mengoptimalkan strategi pengelolaan stok mereka.

Semoga buku ini dapat memberikan manfaat bagi berbagai pihak yang tertarik pada topik analisis data dalam konteks bisnis. Kami mengucapkan terima kasih kepada semua pihak yang telah berkontribusi dalam penyusunan buku ini. Kritik dan saran yang membangun sangat kami harapkan demi penyempurnaan di masa mendatang.



Daftar Isi

Prakata	iii
Pengantar	v
Daftar Isi	v

BAB I

PENDAHULUAN—1

BAB II

DATA MINING—5

Pengertian Data Mining.....	5
Tahapan dalam Proses Data Mining	7

BAB III

ALGORITMA APRIORI—9

Pengertian Algoritma Apriori.....	9
Association Rule	10
Penerapan Algoritma Apriori pada Penjualan Sembako.....	11
Uji Performa Algoritma Apriori	12

BAB IV

METODE K-NEAREST NEIGHBOR (KNN)—23

Pengertian K-Nearest Neighbor.....	23
Uji Performa Metode K-Nearest Neighbor	26

BAB V

RapidMiner sebagai Alat Bantu Data Mining—31

Pengertian RapidMiner.....	31
Implementasi RapidMiner Pada Algoritma Apriori	33
Implementasi RapidMiner Pada Metode K-Nearest Neighbor	37

BAB VI

STUDI KASUS DAN APLIKASI NYATA—43

Studi Kasus di UD. Amora Jaya	43
Kerangka Kerja Kajian	46
Hasil dan Analisis	48

BAB VII

EVALUASI DAN PERBANDINGAN METODE—51

Efektivitas Kelebihan Algoritma Apriori.....	51
Efektivitas Kekurangan Algoritma Apriori	52
Efektivitas Kelebihan Metode K-Nearest Neighbor (KNN).....	52
Efektivitas Kekurangan Metode KNN	53

Daftar Pustaka.....	55
---------------------	----

Tentang Penulis.....	59
----------------------	----



BAB I

PENDAHULUAN

Pengelolaan persediaan merupakan aspek krusial dalam operasional toko grosir sembako. Sebagai usaha yang menyediakan kebutuhan pokok masyarakat, ketersediaan barang yang stabil menjadi faktor utama dalam menjaga kepuasan pelanggan sekaligus memastikan profitabilitas usaha. Namun, tantangan utama dalam pengelolaan persediaan di toko grosir sembako adalah ketidakseimbangan antara permintaan dan stok barang, yang dapat menyebabkan kelebihan atau kekurangan persediaan. Kondisi ini tidak hanya berdampak pada efisiensi bisnis tetapi juga memengaruhi kepercayaan pelanggan dan keberlanjutan usaha.

Salah satu permasalahan umum yang sering terjadi adalah kelebihan stok (*overstocking*). Toko sembako biasanya membeli barang dalam jumlah besar untuk mendapatkan harga yang lebih murah dari pemasok. Namun, tanpa perencanaan berbasis data, kelebihan stok dapat mengakibatkan penumpukan barang yang berujung pada peningkatan biaya penyimpanan dan risiko barang kadaluarsa, terutama untuk produk dengan masa simpan terbatas seperti minyak goreng, beras, dan telur. Kualitas barang yang menurun akibat penyimpanan yang terlalu lama dapat menyebabkan kerugian karena produk menjadi tidak layak jual atau harus dijual dengan harga diskon, yang akhirnya mengurangi margin keuntungan.

Sebaliknya, kekurangan stok (*understocking*) juga menjadi tantangan besar bagi toko grosir sembako. Permintaan konsumen yang bersifat fluktuatif, terutama menjelang hari besar seperti Idul Fitri atau Tahun Baru, sering kali tidak dapat diprediksi dengan baik tanpa sistem yang terorganisir. Kekurangan stok pada produk yang memiliki permintaan tinggi, seperti gula pasir dan tepung terigu, dapat mengakibatkan pelanggan beralih ke kompetitor yang memiliki ketersediaan barang lebih baik. Situasi ini tidak hanya menurunkan tingkat penjualan tetapi juga berpotensi menurunkan loyalitas pelanggan, yang dalam jangka panjang dapat berdampak negatif terhadap citra bisnis.

Permasalahan dalam pengelolaan stok ini semakin kompleks karena banyak toko grosir sembako masih menggunakan pendekatan manual dalam pencatatan transaksi dan inventarisasi. Tanpa sistem yang terkomputerisasi, pemilik toko sering kali mengalami kesulitan dalam melacak pola pembelian pelanggan, memprediksi tren permintaan, serta menentukan strategi stok yang optimal. Akibatnya, keputusan pengadaan barang cenderung berbasis intuisi atau pengalaman pribadi, yang tidak selalu akurat dan dapat menyebabkan inefisiensi operasional.

Salah satu solusi yang dapat diterapkan untuk mengatasi tantangan ini adalah pemanfaatan teknologi data mining, yang memungkinkan analisis mendalam terhadap pola pembelian dan preferensi pelanggan. Dengan menerapkan algoritma seperti *Apriori* dan *K-Nearest Neighbor* (KNN), pemilik toko dapat mengidentifikasi keterkaitan antar produk yang sering dibeli bersamaan serta memprediksi kebutuhan stok berdasarkan tren historis penjualan. Implementasi teknologi ini, terutama melalui perangkat lunak seperti *RapidMiner*, memungkinkan otomatisasi dalam pengambilan keputusan, sehingga pengelolaan stok menjadi lebih efisien, akurat, dan responsif terhadap perubahan permintaan pasar.

Data mining merupakan teknologi yang berperan penting dalam meningkatkan efisiensi pengelolaan persediaan barang, terutama di sektor perdagangan *ritel* seperti toko sembako. Dengan kemampuan untuk mengidentifikasi pola pembelian dan memprediksi kebutuhan stok, data mining memungkinkan pemilik usaha mengelola persediaan dengan lebih akurat,

sehingga dapat menghindari risiko kekurangan atau kelebihan stok. Oleh karena itu, strategi berbasis data menjadi solusi efektif untuk mengatasi tantangan ini.

Salah satu manfaat utama dari penerapan *datamining* dalam optimasi persediaan adalah kemampuannya untuk menemukan pola pembelian pelanggan. Dengan menganalisis data transaksi penjualan dalam periode tertentu, pemilik toko dapat mengidentifikasi produk-produk yang sering dibeli bersamaan. Teknik ini dikenal sebagai *association rule mining*, yang sering diterapkan menggunakan algoritma *Apriori*.

Selain itu, data mining juga membantu dalam prediksi permintaan stok berdasarkan tren penjualan historis. Metode seperti *K-Nearest Neighbor* (KNN) dapat digunakan untuk memperkirakan jumlah stok yang harus disediakan berdasarkan pola belanja sebelumnya. Dengan menerapkan algoritma ini, pemilik toko dapat mengantisipasi lonjakan permintaan pada periode tertentu, misalnya menjelang hari raya atau akhir bulan ketika pembelian sembako cenderung meningkat. Dengan demikian, toko dapat melakukan pemesanan barang kepada pemasok secara lebih tepat waktu dan dalam jumlah yang sesuai, sehingga meminimalkan risiko kehabisan stok tanpa harus menumpuk barang secara berlebihan.

Dalam implementasinya, toko dapat menggunakan *RapidMiner* sebagai perangkat lunak untuk melakukan analisis data secara visual tanpa perlu keahlian pemrograman. *RapidMiner* memungkinkan pemilik usaha untuk menginput data transaksi, menerapkan algoritma *data mining*, dan menghasilkan laporan analisis dalam bentuk grafik atau tabel yang mudah dipahami. Dengan bantuan teknologi ini, pemilik toko yang sebelumnya bergantung pada metode manual dapat mengambil keputusan berbasis data secara lebih akurat dan cepat.

Tujuan utama penulisan buku ini adalah untuk memberikan wawasan yang komprehensif mengenai penerapan algoritma *Apriori*, metode *K-Nearest Neighbor* (KNN), dan perangkat lunak *RapidMiner* dalam menganalisis serta menentukan persediaan barang sembako. Dalam dunia bisnis *ritel*, khususnya toko sembako, pengelolaan persediaan merupakan tantangan yang krusial. Kesalahan dalam menentukan jumlah stok yang

harus disediakan dapat menyebabkan berbagai permasalahan, seperti kelebihan stok (*overstocking*) yang meningkatkan biaya penyimpanan, atau kekurangan stok (*understocking*) yang mengakibatkan kehilangan pelanggan. Oleh karena itu, penerapan metode analisis berbasis data menjadi solusi yang sangat relevan untuk mengoptimalkan pengelolaan persediaan.

Selain itu, buku ini ditulis untuk dapat menjadi referensi yang berguna bagi pemilik toko, akademisi, serta praktisi yang tertarik dalam bidang manajemen persediaan berbasis data mining. Pemahaman yang diperoleh dari buku ini dapat membantu pembaca dalam mengoptimalkan pengelolaan stok barang, meningkatkan efisiensi operasional, serta meningkatkan daya saing usaha mereka dalam industri perdagangan *ritel*. Buku ini juga tidak hanya berperan sebagai panduan akademik tetapi juga sebagai solusi praktis bagi dunia bisnis yang semakin bergantung pada teknologi dan analisis data.



BAB II

DATA MINING

Pengertian Data Mining

Data Mining merupakan salah satu teknik untuk menemukan, mencari, atau menggali informasi atau pengetahuan baru dari sekumpulan data yang sangat besar, dengan integrasi atau penggabungan dengan disiplin ilmu lain seperti statistika, kecerdasan buatan, serta *machine learning*, menjadikan *data mining* sebagai salah satu alat bantu untuk menganalisa data yang kemudian menghasilkan informasi yang berguna. (R. F. Naibaho, S. Z. Harahap, & A. P. Juledi, 2024)

Dalam konteks bisnis *ritel*, *data mining* digunakan untuk menganalisis pola pembelian pelanggan berdasarkan data transaksi yang telah dikumpulkan. Dengan menerapkan algoritma seperti *Apriori* dan *K-Nearest Neighbor* (KNN), pemilik toko dapat memahami preferensi pelanggan, mengidentifikasi kombinasi barang yang sering dibeli bersama, serta memprediksi permintaan barang di masa mendatang. Misalnya, melalui analisis asosiasi menggunakan Algoritma *Apriori*, pemilik toko dapat mengetahui bahwa pelanggan yang membeli beras juga cenderung membeli minyak goreng dan gula. Informasi ini memungkinkan toko untuk menyusun strategi

penempatan produk yang lebih efektif serta merancang promosi berdasarkan kebiasaan belanja pelanggan.

Selain membantu dalam analisis pola pembelian, *data mining* juga memiliki manfaat signifikan dalam manajemen persediaan. Dengan memanfaatkan teknik prediksi, seperti metode *K-Nearest Neighbo* (KNN), toko dapat memperkirakan permintaan barang berdasarkan pola historis dan tren pasar. Misalnya, dengan menganalisis data transaksi selama beberapa bulan terakhir, sistem dapat mengidentifikasi permintaan barang tertentu menjelang hari raya atau akhir bulan. Dengan demikian, toko dapat mengantisipasi peningkatan permintaan tersebut dan menyesuaikan jumlah stok barang, sehingga dapat mengurangi risiko kehabisan barang atau stok yang berlebihan.

Adapun beberapa manfaat dari *data mining*, khususnya dalam pengelolaan persediaan barang :

1. Membantu pemilik toko membuat keputusan yang lebih baik.
2. Membantu pemilik toko menemukan pola baru yang belum pernah terprediksi.
3. Membantu pemilik toko menganalisis data penjualan sebelumnya.
4. Membantu pemilik toko mengetahui tren.
5. Membantu pemilik toko menyusun strategi peningkatan penjualan.

Data Mining terbagi atas beberapa kelompok berdasarkan tugas yang dilakukannya yaitu : (Y. Ananda, B. Sembiring, & E. A. Sembiring, 2023)

1. *Description* (Deskripsi) penulis selalu berusaha menciptakan metode untuk menggambarkan tren tidak terlihat pada suatu data.
2. *Estimation* (Estimasi) hampir sama dengan klasifikasi, kecuali variabel objektifnya target lebih kearah numerik.
3. Prediksi hampir sama dengan klasifikasi. Namun hasilnya adalah sesuatu yang mungkin akan terjadi di masa mendatang.
4. Klasifikasi, pada klasifikasi variabel bertujuan kategori.
5. *Clustering* (Pengklusteran) merupakan pengelompokan pemilihan *record*, observasi pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan.

6. Asosiasi mengenali alur sebuah kejadian yang berjalan pada saat tertentu.

Tahapan dalam Proses Data Mining

Data mining merupakan bagian dari proses yang lebih luas dalam eksplorasi data yang disebut Knowledge Discovery in Database (KDD). Knowledge Discovery in Database (KDD) adalah tahapan sistematis dalam mengekstraksi pengetahuan yang berguna dari kumpulan seleksi data, pemrosesan, transformasi, data mining, dan evaluasi.

Knowledge Discovery in Database (KDD) merupakan salah satu tahapan atau kegiatan yang meliputi kegiatan pengumpulan data dan diikuti pemakaian data historis untuk menemukan sebuah pola yang ada dalam hubungan antar data dalam sebuah *database* yang berukuran besar. (K. Handoko & L. S. Lesmana 2018) *Knowledge Discovery in Database* (KDD) adalah salah satu metode yang bisa digunakan dalam melakukan data mining. Dalam proses *data mining* proses *Knowledge Discovery in Database* (KDD) memiliki beberapa tahapan yaitu sebagai berikut: (Yuli Mardi, 2023)

1. Seleksi Data (*Data Selection*)

Pemilihan data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam *Knowledge Discovery in Database* (KDD) dimulai. Data hasil seleksi yang akan digunakan untuk proses *data mining*, disimpan dalam suatu berkas terpisah dari basis data operasional.

2. Pemrosesan (*Preprocessing*)

Sebelum proses *data mining* dilaksanakan, perlu dilakukan proses *cleaning* pada data yang menjadi fokus *Knowledge Discovery in Database* (KDD). Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang konsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak.

3. Transformasi (*Transformation*)

Coding adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses *data mining*. Proses coding dalam *Knowledge Discovery in Database* (KDD) yaitu proses kreatif

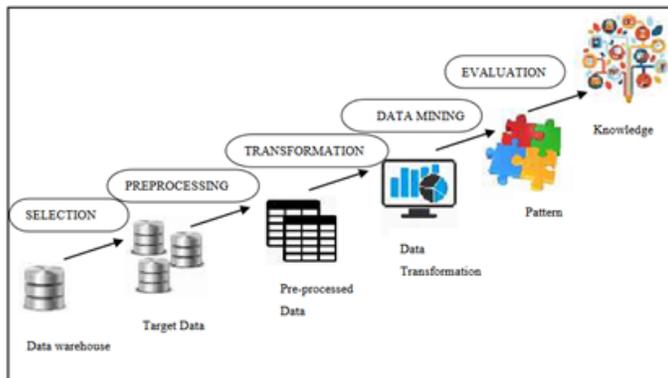
dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

4. Data Mining

Data mining adalah proses mencari pola atau informasi menarik dalam dengan menggunakan teknik atau metode tertentu. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses *Knowledge Discovery in Database* (KDD) secara keseluruhan.

1. Evaluasi (*Evaluation*)

Tahap evaluasi merupakan bagian dari proses *Knowledge Discovery in Database* (KDD) yang disebut *interpretation*. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.



Gambar 2. 1. Knowledge Discovery in Database



BAB III

ALGORITMA APRIORI

Pengertian Algoritma Apriori

Algoritma *Apriori* merupakan salah satu teknik utama dalam data mining yang digunakan untuk menemukan pola asosiasi antar produk dalam kumpulan data transaksi. Algoritma ini dikembangkan oleh Rakesh Agrawal dan Ramakrishnan Srikant pada tahun 1994 sebagai metode untuk mengekstrak aturan asosiasi dari data yang besar dan kompleks. Dalam konteks bisnis ritel dan manajemen persediaan, Algoritma *Apriori* sangat berguna untuk menganalisis kebiasaan belanja pelanggan dan mengidentifikasi hubungan antar produk yang sering dibeli secara bersamaan.

Konsep dasar dari Algoritma *Apriori* berpusat pada dua parameter utama, yaitu minimum *support* dan minimum *confidence*. Minimum *support* mengukur seberapa sering suatu kombinasi barang muncul dalam keseluruhan transaksi, sedangkan minimum *confidence* menunjukkan kemungkinan bahwa pelanggan yang membeli suatu produk juga akan membeli produk lainnya. Misalnya, jika data transaksi menunjukkan bahwa pelanggan yang membeli beras juga sering membeli minyak goreng dengan nilai *confidence* sebesar 80%, maka aturan asosiasi dapat dibentuk sebagai:

“Jika pelanggan membeli beras, maka kemungkinan 80% mereka juga akan membeli minyak goreng.”

Proses kerja Algoritma *Apriori* terdiri dari beberapa tahap utama. Pertama, algoritma mengidentifikasi *frequent itemsets*, yaitu kombinasi barang yang sering muncul dalam transaksi berdasarkan nilai support yang telah ditentukan sebelumnya. Setelah itu, algoritma membentuk aturan-aturan asosiasi dengan menghitung nilai confidence dari kombinasi produk yang telah ditemukan. Tahap terakhir adalah evaluasi dan penyaringan aturan asosiasi berdasarkan parameter tambahan seperti rasio angkat, yang mengukur seberapa besar hubungan antar produk dibandingkan dengan pembelian acak.

Keunggulan utama dari Algoritma *Apriori* adalah kemampuannya dalam mengolah data dalam jumlah besar secara efisien. Namun, algoritma ini juga memiliki tantangan, terutama dalam menangani dataset yang sangat besar dengan jumlah kombinasi *item* yang tinggi. Oleh karena itu, optimalisasi dengan teknik seperti penghitungan item berbasis hash dan partisi sering digunakan untuk meningkatkan kinerja algoritma dalam menganalisis data transaksi skala besar.

Association Rule

Asosiasi adalah teknik untuk mengamati Asosiasi antara pasangan *item* yang diberikan. (Indah & I. Ali, 2024) Metode *association rule* merupakan metode untuk mencari pola atau hubungan yang menarik pada suatu data, yang dapat membantu dalam menemukan pola yang sering terjadi, asosiasi, hubungan, dan korelasi dari kumpulan data pada database transaksional dikenal juga dengan istilah analisis asosiasi. Aturan Asosiasi merupakan teknik *data mining* yang digunakan untuk mengidentifikasi pola dan hubungan antar *item* dalam dataset. Tujuan dari aturan asosiasi adalah untuk mengidentifikasi aturan yang mengatur hubungan antar *item*.

Dalam menentukan nilai *minimum support* sebuah *item* dapat menggunakan rumus persamaan seperti dibawah ini :

$$1. \quad \text{Support (A)} = \frac{\text{Jumlah Transaksi Mengandung A}}{\text{Total Transaksi}} \times 100\%$$

Persamaan (1) merupakan rumus menentukan *support*.

$$2. \quad \text{Support (A, \cap B)} = \frac{\sum \text{Jumlah Transaksi Mengandung A dan B}}{\text{Total Transaksi}} \times 100\%$$

Persamaan (2) merupakan rumus menentukan *support 2 item*.

Sedangkan untuk menentukan nilai *minimum confidence* sebuah *item* dapat menggunakan rumus persamaan seperti dibawah ini :

$$1. \quad \text{Confidence P (B|A)} = \frac{\sum \text{Jumlah Transaksi Mengandung A dan B}}{\text{Jumlah Transaksi Mengandung A}} \times 100\%$$

Persamaan (3) merupakan rumus mencari *confidence*.

Penerapan Algoritma Apriori pada Penjualan Sembako

Sebagai contoh studi kasus, sebuah toko sembako Ud. Amora Jaya ingin menganalisis transaksi pelanggan selama enam bulan terakhir untuk mengidentifikasi barang-barang yang sering dibeli secara bersamaan. Data yang digunakan mencakup daftar transaksi yang terdiri dari berbagai produk sembako, seperti beras, minyak goreng, gula, tepung terigu, telur, dan kebutuhan pokok lainnya. Proses analisis dimulai dengan menetapkan nilai ambang batas *support* dan *Confidence*, yang akan digunakan untuk menyaring aturan asosiasi yang paling relevan.

Penerapan Algoritma *Apriori* juga dapat membantu dalam manajemen persediaan barang. Dengan mengetahui pola pembelian pelanggan, toko dapat mencantumkan jumlah stok yang dibutuhkan untuk setiap produk dan menghindari risiko kekurangan atau kelebihan stok. Sebagai contoh, jika pola transaksi menunjukkan bahwa tepung terigu dan gula sering dibeli bersama menjelang hari raya, toko dapat meningkatkan stok kedua barang ini pada periode yang sama di tahun berikutnya untuk memenuhi permintaan pelanggan.

Uji Performa Algoritma Apriori

Tabel 3. 1. DataPenjualan (Mei 2024 – Oktober 2024)

Kode	Tanggal	Nama Item	Jumlah Item Terjual
1	Rabu, 01 Mei 2024	Gula Putih, Telur, Mie Instan, Sabun Batang, Kacang Hijau, Gula Merah, Pepsodent, Teh, Roti, Downy, Deterjen Liqiud, Tepung Beras, Royco, Pks, Vape, Sarden, Garam, Mie Putih.	18
2	Kamis, 02 Mei 2024	Beras, Gula Putih, Minyak Goreng, Gula Merah, Kacang Hijau, Susu Kaleng, Sunlight, Royco, Garam, Shampo, Kopi Bintang, Mie Lidi, Sikat Gigi, Saos Dena.	14
3	Jumat, 03 Mei 2024	Telur, Mie Putih, Mie Lidi, Bumbu Racik, Tepung Terigu, Kacang Tanah, Kecap Bango, Royco, Pks, Rapika, Wipol, Tepung Beras, Sarden, Gula Putih, Pepsodent, Roti, Teh, Susu Kaleng, Garam, Sabun Batang.	20
4	Sabtu, 04 Mei 2024	Minyak Goreng, Gula Merah, Teh, Downy, Deterjen Liqiud, Tepung Beras, Royco, Pks, Vape, Sarden, Garam, Pepsodent, Roti, Susu Kaleng, Mie Putih.	15

Kode	Tanggal	Nama Item	Jumlah Item Terjual
5	Minggu, 05 Mei 2024	Tepung Tapioka, Sabun Lifeboy, Gula Putih, Telur, Mie Instan, Sabun Batang, Minyak Goreng, Gula Merah, Kopi Bintang, Mie Putih, Tepung Beras, Royco, Pks, Vape, Sarden, Garam, Mie Lidi.	17
6	Senin, 06 Mei 2024	Susu Kaleng, Mie Putih, Tepung Terigu, Kacang Tanah, Kecap Bango, Tepung Beras, Royco, Pks, Vape, Gula Putih, Telur, Mie Instan, Sunlight, Garam, Pepsodent, Roti, Tepung Tapioka, Sabun Lifeboy, Sikat Gigi, Saos Dena, Beras, Downy, Wipol.	23
7	Selasa, 07 Mei 2024	Sabun Batang, Gula Putih, Telur, Beras, kopi Bintang, Sunlight, Kecap Bango, Deterjen Liquid, Sabun Lifeboy, Garam, Teh, Pks, Tepung Tapioka, Wipol, Royco, Rapika.	16
8	Rabu, 08 Mei 2024	Mie Instan, Minyak Goreng, Kecap Bango, Rapika, Teh, Gula Putih, Sarden, Tepung Beras, Royco, Pks, Vape, Gula Putih, Telur, Sunlight, Garam, Pepsodent, Roti, Tepung Tapioka, Sabun Lifeboy, Sikat Gigi, Saos Dena, Beras, Downy, Wipol, Kopi Bintang.	25

Kode	Tanggal	Nama Item	Jumlah Item Terjual
9	Kamis, 09 Mei 2024	Beras, Gula Putih, Minyak Goreng, Telur, Mie Instan, Tepung Terigu, Tepung Beras, Tepung Tapioka, Gula Merah, Kacang Hijau, Kacang Tanah, Garam, Teh, Kopi Bintang, Pepsodent, Roti, Sunlight, Susu Kaleng, Kecap Bango, Royco, Sabun Lifeboy, Sarden, Deterjen Liquid, Vape, Sabun Batang, Downy, Wipol, Sikat Gigi, Shampo, Saos Dena, Pks, Rapika, Bumbu Racik, Mie Putih, Mie Lidi.	35
10	Jumat, 10 Mei 2024	Rapika, Bumbu Racik, Mie Lidi, Gula Merah, Kacang Hijau, Kacang Tanah, Garam, Teh, Kopi Bintang, Mie Instan, Tepung Terigu, Tepung Beras, Tepung Tapioka, Pks, Royco, Sabun Batang, Vape, Wipol, Telur.	19
.....



Kode	Tanggal	Nama Item	Jumlah Item Terjual
184	Kamis, 31 Oktober 2024	Mie Lidi, Gula Putih, Minyak Goreng, Telur, Mie Instan, Tepung Terigu, Tepung Beras, Tepung Tapioka, Gula Merah, Kacang Hijau, Kacang Tanah, Garam, Teh, Kopi Bintang, Pepsodent, Roti, Sunlight, Susu Kaleng, Kecap Bango, Royco, Sabun Lifeboy, Sarden, Deterjen Liquid, Vape, Sabun Batang, Downy, Wipol, Sikat Gigi, Shampo, Saos Dena, Pks, Rapika, Bumbu Racik, Mie Putih, Beras.	35

Untuk menghasilkan aturan asosiasi yang relevan, tahap selanjutnya pada tahap awal perhitungan adalah menentukan nilai minimum *support* dan minimum *confidence*. Dalam penelitian ini, nilai minimum *support* yang ditetapkan adalah 85% dan nilai minimum *confidence* adalah 80%. Setelah menetapkan nilai minimum *support* dan *confidence*, langkah berikutnya adalah mencari *support* 1 itemset dan menghitung nilai *support* nya, dengan menggunakan rumus:



Tabel 3. 2. Data Support 1 Itemset

Item	Transaksi	Support 1 Itemset
Beras	138	75%
Gula Putih	118	64,13%
Minyak Goreng	107	58,15%
Telur	147	79,89%
Mie Instan	141	76,63%
Tepung Terigu	103	55,97%

Item	Transaksi	Support 1 Itemset
Tepung Beras	126	68,47%
Tepung Tapioka	100	54,34%
Gula Merah	108	58,69%
Kacang Hijau	88	47,82%
Kacang Tanah	98	53,26%
Garam	143	77,71%
Teh	130	70,65%
Kopi Bintang	135	73,36%
Pepsodent	168	91,30%
Roti	156	84,78%
Sunlight	158	85,86%
Susu Kaleng	133	72,28%
Kecap Bango	136	73,91%
Royco	171	92,93%
Sabun Lifeboy	158	85,86%
Sarden	156	84,78%
Deterjen Liquid	140	76,08%
Vape	163	88,58%
Sabun Batang	148	80,43%
Downy	156	84,78%
Wipol	153	83,15%
Sikat Gigi	147	79,89%
Shampo	122	66,30%
Saos Dena	139	75,54%
Pks	159	86,41%
Rapika	108	58,69%
Bumbu Racik	107	58,15%
Mie Putih	111	60,32%
Mie Lidi	97	52,71%

Pada tahap berikutnya memlih data yang memenuhi minimum *support* yaitu 85%, data yang kurang dari 85% akan dihapus dari tabel. Dari proses diatas terdapat 6 produk yang masih memenuhi minimum *support*. Setiap

kandidat 1 *itemset* yang memenuhi nilai minimum ini akan dianggap sebagai support 1 *itemset*. Data hasil support 1 *itemset* ditunjukkan pada tabel dibawah ini:

Tabel 3. 3. Data Hasil Support 1 Itemset

Item	Transaksi	Support 1 Itemset
Pepsodent	168	91,30%
Sunlight	158	85,86%
Royco	171	92,93%
Sabun Lifeboy	158	85,86%
Vape	163	88,58%
Pks	159	86,41%

Setelah mendapatkan *support 1 itemset*, langkah berikutnya adalah melakukan tahap kedua yaitu menggabungkan atau mengkombinasikan 2 produk yang memenuhi minimum *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:

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

Tabel 3. 4. Data Support 2 Itemset

No	Item	Frekuensi Kemunculan	Transaksi	Support 2 Itemset
1	Pepsodent, Sunlight	149	$(149/184) \times 100$	81.97%
2	Pepsodent, Royco	157	$(157/184) \times 100$	85.32%
3	Pepsodent, Sabun Lifeboy	147	$(147/184) \times 100$	79.89%

No	Item	Frekuensi Kemunculan	Transaksi	Support 2 Itemset
4	Pepsodent, Vape	154	$(154/184)*100$	83.69%
5	Pepsodent, Pks	147	$(147/184)*100$	79.89%
6	Sunlight, Royco	147	$(149/184)*100$	79.89%
7	Sunlight, Sabun Lifeboy	149	$(149/184)*100$	80.97%
8	Sunlight, Vape	140	$(140/184)*100$	76.08%
9	Sunlight, Pks	135	$(135/184)*100$	73.36%
10	Royco, Sabun Lifeboy	152	$(152/184)*100$	82.60%
11	Royco, Vape	155	$(155/184)*100$	84.23%
12	Royco, Pks	151	$(151/184)*100$	82.06%
13	Sabun Lifeboy, Vape	145	$(145/184)*100$	78.80%
14	Sabun Lifeboy, Pks	137	$(137/184)*100$	74.45%
15	Vape, Pks	142	$(142/184)*100$	77.17%

Seperti langkah sebelumnya, tahap berikutnya memilih data yang memenuhi minimum *support* yaitu 80%, data yang kurang dari 80% akan dihapus dari tabel. Dari proses diatas terdapat 7 produk yang masih memenuhi minimum *support*. Setiap kandidat 2 *itemset* yang memenuhi nilai minimal ini akan dianggap sebagai *support 2 itemset*. Data hasil *support 2 itemset* ditunjukkan pada tabel dibawah ini:

Tabel 3. 5. Data Hasil Support 2 Itemset

No	Item	Transaksi	Support 2 Itemset
1	Pepsodent, Sunlight	149	80,97%
2	Pepsodent, Royco	157	85,32%

No	Item	Transaksi	Support 2 Itemset
3	Pepsodent, Vape	154	83,69%
4	Sunlight, Sabun Lifeboy	149	80,97%
5	Royco, Sabun Lifeboy	152	82,60%
6	Royco, Vape	155	84,23%
7	Royco, Pks	151	82,06%

Setelah mendapatkan nilai minimum 2 *itemset*, langkah berikutnya adalah menggabungkan atau mengkombinasikan 3 produk yang memenuhi minimum *support*. Disini 3 *itemset* tersebut saling dipasangkan dan tidak boleh ada pasangan yang sama. *Support 3 itemset* dihasilkan melalui perhitungan kombinasi dari 3 produk yaitu perhitungan dukungannya tetap sama seperti pada langkah sebelumnya saat mencari minimum *support 2 itemset*. Untuk menghitung dukungan kombinasi dari 3 *itemset*, dengan menggunakan rumus:

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

Tabel 3. 6. Data Hasil Support 3 Itemset

No	Item	Frekuensi Kemunculan	Transaksi	Support 3 Itemset
1	Pepsodent, Sunlight, Royco	138	(138/184)*100	75%
2	Sunlight, Royco, Sabun Lifeboy	144	(144/184)*100	78.26%
3	Royco, Sabun Lifeboy, Vape	140	(140/184)*100	76.08%
4	Sabun Lifeboy, Vape, Pks	125	(125/184)*100	67.93%
5	Vape, Pks, Pepsodent	136	(136/184)*100	73.91%

Jika nilai minimum *support* telah didapatkan maka selanjutnya kita menghitung nilai minimum *confidence*. Untuk mendapatkan nilai minimum *confidence* kita dapat menghitung dengan menggunakan jumlah transaksi yang mengandung A dan B dibagi jumlah transaksi yang mengandung A. Contoh rumusnya sebagai berikut:

Tabel 3. 7. Data Nilai Confidence

No	Item	Frekuensi A	Frekuensi A∩B	Transaksi	Confiden
1	Pepsodent, Sunlight	168	149	$(149/168)*100$	88.69%
2	Sunlight, Pepsodent	158	149	$(149/158)*100$	94.30%
3	Pepsodent, Royco	168	157	$(157/168)*100$	93.45%
4	Royco, Pepsodent	171	157	$(157/171)*100$	91.81%
5	Pepsodent, Vape	168	154	$(154/168)*100$	91.66%
6	Vape, Pepsodent	163	154	$(154/163)*100$	94.47%
7	Sunlight, Sabun Lifeboy	158	149	$(149/158)*100$	94.30%
8	Sabun Lifeboy, Sunlight	158	149	$(149/158)*100$	94.30%
9	Royco, Sabun Lifeboy	171	152	$(152/171)*100$	88.88%
10	Sabun Lifeboy, Royco	158	152	$(152/158)*100$	96.20%
11	Royco, Vape	171	155	$(155/171)*100$	90.64%
12	Vape, Royco	163	155	$(155/163)*100$	95.09%
13	Royco, Pks	171	151	$(151/171)*100$	88.30%

No	Item	Frekuensi A	Frekuensi $A \cap B$	Transaksi	Confiden
14	Pks, Royco	159	151	$(151/159) \times 100$	94.96%

Akhir dari penentuan aturan asosiasi diatas, menunjukkan barang yang sering terjual adalah Sabun Lifeboy dan Royco dengan nilai support 82,60% serta confidence 96,20%, sehingga pemilik mengetahui barang yang banyak terjual.





BAB IV

METODE K-NEAREST NEIGHBOR (KNN)

Pengertian K-Nearest Neighbor

K-Nearest Neighbor (KNN) adalah suatu metode yang menggunakan algoritma *supervised* dimana hasil dari *query instance* yang baru diklasifikasikan berdasarkan mayoritas dari *label class* pada K-NN. Metode *K-Nearest Neighbor* (KNN) merupakan sebuah metode untuk melakukan klasifikasi terhadap obyek baru berdasarkan (K) tetangga terdekatnya. KNN termasuk algoritma *supervised learning*, yang mana hasil dari *query instance* baru, diklasifikasikan berdasarkan mayoritas dari kategori pada KNN. Kelas yang paling banyak muncul, yang akan menjadi kelas hasil klasifikasi.

Tujuan dari algoritma K-NN adalah mengklasifikasikan objek baru berdasarkan atribut dan training data. Sehingga penelitian ini dengan penerapan algoritma *K-Nearest Neighbor* dapat mempermudah UD Amora Jaya pada penjualan produk sembako mengambil objek baru berdasarkan data yang letaknya terdekat dari data baru tersebut. Algoritma K-NN menggunakan klasifikasi ketetanggaan sebagai nilai prediksi dari sampel uji yang baru. (N. E. Pratiwi, L. Suryadi, F. Ardhy & P. Riswanto, 2022)

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* : (S. A. Pradana & M. A. Adiguna, 2024)

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

Keterangan :

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

y_{xi} = 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 : (A. A. Putri, 2021)

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.

Berikut adalah rumus umum untuk uji performa metode KNN, terutama dalam konteks klasifikasi: (N. W. Mardiyah, N. Rahaningsih, & I. Ali, 2024)

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

Untuk mengukur sejauh mana model dapat mengklasifikasikan data dengan benar.

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

Untuk mengukur sejauh mana kelas yang diklasifikasikan sebagai positif oleh model adalah benar-benar positif.

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

Mengukur sejauh mana model dapat menangkap atau mendeteksi seluruh kasus positif yang harusnya terdeteksi.

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

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 4. 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.

Uji Performa Metode K-Nearest Neighbor

Tabel 4. 2. Data Uji

No	Nama Produk	Kuantitas Produk	Kuantitas Terjual	Kategori
1	Beras	150	138	Laris
2	Gula Putih	135	118	Laris
3	Minyak Goreng	120	107	Laris
4	Telur	155	147	Laris
5	Mie Instan	200	141	Tidak Laris
6	Tepung Terigu	160	103	Tidak Laris
7	Tepung Beras	130	126	Laris
8	Tepung Tapioka	125	100	Laris
9	Gula Merah	170	108	Tidak Laris
10	Kacang Hijau	122	108	Laris
11	Kacang Tanah	100	98	Laris
12	Garam	165	143	Laris
13	Teh	140	130	Laris
14	Kopi Bintang	145	135	Laris
15	Pepsodent	210	168	Laris
16	Roti	175	156	Laris
17	Sunlight	185	158	Laris
18	Susu Kaleng	152	133	Laris
19	Kecap Bango	147	136	Laris
20	Royco	205	171	Laris
21	Sabun Lifeboy	177	158	Laris
22	Sarden	166	156	Laris
23	Deterjen Liquid	186	140	Laris
24	Vape	172	163	Laris
25	Sabun Batang	157	148	Laris
26	Downy	190	156	Tidak Laris
27	Wipol	183	153	Laris
28	Sikat Gigi	173	147	Laris
29	Shampo	134	122	Laris
30	Saos Dena	152	139	Laris

No	Nama Produk	Kuantitas Produk	Kuantitas Terjual	Kategori
31	Pks	172	159	Laris
32	Rapika	133	108	Laris
33	Bumbu Racik	128	107	Laris
34	Mie Putih	143	111	Laris
35	Mie Lidi	162	97	Tidak Laris

Berdasarkan data uji tersebut, langkah selanjutnya yaitu melakukan normalisasi data dengan menggunakan normalisasi min max. Berikut rumus pada normalisasi min max yaitu:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

Keterangan:

x' = nilai skala baru

x = nilai data asli sebelum dilakukan normalisasi

Min = nilai terendah dari data

Max = nilai tertinggi dari data

Berikut penyelesaian perhitungannya dalam bentuk tabel:

Nilai Min Kuantitas Produk : 100

Nilai Max Kuantitas Produk : 210

Nilai Min Kuantitas Terjual : 98

Nilai Max Kuantitas Terjual : 171

Tabel 4. 3. Data Normalisasi Min Max

Min Max Kuantitas Produk	Min Max Kuantitas Terjual
0,4545	0,5479
0,3181	0,2739
0,1818	0,1232
0,5	0,6712
0,9090	0,5890
0,5454	0,0684
0,2727	0,3835
0,2272	0,0273

Min Max Kuantitas Produk	Min Max Kuantitas Terjual
0,6363	0,1369
0,2	0,1369
0	0
0,5909	0,6164
0,3636	0,4383
0,4090	0,5068
1	0,9589
0,6818	0,7945
0,7727	0,8219
0,4727	0,4794
0,4272	0,5205
0,9545	1
0,7	0,8219
0,6	0,7945
0,7818	0,5753
0,6545	0,8904
0,5181	0,6849
0,8181	0,7945
0,7545	0,7534
0,6636	0,6712
0,3090	0,3287
0,4727	0,5616
0,6545	0,8356
0,3	0,1369
0,2545	0,1232
0,3909	0,1780
0,5636	-0,0136

Tahap selanjutnya yaitu menentukan Jarak *Euclidean Distance*, *Ascending* (ranking), dan parameter K = jumlah tetangga terdekat. Pada penelitian ini nilai K yang digunakan adalah 5. Menghitung jarak antara data testing (uji) dengan semua data training. (S. P. Dewi & E. Rahayu, 2022)

Tabel 4. 4. Hasil Nilai K

Jarak Euclidean	Ranking	K=1	K=3	K=5
42,72001873	14			
34,20526275	5			Laris
43,17406629	15			
50,48762225	20			
58,13776741	23			
6,32455532	1	Tidak Laris	Tidak Laris	Tidak Laris
43,18564576	17			
37,12142239	7			
13,60147051	2		Tidak Laris	Tidak Laris
41,48493703	11			
62,00806399	27			
46,09772229	18			
39,66106403	10			
41,62931659	12			
85,70297544	34			
60,41522987	26			
65,19202405	30			
37,36308338	8			
41,78516483	13			
85,58621384	33			
62,8171951	29			
59,13543777	24			
49,24428901	19			
66,75327707	32			
51,2445119	22			
65,30696747	31			
59,80802622	25			
51,19570294	21			
37,53664876	9			

Jarak Euclidean	Ranking	K=1	K=3	K=5
43,17406629	15			
62,80127387	28			
31,01612484	4			Laris
35,44009029	6			
23,60084744	3		Laris	Laris

Berdasarkan hasil perhitungan jarak diatas, maka didapatkan suatu hasil dengan nilai $K = 5$ memiliki keputusan yaitu Terlaris = 4 dan Tidak Terlaris = 5



BAB V

RapidMiner sebagai Alat Bantu Data Mining

Pengertian RapidMiner

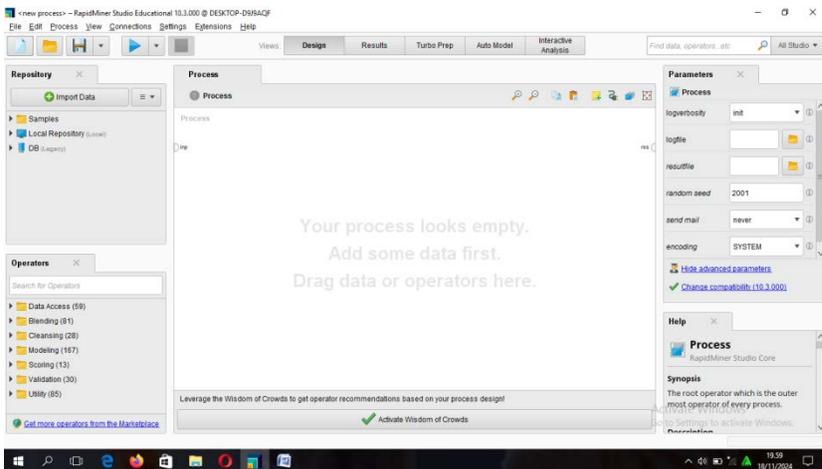
RapidMiner adalah *platform* perangkat lunak yang kuat untuk ilmu data dan pembelajaran mesin. Ini menyediakan beragam alat untuk persiapan data, pemodelan, evaluasi, dan implementasi. *RapidMiner* dirancang untuk mudah digunakan dan memungkinkan pengguna untuk dengan mudah membangun dan menguji berbagai model, bahkan tanpa pengalaman pemrograman. (M. R. Nahjan & N. Heryana) *RapidMiner* mempermudah pengguna untuk menggabungkan Algoritma *Apriori* dan metode *K-Nearest Neighbor* (KNN) dalam satu analisis yang terintegrasi. Proses *preprocessing* data yang efisien dapat dilakukan untuk membersihkan, mengubah format, dan menyiapkan data untuk analisis lebih lanjut. Algoritma *Apriori* dapat digunakan untuk mengatasi munculnya *frequentitem* atau *itemset* dalam pencarian nilai *support* dan nilai *confidence* pada database yang cukup besar.

RapidMiner merupakan *platform* analitik data yang dirancang untuk memfasilitasi proses analisis data secara efisien tanpa memerlukan kemampuan pemrograman yang mendalam. Dengan antarmuka yang intuitif dan

fitur *drag-and-drop*, *RapidMiner* memungkinkan pengguna dari berbagai latar belakang untuk merancang, menjalankan, dan mengelola alur kerja analitik dengan mudah. *Platform* ini mendukung berbagai tugas analisis data, mulai dari persiapan data hingga penerapan model prediktif, sehingga menjadi alat yang sangat berguna bagi pemula maupun profesional di bidang *data science*.

Salah satu keunggulan utama *RapidMiner* adalah kemampuannya untuk mengintegrasikan berbagai sumber data dan sistem *IT* yang beragam. Hal ini menjadikannya solusi yang fleksibel dan berskalabel untuk berbagai kebutuhan analisis data dalam bisnis dan industri lainnya. Selain itu, *RapidMiner* menyediakan sekitar 500 operator *data mining*, termasuk operator *input*, *output*, *preprocessing* data, dan visualisasi, yang memungkinkan pengguna untuk melakukan analisis data yang kompleks tanpa perlu menulis kode secara manual.

Dalam penerapannya, *RapidMiner* memungkinkan pengguna untuk mengakses, mengolah, dan menganalisis data dengan antarmuka visual yang sederhana dan inovatif. *Platform* ini juga menyediakan alat untuk menyimpan data, pembuatan model, evaluasi, dan visualisasi, serta kemampuan integrasi dengan berbagai sumber data. Dengan demikian, *RapidMiner* cocok untuk para pemula hingga ahli di bidang ilmu data karena kemudahannya dalam mengelola alur kerja analitik yang kompleks.

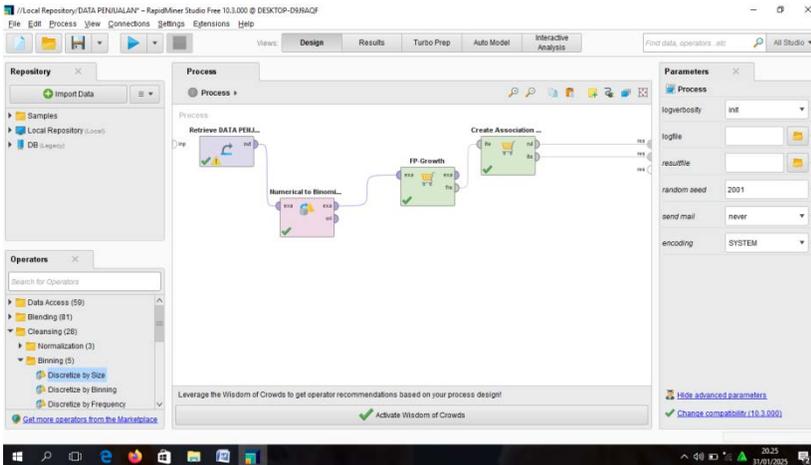


Gambar 5. 1. Aplikasi Rapid Miner

Implementasi RapidMiner Pada Algoritma Apriori

Implementasikan algoritma *apriori* pada data tersebut dengan menggunakan aplikasi *rapidminer*. Langkah pertama dalam analisis adalah memasukkan data transaksi ke dalam RapidMiner. Data transaksi yang diperoleh dari toko UD. Amora Jaya disusun ke dalam bentuk format *Excel*, yang kemudian diunggah atau diimport ke RapidMiner. Setelah itu, dilakukan proses *preprocessing*, yang mencakup pembersihan data, normalisasi nilai numerik, serta transformasi data agar dapat digunakan dalam analisis algoritma Apriori.

Setelah semua operator yang kita butuhkan sudah berada dihalaman main proses selanjutnya hubungkan Dataset Data Penjualan Ud. Amora Jaya.xlsx, output hubungkan ke *exa* ke *exaNumerical to Binominal*, dari *Numerical to Binominal* ke *exa Fp-Growth* dan *Frequent Fp-Growth* ke *Itemset Create Association Rules*, dan *rule* dan *itemset Create Association Rules* ke *Result* yang dapat dilihat pada gambar berikut.



Gambar 5. 2. Susunan Antar Operator dan Data

Gambar ini menunjukkan bagaimana operator dalam perangkat lunak RapidMiner terhubung untuk menjalankan algoritma Apriori dalam analisis pola pembelian pelanggan di UD. Amora Jaya. Proses yang ditampilkan dalam gambar meliputi :

1. Memasukkan Dataset

Data transaksi penjualan diimpor dari berkas *Data Penjualan UD. Amora Jaya.xlsx* ke dalam RapidMiner.

2. Transformasi Data

Data numerik yang bersifat kuantitatif konversi menjadi bentuk kategorikal menggunakan operator *Numerical to Binominal*. Transformasi ini bertujuan untuk mengubah data transaksi menjadi bentuk biner.

3. Penerapan Algoritma Fp-Growth

Operator *Fp-Growth* digunakan untuk mengekstrak itemset yang sering muncul dalam transaksi penjualan berdasarkan nilai *support* yang telah ditentukan. Algoritma ini bertujuan untuk menemukan pola produk yang sering dibeli secara bersamaan.

4. Pembentukan Aturan Asosiasi

Setelah pola itemset ditemukan, operator *Create Association Rules* digunakan untuk membentuk aturan asosiasi berdasarkan nilai minimum *support* dan *confidence* yang telah ditetapkan. Aturan yang terbentuk akan menampilkan hubungan antarproduk, seperti kecenderungan pelanggan membeli produk tertentu secara bersamaan.

5. Visualisasi Hasil

Hasil akhir dari penyiaran ditampilkan dalam bentuk tabel asosiasi yang menampilkan hubungan antar itemset serta nilai *support* dan *confidence* -nya. Tahap selanjutnya yaitu mengklik tombol *Run* sehingga aplikasi menampilkan hasil dari input nilai minimum *support* yang dapat dilihat pada gambar berikut.

No. of Sets	Size	Support	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
1	1	0.859	Sabun Lifeboy					
1	1	0.859	Sunlight					
1	1	0.832	Wipol					
1	1	0.804	Sabun Batang					
1	1	0.799	Sikat Gigi					
1	1	0.755	Saos Dena					
1	1	0.750	Beras					
1	1	0.739	Kecap Bango					
1	1	0.734	Kopi Bintang					
1	1	0.723	Susu Kaleng					
1	1	0.663	Shampo					
1	1	0.587	Rapika					
1	1	0.582	Bumbu Raskit					
1	1	0.582	Minyak Goreng					
1	1	0.560	Tepung Terigu					
2	2	0.810	Sabun Lifeboy	Sunlight				

Gambar 5. 3. Hasil Minimum Support

Association Rule memperlihatkan hasil dari kombinasi *itemset* dengan keputusan dan nilai *support* beserta *confidence*. Tahapan selanjutnya yaitu melihat dengan mengklik menu *Description* dan dapat dilihat pada gambar berikut.

AssociationRules

Association Rules

- [Sunlight, Wipol] --> [Sabun Batang] (confidence: 0.850)
- [Sunlight, Wipol] --> [Sabun Lifeboy, Sabun Batang] (confidence: 0.850)
- [Sabun Lifeboy, Wipol, Sikat Gigi] --> [Kecap Bango] (confidence: 0.850)
- [Sikat Gigi] --> [Sabun Batang] (confidence: 0.850)
- [Sikat Gigi] --> [Sabun Lifeboy, Sabun Batang] (confidence: 0.850)
- [Sunlight, Susu Kaleng] --> [Kopi Bintang] (confidence: 0.850)
- [Sunlight, Kecap Bango] --> [Sabun Lifeboy, Susu Kaleng] (confidence: 0.850)
- [Sunlight, Susu Kaleng] --> [Sabun Lifeboy, Kecap Bango] (confidence: 0.850)
- [Wipol, Sikat Gigi] --> [Kecap Bango] (confidence: 0.851)
- [Sabun Lifeboy, Sunlight, Beras] --> [Kecap Bango] (confidence: 0.851)
- [Sunlight, Wipol, Kecap Bango] --> [Sabun Batang] (confidence: 0.851)
- [Sunlight, Wipol, Kecap Bango] --> [Sabun Lifeboy, Sabun Batang] (confidence: 0.851)
- [Sunlight, Wipol, Kecap Bango] --> [Sabun Lifeboy, Sikat Gigi, Susu Kaleng] (confidence: 0.851)
- [Sikat Gigi, Kecap Bango] --> [Sabun Lifeboy, Sunlight, Wipol, Susu Kaleng] (confidence: 0.851)
- [Sabun Batang] --> [Sabun Lifeboy, Wipol] (confidence: 0.851)
- [Sabun Lifeboy, Beras] --> [Kecap Bango] (confidence: 0.852)
- [Sunlight, Sabun Batang] --> [Sabun Lifeboy, Susu Kaleng] (confidence: 0.852)
- [Sabun Lifeboy, Beras] --> [Sunlight, Sikat Gigi] (confidence: 0.852)
- [Kopi Bintang] --> [Sabun Lifeboy, Sabun Batang] (confidence: 0.852)
- [Shampo] --> [Wipol, Saos Dena] (confidence: 0.852)
- [Shampo] --> [Sabun Lifeboy, Sunlight, Saos Dena] (confidence: 0.852)
- [Shampo] --> [Sabun Lifeboy, Wipol, Saos Dena] (confidence: 0.852)
- [Sunlight, Sikat Gigi, Saos Dena] --> [Sabun Batang] (confidence: 0.852)
- [Shampo] --> [Sunlight, Sabun Batang, Saos Dena] (confidence: 0.852)
- [Wipol, Sabun Batang, Sikat Gigi] --> [Kecap Bango] (confidence: 0.852)
- [Wipol, Sabun Batang, Sikat Gigi] --> [Kopi Bintang] (confidence: 0.852)

Gambar 5. 4. Data Association Rule

Berdasarkan implementasi diatas sesuai aturan yang telah dibuat dengan mengambil 3 itemset dengan minimum *support* 85% dan minimum *Confidence* 80%, maka *AssociationRule* yang didapat dilihat pada gambar berikut.

Tabel 5. 1. Data yang Memenuhi Min Support dan Min Confidence

Jika Membeli	Maka Akan Membeli	Support	Confidence
Pepsodent	Sunlight	80,97%	88,69%
Pepsodent	Royco	85,32%	93,45%
Pepsodent	Vape	83,69%	91,66%
Sunlight	Sabun Lifeboy	80,97%	94,30%
Royco	Sabun Lifeboy	82,60%	88,88%
Royco	Vape	84,23%	90,64%
Royco	Pks	82,06%	88,30%

Berdasarkan perhitungan manual dan pengujian menggunakan aplikasi *rapidminer* maka *rule* atau hasil yang diperoleh sesuai ketentuan dua itemset dengan minimum *support* 85% dan minimum *confidence* 80%. Maka dapat dibuat *rule-rule* yang akan digunakan bagi pemilik toko Ud. Amora Jaya sebagai acuan untuk meningkatkan penjualan sembako sehingga diharapkan dapat meningkatkan penghasilan adalah sebagai berikut :

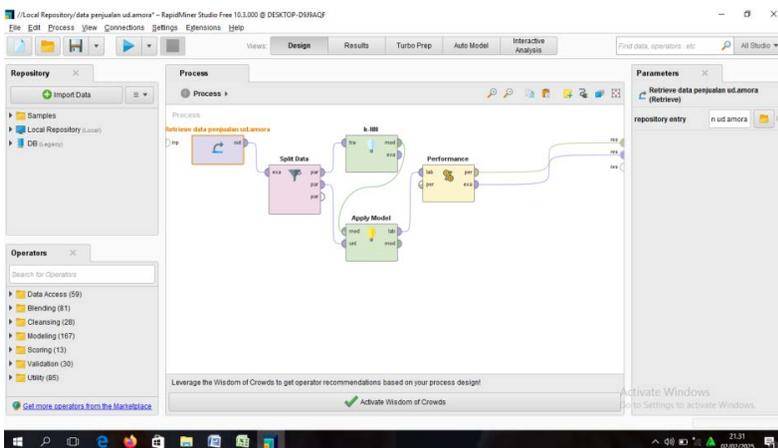
1. Jika membeli pepsodent maka membeli sunlight juga dengan tingkat kepercayaan mencapai 88,69% dan didukung oleh 80,97% dari data keseluruhan.
2. Jika membeli pepsodent maka membeli royco juga dengan tingkat kepercayaan mencapai 93,45% dan didukung oleh 85,32% dari data keseluruhan.
3. Jika membeli pepsodent maka membeli vape juga dengan tingkat kepercayaan mencapai 91,66% dan didukung oleh 83,69% dari data keseluruhan.
4. Jika membeli sunlight maka membeli sabun lifeboy juga dengan tingkat kepercayaan mencapai 94,30% dan didukung oleh 80,97% dari data keseluruhan.
5. Jika membeli royco maka membeli sabun lifeboy juga dengan tingkat kepercayaan mencapai 88,88% dan didukung oleh 82,60% dari data keseluruhan.

6. Jika membeli royco maka membeli vape juga dengan tingkat kepercayaan mencapai 90,64% dan didukung oleh 84,23% dari data keseluruhan.
7. Jika membeli royco maka membeli pks juga dengan tingkat kepercayaan mencapai 88,30% dan didukung oleh 82,06% dari data keseluruhan.

Implementasi RapidMiner Pada Metode K-Nearest Neighbor

Implementasikan metode *K-Nearest Neighbor* pada data tersebut dengan menggunakan aplikasi *rapidminer*. Langkah pertama dalam analisis adalah memasukkan data transaksi ke dalam RapidMiner. Data transaksi yang diperoleh dari toko UD. Amora Jaya disusun ke dalam bentuk format *Excel*, yang kemudian diunggah atau diimport ke RapidMiner.

Setelah itu, dilakukan proses *preprocessing*, yang mencakup pemberian data, normalisasi nilai numerik, serta transformasi data agar dapat digunakan dalam analisis metode KNN. Setelah semua operator yang kita butuhkan sudah berada di halaman main proses selanjutnya hubungkan Dataset Data Penjualan Ud. Amora Jaya.xlsx, output hubungkan ke *exa* ke *exaSplit Data*, dari *Split Datapar* ke *tra KNN*, *mod KNN* ke *mod Apply Model*, dan *par Split Data* ke *unl Apply*, dari *lab Apply Model* ke *lab Performance*, dan *per* dan *exa* ke *Result* yang dapat dilihat pada gambar berikut.



Gambar 5. 5. Susunan Antar Operator dan Data KNN

Gambar ini menggambarkan tahapan utama dalam penerapan algoritma K-Nearest Neighbor (KNN) menggunakan perangkat lunak RapidMiner untuk menganalisis pola penjualan di UD. Amora Jaya. Secara garis besar, gambar ini mencakup langkah-langkah berikut:

1. Memasukkan Dataset

Data transaksi dari *Data Penjualan UD. Amora Jaya.xlsx* diunggah ke RapidMiner sebagai input utama untuk analisis.

2. Praproses Data

Dilakukan proses pembersihan data, normalisasi nilai numerik, serta transformasi data agar siap digunakan dalam metode analisis KNN.

3. Pemrosesan Data dengan KNN

Operator *Split Data* digunakan untuk membagi dataset menjadi data pelatihan dan data pengujian. Operator *KNN* diterapkan pada pelatihan data untuk membentuk klasifikasi model berdasarkan nilai *K* yang telah ditentukan.

4. Penerapan Model KNN

Model KNN yang telah dibor diterapkan pada pengujian data menggunakan operator *Apply Model*. Hasil klasifikasi dikirim ke operator *Performance* untuk mengukur akurasi model.

5. Evaluasi dan Hasil Prediksi

Kinerja Operator menghitung tingkat akurasi prediksi berdasarkan jumlah data yang diklasifikasikan dengan benar dibandingkan jumlah total data. Hasil akhir berupa nilai akurasi sebesar 85,71%, yang menunjukkan sejauh mana model KNN dapat memprediksi pola penjualan di toko

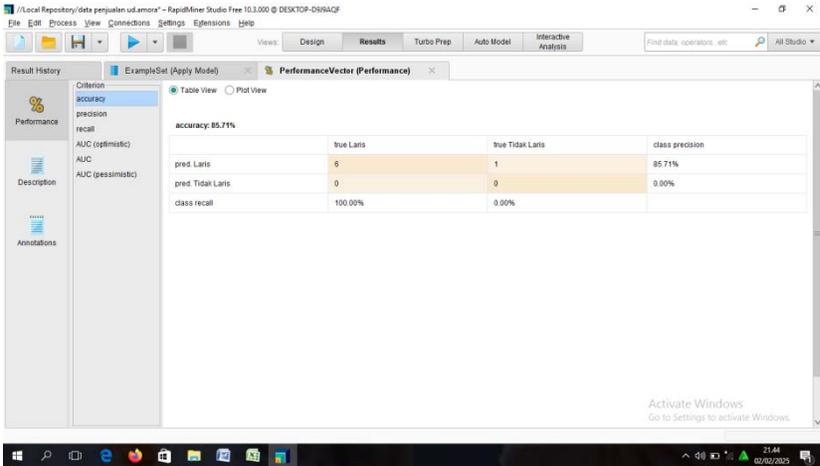
Tahap selanjutnya yaitu mengklik tombol *Run* maka akan muncul sebuah tab *result*, yang isinya sebuah prediksi dari seluruh data yang memenuhi nilai $k=5$, didapatkan hasil prediksi penjualan seperti gambar berikut.

Row No.	Kategori	prediction(K...	confidence_...	confidence_...	Nama Produk	Kuantitas Pr...	Kuantitas Te...
1	Laris	Laris	1	0	Beras	150	138
2	Laris	Laris	1.000	0	Tepung Beras	130	126
3	Laris	Laris	1	0	Teh	140	130
4	Laris	Laris	1	0	Kopi Bintang	145	135
5	Laris	Laris	1	0	Kecap Bango	147	136
6	Laris	Laris	1	0	Sabun Batang	157	148
7	Tidak Laris	Laris	0.546	0.454	Mie Lidi	162	97

Gambar 5. 6. Hasil Prediksi Data

Gambar ini menunjukkan bagaimana metode KNN dapat digunakan dalam analisis data penjualan untuk mengklasifikasikan produk berdasarkan tingkat kelarisannya. Algoritma KNN bekerja dengan membandingkan data transaksi baru dengan data historis dan menentukan kategori berdasarkan mayoritas tetangga terdekat dalam ruang fitur. Misalnya, jika produk seperti beras dan minyak goreng sering dibeli bersamaan dengan pola tertentu, maka data baru yang memiliki karakteristik serupa dapat dikategorikan sebagai produk dengan tingkat penjualan tinggi. Dengan nilai K yang dipilih, model ini dapat memberikan rekomendasi yang membantu pemilik toko dalam menentukan strategi stok dan pemasaran.

Keakuratan sebesar 85,71% menunjukkan bahwa model ini cukup efektif dalam memprediksi pola penjualan di UD. Amora Jaya. Dengan informasi ini, pemilik toko dapat lebih bijaksana dalam mengelola stok, mengurangi risiko produk yang tidak laku, serta meningkatkan efisiensi operasional secara keseluruhan. Selanjutnya yaitu mencari nilai hasil akurasi yang ada pada tampilan aplikasi rapidminer. Adapun hasil *accuracy* dari *performance vektor* sebesar 85,71% seperti gambar berikut ini.

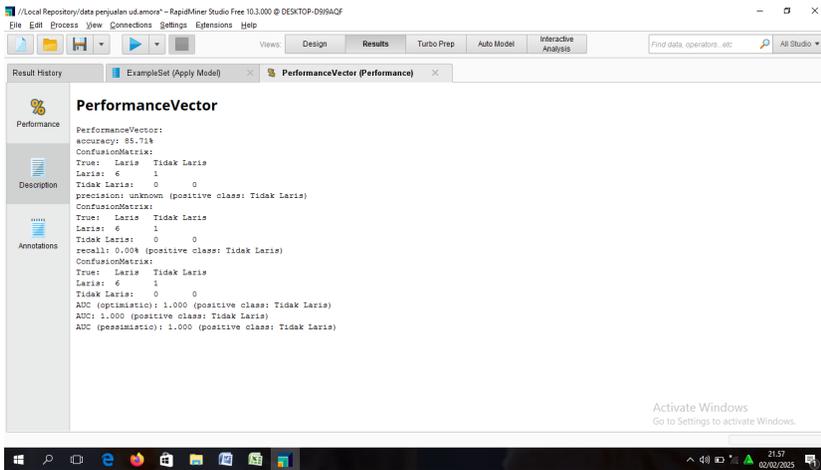


Gambar 5. 7. Hasil Perhitungan Nilai Akurasi

Model klasifikasi KNN yang diterapkan dalam penelitian ini dievaluasi menggunakan metrik akurasi. Berdasarkan hasil pemodelan yang dilakukan dalam RapidMiner, diperoleh akurasi sebesar 85,71% , yang menunjukkan bahwa model memiliki tingkat keakuratan yang cukup tinggi dalam mengklasifikasikan barang.

Selain akurasi, kinerja model juga dapat diukur dengan presisi , recall , dan F1-score , yang berguna untuk mengetahui sejauh mana model dapat membedakan antara kelas “Laris” dan “Tidak Laris” dengan baik.

Dengan hasil yang diperoleh, model klasifikasi KNN ini dapat menjadi alat bantu bagi UD. Amora Jaya dalam mengidentifikasi produk yang memiliki kecenderungan laris di pasar. Hal ini memungkinkan pengelola toko untuk mengambil keputusan berdasarkan data dalam menentukan jumlah stok barang dan strategi pemasaran yang lebih efektif.



Gambar 5. 8. Model Klasifikasi KNN





BAB VI

STUDI KASUS DAN APLIKASI NYATA

Studi Kasus di UD. Amora Jaya

UD. Amora Jaya merupakan salah satu toko grosir yang bergerak di bidang penjualan sembako dan kebutuhan rumah tangga di Kabupaten Labuhanbatu Selatan, Provinsi Sumatera Utara. Toko ini berlokasi di Jl. Bedagai, Kotapinang, yang merupakan kawasan strategis dengan lalu lintas perdagangan yang cukup tinggi. Sebagai penyedia kebutuhan pokok bagi masyarakat sekitar, UD. Amora Jaya melayani pelanggan dari berbagai kalangan, termasuk rumah tangga, warung kecil, serta pedagang eceran yang membeli dalam jumlah besar untuk dijual kembali. Sejak didirikan pada tahun 2020, toko ini telah berkembang menjadi salah satu pemasok utama sembako di wilayahnya, dengan daftar produk yang mencakup beras, gula pasir, minyak goreng, telur, tepung terigu, mie instan, serta berbagai bahan kebutuhan pokok lainnya.

Sebagai usaha yang bergerak di sektor perdagangan bahan pokok, UD. Amora Jaya menghadapi berbagai tantangan dalam mengelola stok barang. Salah satu permasalahan utama adalah fluktuasi permintaan yang tidak menentu, terutama saat terjadi lonjakan kebutuhan menjelang

hari-hari besar seperti bulan Ramadan, Idul Fitri, Natal, dan Tahun Baru. Ketika permintaan melonjak, toko sering mengalami kekurangan stok pada barang-barang yang paling diminati, seperti beras, minyak goreng, dan gula pasir. Akibatnya, pelanggan yang tidak dapat memperoleh barang yang mereka cari akan beralih ke toko lain, yang berpotensi menyebabkan kehilangan pendapatan bagi pemilik toko.

Di sisi lain, terdapat pula risiko kelebihan stok untuk barang-barang tertentu yang permintaannya tidak stabil. Misalnya, produk seperti gula merah, sarden, dan kecap terkadang mengalami penurunan permintaan dalam beberapa bulan tertentu, sehingga stok yang tidak terjual dalam waktu lama dapat menyebabkan penyimpanan yang berlebihan dan potensi kerugian akibat kerusakan produk. Selain itu, barang yang disimpan terlalu lama dalam gudang dapat mengalami penurunan kualitas, terutama untuk produk yang memiliki masa kedaluwarsa terbatas.

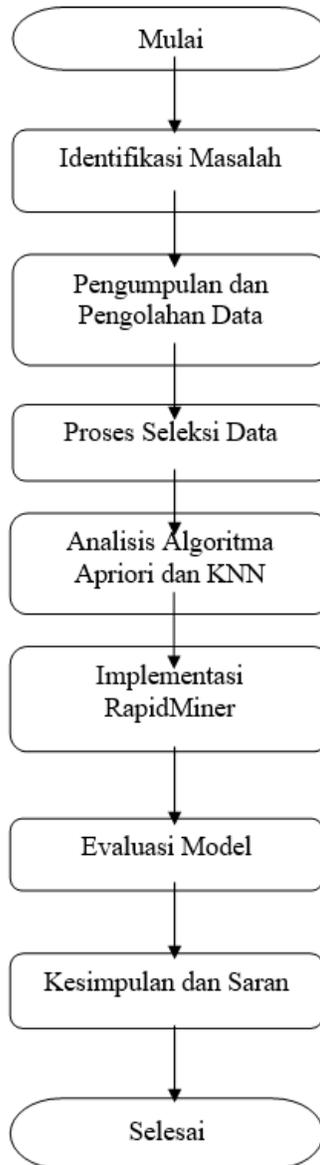
Masalah lain yang dihadapi oleh UD. Amora Jaya adalah pencatatan transaksi yang masih dilakukan secara manual. Sebagian besar data penjualan masih dicatat menggunakan buku atau sistem sederhana yang tidak memungkinkan pemilik toko untuk melakukan analisis mendalam terhadap pola pembelian pelanggan. Akibatnya, sulit bagi pemilik untuk membuat keputusan yang berbasis data terkait jumlah stok yang harus disediakan. Dalam banyak kasus, keputusan pengadaan barang masih mengandalkan intuisi dan pengalaman pribadi, yang tidak selalu menghasilkan estimasi yang akurat.

Untuk mengatasi permasalahan tersebut, UD. Amora Jaya mulai menerapkan teknologi data mining dengan menggunakan algoritma *Apriori* dan *K-Nearest Neighbor* (KNN) melalui perangkat lunak *Rapid Miner*. Penerapan teknologi ini bertujuan untuk mengidentifikasi pola pembelian pelanggan dan melakukan prediksi kebutuhan stok berdasarkan data transaksi sebelumnya. Dengan algoritma *Apriori*, toko dapat menemukan pola keterkaitan antarproduk yang sering dibeli bersama, sehingga dapat mengoptimalkan penataan stok barang dan strategi pemasaran. Sementara itu, metode KNN digunakan untuk memprediksi kebutuhan stok di masa mendatang

berdasarkan pola transaksi historis, sehingga toko dapat melakukan pengadaan barang secara lebih efisien.

Melalui penerapan teknologi ini, UD. Amora Jaya diharapkan dapat mengoptimalkan pengelolaan persediaan barang, mengurangi risiko kelebihan dan kekurangan stok, serta meningkatkan efisiensi operasional toko secara keseluruhan. Dengan adanya analisis berbasis data, pemilik toko dapat membuat keputusan yang lebih cerdas dalam menentukan jumlah barang yang harus disediakan setiap bulannya, sehingga dapat meningkatkan kepuasan pelanggan dan memaksimalkan keuntungan usaha.

Kerangka Kerja Kajian



Gambar 6. 1. Kerangka Kerja Kajian

Dari Gambar 6.1 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 UD. Amora Jaya dalam pengelolaan persediaan barang sembako. Masalah seperti kekurangan stok barang saat permintaan tinggi atau kelebihan stok barang yang dapat menyebabkan kerugian akibat kerusakan diidentifikasi melalui observasi langsung dan wawancara. Tujuannya adalah untuk memahami kebutuhan nyata dari toko, sehingga solusi yang ditawarkan dapat tepat sasaran.

2. Pengumpulan dan Pengolahan Data

Data transaksi penjualan selama 6 bulan terakhir dikumpulkan sebagai bahan analisis utama. Data ini diperoleh dari catatan transaksi manual yang digunakan UD. Amora Jaya. Data yang dikumpulkan dalam rentan waktu tertentu, lalu data diolah agar tidak ada data yang kosong maupun data yang tidak jelas.

3. Proses Seleksi Data

Pada tahap ini, dilakukan pemilihan data yang relevan dengan penelitian. Misalnya, atribut yang berkaitan dengan jenis barang, jumlah pembelian, dan waktu transaksi diseleksi untuk memastikan analisis lebih fokus dan efisien. Data yang telah diseleksi kemudian dipersiapkan dalam format yang dapat diproses oleh perangkat lunak seperti *RapidMiner*.

4. Analisis Algoritma *Apriori* dan KNN

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 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.

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 algoritma *Apriori* dan KNN. 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 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 algoritma *Apriori* dan KNN 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.

Hasil dan Analisis

Dari hasil analisis, ditemukan bahwa beberapa kombinasi barang memiliki asosiasi yang sangat tinggi yaitu :

1. Jika membeli pepsodent maka membeli sunlight juga dengan tingkat kepercayaan mencapai 88,69% dan didukung oleh 80,97% dari data keseluruhan.
2. Jika membeli pepsodent maka membeli royco juga dengan tingkat kepercayaan mencapai 93,45% dan didukung oleh 85,32% dari data keseluruhan.
3. Jika membeli pepsodent maka membeli vape juga dengan tingkat kepercayaan mencapai 91,66% dan didukung oleh 83,69% dari data keseluruhan.

4. Jika membeli sunlight maka membeli sabun lifeboy juga dengan tingkat kepercayaan mencapai 94,30% dan didukung oleh 80,97% dari data keseluruhan.
5. Jika membeli royco maka membeli sabun lifeboy juga dengan tingkat kepercayaan mencapai 88,88% dan didukung oleh 82,60% dari data keseluruhan.
6. Jika membeli royco maka membeli vape juga dengan tingkat kepercayaan mencapai 90,64% dan didukung oleh 84,23% dari data keseluruhan.
7. Jika membeli royco maka membeli pks juga dengan tingkat kepercayaan mencapai 88,30% dan didukung oleh 82,06% dari data keseluruhan.

Sedangkan metode *KNN* berdasarkan hasil perhitungan maka didapatkan hasil prediksi penjualan produk terlaris sebanyak 6 *item* yaitu beras, tepung beras, teh, kopi bintang, kecap bango, dan sabun batang dan tidak laris 1 *item* yaitu mie lidi. Hasil nilai akurasi sebesar 85,71%.





BAB VII

EVALUASI DAN PERBANDINGAN METODE

Efektivitas Kelebihan Algoritma Apriori

Salah satu keunggulan utama dari algoritma *Apriori* adalah kemampuannya dalam mengidentifikasi pola asosiasi antar produk berdasarkan data transaksi historis. Dengan adanya aturan asosiasi yang dihasilkan, pemilik toko dapat memahami kombinasi barang yang sering dibeli bersamaan, sehingga strategi pemasaran dan penataan stok dapat dioptimalkan. Misalnya, jika analisis menunjukkan bahwa minyak goreng sering dibeli bersamaan dengan gula pasir, maka toko dapat menempatkan kedua barang tersebut dalam rak yang berdekatan untuk meningkatkan kenyamanan pelanggan dan mendorong penjualan lebih lanjut.

Selain itu, *Apriori* memungkinkan toko untuk menerapkan strategi promosi berbasis pola pembelian pelanggan. Dengan mengetahui hubungan antar produk, toko dapat menyusun paket promosi seperti diskon untuk pembelian kombinasi tertentu, yang pada akhirnya dapat meningkatkan volume penjualan. Keunggulan lainnya adalah bahwa *Apriori* relatif mudah diterapkan dalam sistem berbasis data besar, terutama jika menggunakan

perangkat lunak seperti *RapidMiner* yang sudah memiliki fitur bawaan untuk aturan asosiasi.

Efektivitas Kekurangan Algoritma Apriori

Meskipun efektif dalam mengidentifikasi hubungan antar barang, *Apriori* memiliki keterbatasan utama dalam kompleksitas komputasi yang tinggi. Algoritma ini harus menghasilkan kandidat kombinasi barang terlebih dahulu sebelum menemukan pola asosiasi yang signifikan. Jika jumlah item dalam dataset sangat besar, maka jumlah kombinasi yang harus diuji juga meningkat secara eksponensial, yang dapat menyebabkan waktu pemrosesan yang lama dan konsumsi sumber daya yang tinggi.

Selain itu, *Apriori* tidak mempertimbangkan faktor waktu atau tren musiman. Algoritma ini hanya mendasarkan hasil analisisnya pada frekuensi kemunculan barang dalam transaksi tanpa mempertimbangkan perubahan pola pembelian dari waktu ke waktu. Hal ini menjadi tantangan bagi bisnis ritel yang memiliki pola permintaan yang berubah-ubah berdasarkan musim atau peristiwa tertentu, seperti meningkatnya permintaan bahan pokok menjelang hari raya.

Efektivitas Kelebihan Metode K-Nearest Neighbor (KNN)

Di sisi lain, metode KNN menawarkan keunggulan dalam prediksi kebutuhan stok berdasarkan data historis. Dengan membandingkan pola pembelian pelanggan sebelumnya, KNN dapat mengklasifikasikan dan memprediksi jumlah barang yang perlu disediakan untuk periode berikutnya. Keakuratan prediksi ini sangat bergantung pada pemilihan jumlah tetangga terdekat (nilai K) serta kualitas data yang digunakan.

Keunggulan lain dari KNN adalah kemampuannya dalam menangani dataset yang tidak memiliki distribusi tertentu. Berbeda dengan metode regresi yang mengasumsikan adanya pola linier dalam data, KNN dapat digunakan untuk data yang memiliki pola non-linier, yang sering ditemukan dalam transaksi penjualan ritel. Selain itu, KNN relatif mudah

diimplementasikan dan dapat diterapkan pada berbagai jenis data tanpa memerlukan asumsi distribusi tertentu.

Efektivitas Kekurangan Metode KNN

Meskipun memiliki banyak keunggulan, metode KNN juga memiliki beberapa keterbatasan. Salah satu kekurangan utamanya adalah sensitivitas terhadap jumlah data. Jika jumlah transaksi yang digunakan sebagai referensi terlalu sedikit, maka hasil prediksi yang dihasilkan bisa menjadi kurang akurat. Selain itu, KNN memerlukan proses komputasi yang cukup tinggi karena harus menghitung jarak antara setiap data uji dengan seluruh data latih untuk menentukan klasifikasi atau prediksi yang paling sesuai.

Selain itu, hasil prediksi KNN sangat dipengaruhi oleh pemilihan nilai K yang optimal. Jika nilai K terlalu kecil, maka model menjadi terlalu sensitif terhadap variasi data dan dapat menghasilkan prediksi yang kurang stabil (*overfitting*). Sebaliknya, jika nilai K terlalu besar, maka model bisa kehilangan kemampuan dalam menangkap pola lokal yang spesifik (*underfitting*), sehingga akurasi prediksi bisa menurun.

Perbandingan Analisis Manual dan RapidMiner

Salah satu aspek utama dalam perbandingan ini adalah waktu dan tenaga yang dibutuhkan dalam proses analisis data. Metode manual cenderung memerlukan lebih banyak waktu karena setiap transaksi harus dicatat dan dianalisis secara individu. Proses ini sering kali menyebabkan keterlambatan dalam pengambilan keputusan, terutama ketika volume transaksi yang besar harus dianalisis untuk menentukan pola pembelian dan kebutuhan stok. Selain itu, kemungkinan kesalahan manusiada dalam pencatatan atau pengolahan data juga cukup tinggi, yang dapat berdampak pada ketidakakuratan hasil analisis.

Sebaliknya, penggunaan *RapidMiner* memungkinkan otomatisasi proses analisis data, sehingga toko dapat memperoleh wawasan bisnis dalam waktu yang lebih singkat. Dengan hanya memasukkan dataset transaksi, *RapidMiner* dapat secara otomatis menghasilkan aturan asosiasi melalui algoritma *Apriori* serta memprediksi kebutuhan stok melalui algoritma KNN dalam hitungan menit. Selain itu, penggunaan perangkat lunak ini

mengurangi ketergantungan pada pengalaman subjektif pemilik toko, sehingga keputusan yang diambil lebih berbasis data yang akurat.

Dalam hal keakuratan, *RapidMiner* memberikan hasil yang lebih presisi dibandingkan analisis manual. Metode manual sering kali tidak dapat mengidentifikasi pola transaksi yang kompleks, karena keterbatasan dalam memproses data dalam jumlah besar. Sebagai contoh, tanpa alat analitik yang canggih, sulit bagi pemilik toko untuk memahami hubungan antar-produk dalam pola pembelian pelanggan atau memprediksi jumlah stok yang dibutuhkan di masa depan dengan tepat.

Sementara itu, *RapidMiner* mampu menemukan hubungan tersembunyi antar barang dalam transaksi dengan bantuan algoritma *Apriori*. Analisis ini memungkinkan pemilik toko mengetahui produk mana yang sering dibeli bersamaan, sehingga mereka dapat mengoptimalkan strategi penyimpanan dan promosi. Selain itu, dengan algoritma KNN, perangkat lunak ini dapat memprediksi jumlah stok yang harus disediakan berdasarkan tren historis, dengan tingkat keakuratan yang lebih tinggi dibandingkan metode manual.

Hasil evaluasi menunjukkan bahwa metode *RapidMiner* memiliki tingkat akurasi hingga 85% dalam memprediksi kebutuhan stok, sementara metode manual sering kali menghasilkan estimasi yang kurang presisi. Faktor ini disebabkan oleh keterbatasan analisis manusia dalam mengidentifikasi pola data yang kompleks serta kurangnya pendekatan berbasis statistik dalam metode manual.

Daftar Pustaka

- [1] P. A. Lestari, M. Nasution, and S. Z. Harahap, "Analisis Data Penjualan Pada Apotek Ritonga Farma Menggunakan Data Mining Apriori," *Informatika*, vol. 12, no. 2, pp. 180–189, 2024.
- [2] K. Handoko and L. S. Lesmana, "Data Mining Pada Jumlah Penumpang Menggunakan Metode Clustering," *Semin. Nas. Ilmu Sos. dan Teknol.*, no. 1, pp. 97–102, 2018.
- [3] I. A. A. Sabri, M. Man, W. A. W. A. Bakar, and A. N. M. Rose, "Web Data Extraction Approach for Deep Web using WEIDJ," *Int. Learn. Technol. Conf.*, no. July, 2019.
- [4] Y. Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4 . 5," *J. Edik Inform.*, vol. 2, pp. 213–219, 2023.
- [5] R. F. Naibaho, S. Z. Harahap, and A. P. Juledi, "Implementasi Data Mining Menggunakan Metode Algoritma Apriori Pada Toko IBR Jaya Untuk Meningkatkan Penjualan," *Informatika*, vol. 12, no. 3, pp. 504–513, 2024.
- [6] V. Jessfry and M. Siddik, "PENERAPAN DATA MINING MENGGUNAKAN ALGORITMA APRIORI DALAM MEMBANGUN SISTEM PERSEDIAAN BARANG," *J. Inf. Syst. Informatics Eng.*, vol. 8, no. 1, pp. 187–199, 2024.

- [7] Y. Ananda, B. Sembiring, and E. A. Sembiring, "Implementasi Data Mining Menggunakan Algoritma Apriori Dalam Menentukan Persediaan Barang," *J. Inf. Syst. Res.*, vol. 1, no. 1, pp. 1–8, 2023.
- [8] E. Widarti, Joosten, and D. Putu Yudia Pratiwi, *Buku ajar pengantar sistem informasi*, no. January. 2024.
- [9] D. Sitanggang, M. Turnip, Y. Laia, and S. P. Tamba, *Algoritma Apriori*. 2023.
- [10] Indah and I. Ali, "PENERAPAN ALGORITMA FP GROWTH UNTUK Mendukung Pola Pembelian Sembako di Toko UCI," *J. Mhs. Tek. Inform.*, vol. 8, no. 2, pp. 1643–1650, 2024.
- [11] N. Purwati and S. Karnila, "Strategi Peningkatan Penjualan Produk Menggunakan Market Basket Analysis," *J. Sist. Inf. Bisnis*, vol. 02, pp. 96–103, 2023, doi: 10.21456/vol13iss2pp96-103.
- [12] N. E. Pratiwi, L. Suryadi, N. E. Pratiwi, F. Ardhy, and P. Riswanto, "PENERAPAN DATA MINING PREDIKSI PENJUALAN MEBEL TERLARIS MENGGUNAKAN METODE K-NEAREST NEIGHBOR(K-NN) (STUDI KASUS : TOKO ZERITA MEUBEL)," *JUSIM (J. Sist. Inf. Musirawas)*, vol. 7, no. 2, pp. 174–184, 2022.
- [13] S. A. Pradana and M. A. Adiguna, "Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode K-Nearest Neighbor (Studi Kasus : Rakuni Bakery, Pastry And Cake)," *J. Ilmu Komput. dan Sci.*, vol. 3, no. 6, pp. 1596–1610, 2024, doi: 10.47065/bits.v3i4.1408.
- [14] A. A. Putri, "Penerapan Data Mining Untuk Memprediksi Penjualan Buah Dan Sayur Menggunakan Metode K-Nearest Neighbor (Studi Kasus : PT . Central Brastagi Utama)," vol. 1, no. 6, pp. 354–361, 2021.
- [15] N. W. Mardiyah, N. Rahaningsih, I. Ali, and K. Neighbor, "PENERAPAN DATA MINING MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR PADA PREDIKSI PEMBERIAN KREDIT DI SEKTOR FINANSIAL," *J. Mhs. Tek. Inform.*, vol. 8, no. 2, pp. 1491–1499, 2024.
- [16] M. R. Nahjan, N. Heryana, A. Voutama, F. I. Komputer, U. S. Karawang, and R. Miner, "IMPLEMENTASI RAPIDMINER DENGAN METODE CLUSTERING K-MEANS UNTUK ANALISA PENJUALAN PADA TOKO OJ CELL," *J. Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 101–104, 2023.

- [17] R. Takdirillah, “Penerapan Data Mining Menggunakan Algoritma Apriori Terhadap Data Transaksi Sebagai Pendukung Informasi Strategi Penjualan,” *J. Pendidik. Inform.*, vol. 4, no. 1, pp. 37–46, 2020, doi: 10.29408/edumatic.v4i1.2081.
- [18] A. P. Hakim *et al.*, “Perancangan Model Arsitektur Sistem Informasi Monitoring Keuangan Warga Menggunakan,” *J. Tek. Inform. dan Sist. Inf.*, vol. 10, no. 1, pp. 323–338, 2023.
- [19] S. P. Dewi and E. Rahayu, “Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode K-Nearest Neighbor,” vol. 3, no. 4, pp. 639–648, 2022, doi: 10.47065/bits.v3i4.1408.

Tentang Penulis

Ega Kurnia Lahir di Sei Rumbia (Kota Pinang) 8 Mei 2002. Saat ini penulis tinggal di Kampung Buluhait Kotapinang. Dalam menempuh pendidikan dimulai dari Sekolah Dasar SDN 115502 Sei Rumbia tamat pada tahun 2015, SMPN 2 Kota Pinang tamat pada tahun 2018, dan di SMAN 1 Kota Pinang tamat pada tahun 2021. Lalu melanjutkan pendidikan S-1 ke Perguruan Tinggi Swasta yaitu Universitas Labuhanbatu di Fakultas Sains dan Teknologi dengan jurusan Sistem Informasi. Penulis baru membuat buku pertama kalinya dengan judul “PENERAPAN DATA MINING MENGGUNAKAN ALGORITMA APRIORI DAN METODE K-NEAREST NEIGHBOR DALAM MENENTUKAN PERSEDIAAN BARANG SEMBAKO UD. AMORA JAYA”. Penulis berharap awal mula buku ini menjadi berkat untuk mencapai karier penulis dihari berikutnya, sehingga penulis mampu untuk mengembangkan bakat sebagai seorang penulis sesuai dengan harapan dan cita-cita penulis sejak kecil.

Angga Putra Juledi, S.Kom., M.Kom. Lahir di kota Padang pada tanggal 19 Juli 1994. Dalam menempuh pendidikan dimulai dari Sekolah Dasar SDN 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 S-1 (Sarjana Universitas Putra Indonesia “YPTK” padang lulus pada tahun 2018 dengan jurusan sistem informasi, dan melanjutkan program Pasca Sarjana (S2) di Universitas Putra Indonesia “YPTK” padang tahun 2019 program studi Teknik Informatika. Konsentrasi system informasi. Saya mengabdikan diri sebagai salah satu dosen di bimbing ilmu komputer pada Fakultas Sains dan Teknologi dengan program studi system informasi di Universitas Labuhanbatu dan menjadi dosen tetap pada tahun 2020 pada kampus tersebut. Pada saat 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 ketiga terbit pada 9 Mei 2024 dengan judul “Dasar-Dasar System Basis Data” dan sampai saat ini masih menulis buku.

Ibnu Rasyid Munthe, S.T., M.Kom. Lahir di Rantau Prapat, Kabupaten Labuhanbatu pada tahun 1987. Ia menempuh pendidikan S-1 di Universitas Nurtanio Bandung dan melanjutkan Pendidikan S2 di Universitas Putra Indonesia (YPTK) Padang. Sebagai sosok yang produktif, dengan aktif menulis artikel ilmiah dan buku. Dedikasinya dalam dunia pendidikan tercermin dari perannya sebagai dosen di Universitas Labuhanbatu. Tidak hanya diruang kelas, Ibnu juga memanfaatkan teknologi digital untuk menyebarkan ilmu pengetahuan. Melalui channel Youtube @manjadda-wajada2022. Ia berbagi berbagai informasi pembelajaran yang bermanfaat. Channel ini menjadi wadah untuk menyampaikan pengetahuan kepada khalayak luas, dengan harapan dapat memberikan dampak positif dan inspirasi bagi banyak orang. Terus berinovasi dalam metode pengajaran dan berbagi ilmu, berusaha membawa perubahan positif dalam dunia pendidikan, khususnya di daerah asalnya. Labuhanbatu. Dengan semangat yang tinggi, ia berharap dapat berkontribusi lebih banyak lagi untuk kemajuan pendidikan di Indonesia.

Marnis Nasution S.Kom., M.Kom. Lahir di Bengkulu pada 30 Maret 1990. Merupakan lulusan Sarjana Komputer jurusan Sistem Informasi Universitas Putra Indonesia “YPTK” Padang pada tahun 2012 dan melanjutkan Magister pada universitas yang sama. Saat ini aktif sebagai dosen pada Universitas Labuhanbatu Fakultas Sains dan Teknologi Jurusan Sistem Informasi dan juga menulis artikel sesuai bidang keilmuan.