JURNAL ILMU KOMPUTER
Volume 3 Nomor 1, Juli, 2025
ISSN: 3031-125X (ONLINE)
Penerbit: Program Studi Teknik Informatika S-2 Universitas Pamulang

Analisis Transaksi Pembayaran Tiket Kereta Api (KAI) Dengan Pembayaran Via Bank Menggunakan Metode K-Nearest Neighbors Dan Naïve Bayes Studi kasus PT XYZ

Rizki Agustian[1], Murni Handayani[2], Abu Khalid Rivai[3]

[1,2,3)]Program Studi Teknik Informatika S-2,Universitas Pamulang

Email: [1]agustianrizky740@gmail.com, [2]murnie_h@yahoo.com, [3]dosen01591@unpam.ac.id

ABSTRACT

In the digital era, train ticket payment patterns are increasingly complex with the increasing use of bank payment methods. PT XYZ, as a train ticket service provider, faces challenges in understanding customer behavior based on available transaction data. The main problem in this study is how to effectively group customer data based on their transaction characteristics to support service improvement and marketing strategies. This study implements two data mining classification algorithms, namely K-Nearest Neighbors (KNN) and Naïve Bayes, to analyze train ticket payment transaction patterns. Processing is carried out through the RapidMiner application, with an approach based on historical transaction data collected and processed using Microsoft Excel. The research methodology includes the stages of data collection, preprocessing, classification modeling, and model performance evaluation based on accuracy, precision, and recall metrics. The results show that the Naïve Bayes algorithm has superior performance compared to KNN, with an accuracy of 99.10%, a precision of 99.07%, and a recall of 99.14%. This indicates that Naïve Bayes is more effective in classifying customer transaction data. Companies can implement the Naïve Bayes algorithm in internal analytics systems to support data-driven decision-making, particularly in marketing strategies and customer service personalization.

Keywords: K-Nearest Neighbors (KNN), Naïve Bayes, data classification, train ticket transactions, RapidMiner.

ABSTRAK

Dalam era digital, pola pembayaran tiket kereta api semakin kompleks seiring meningkatnya penggunaan metode pembayaran via bank. PT XYZ, sebagai penyedia layanan tiket kereta api, menghadapi tantangan dalam memahami perilaku pelanggan berdasarkan data transaksi yang tersedia. Permasalahan utama dalam penelitian ini adalah bagaimana mengelompokkan data pelanggan secara efektif berdasarkan karakteristik transaksi mereka untuk mendukung strategi peningkatan layanan dan pemasaran. penelitian ini mengimplementasikan dua algoritma klasifikasi data mining, yaitu *K-Nearest Neighbors* (KNN) dan *Naïve Bayes*, untuk menganalisis pola transaksi pembayaran tiket kereta api. Pengolahan dilakukan melalui aplikasi *RapidMiner*, dengan pendekatan berbasis data historis transaksi yang dikumpulkan dan diolah menggunakan *Microsoft Excel*. Metodologi penelitian mencakup tahapan pengumpulan data, *preprocessing*, pemodelan klasifikasi, serta evaluasi performa model berdasarkan metrik *accuracy*, *precision*, dan *recall*. Hasil penelitian menunjukkan bahwa algoritma *Naïve Bayes* memiliki performa lebih unggul dibandingkan KNN, dengan *accuracy* sebesar 99,10%, *precision* 99,07%, dan *recall* 99,14%. Ini menunjukkan bahwa *Naïve Bayes* lebih efektif dalam mengklasifikasikan data transaksi pelanggan. perusahaan dapat menerapkan algoritma *Naïve Bayes* dalam sistem analitik internal untuk mendukung pengambilan keputusan berbasis data, khususnya dalam strategi pemasaran dan personalisasi layanan pelanggan.

Kata Kunci: *K-Nearest Neighbors* (KNN), *Naïve Bayes*, klasifikasi data, transaksi tiket kereta api, *RapidMiner.*

1. PENDAHULUAN

Perkembangan teknologi informasi sudah sedemikian pesat salah satunya adalah Internet. Teknologi Internet menghubungkan ribuan jaringan komputer individual dan organisasi di seluruh dunia. Ada banyak kemajuan teknologi yang tersedia dengan komputer. Komputer tidak hanya membantu kita dalam menyelesaikan pekerjaan, tetapi juga dapat menjadi alat yang menyenangkan dan berguna untuk digunakan. Untuk lebih memaksimalkan pekerjaan teknologi berbasis internet sangat mendukung pekerjaan secara online [14]. Penjualan merupakan kegiatan utama dalam dunia bisnis yang berperan besar dalam mencapai tujuan perusahaan. Secara umum penjualan dilakukan dengan mempunyai tujuan eksklusif yaitu mendatangkan suatu keuntungan kepada seseorang yang memasarkan produk-produk atau jasa-jasa tertentu [26].

Transportasi merupakan sebuah sarana umum dengan apapun jenisnya dan dimanapun tempatnya, sangat diperlukan bagi setiap orang yang hendak bepergian, apalagi ke tempat yang tidak mungkin untuk dijangkau hanya dengan berjalan kaki sehingga transportasi menjadi kebutuhan utama. Saat ini begitu banyak transportasi umum yang disediakan baik oleh pemerintah dan swasta maupun perorangan sehingga memudahkan masyarakat untuk mencari alternatif pilihan mode transportasi yang terbaik sesuai dengan kebutuhan dan kemampuan mereka. Salah satu contoh mode transpostasi misalnya kereta api, alat transportasi ini selain memberikan penawaran dan kenyamanan juga memberikan penawaran berupa tarif yang terjangkau. Industri transportasi, khususnya kereta api,mengalami transformasi signifikan dalam cara pembayaran tiket dilakukan. Pembayaran melalui bank menjadi salah satu metode yang paling populer, menawarkan kenyamanan dan efisiensi bagi pelanggan. Namun, dengan meningkatnya volume transaksi, penting bagi perusahaan seperti PT XYZ untuk memahami pola perilaku pelanggan dalam menggunakan layanan pembayaran ini. Analisis transaksi pembayaran dapat memberikan wawasan berharga mengenai preferensi dan kebiasaan pelanggan. Klasifikasi merupakan salah satu proses pada data mining yang bertujuan untuk menemukan hubungan dan menentukan atribut atau kelas label dari sampel yang akan diklasifikasi. Klasifikasi merupakan suatu proses yang bersifat *supervised learning* dan digunakan untuk membedakan kelas label data dengan melalui pencarian model atau fitur yang dapat memprediksi kelas dari suatu objek dengan tepat [2].

Tugas utama pada data mining diantaranya yaitu klasifikasi yang merupakan suatu teknik dengan melihat pada kelakuan dan atribut dari kelompok yang telah didefinisikan [27]. *K-Nearest Neighbor* (KNN) merupakan algoritma yang digunakan dalam klasifikasi teks atau data yang bersifat *supervised*. Algoritma *K-Nearest Neighbor* (KNN) bekerja berdasarkan jarak terpendek dari *query instance* ke data uji atau data train untuk menentukan klasifikasi ketetanggaanya [2]. Salah satu metode yang dapat digunakan dalam data mining ialah klasifikasi *Naive Bayes*. *Naive Bayes Classifier* merupakan metode klasifikasi yang berakar pada *teorema Bayes. Naive bayes* merupakan metode pengklasifikasian berdasarkan pada probabilitas sederhana dan dirancang agar dapat dipergunakan dengan asumsi antar variabel penjelas saling bebas (*independen*). Pada algoritma ini pembelajaran lebih ditekankan pada pengestimasian probabilitas. tujuan dari metode *Naïve Bayes* merupakan untuk menemukan probabilitas ketika kita mengetahui probabilitas tertentu lainnya. Hasil dari perhitungan data mining menggunakan metode klasifikasi Naïve Bayes akan makin berguna jika penyajiannya menarik dan dapat dipahami dengan baik oleh penerima data [14].

Meskipun terdapat potensi besar dalam analisis data ini, tantangan seperti pemilihan fitur yang relevan untuk mencapai hasil yang akurat dan bermanfaat. Oleh karena itu, penelitian ini bertujuan untuk mengeksplorasi dan menerapkan metode *nearest neighbors* (knn) dan *naïve bayes* dalam konteks analisis transaksi pembayaran tiket kereta api di PT XYZ. Data ini bisa sangat berguna dalam berbagai konteks, seperti dalam pengolahan transaksi keuangan, pengelompokan data, dan analisis lainnya.Berikut adalah daftar nama transaksi pembelian tiket kereta api yang digunakan dalam sistem. Banyak data transaksi keuangan disimpan dalam bentuk dokumen, seperti *Microsoft Excel* atau *format spreadsheet* lainnya. Meskipun format ini memudahkan penyimpanan dan pengelolaan data secara manual, pengolahan dan analisis data yang lebih kompleks sering kali memerlukan teknik yang lebih canggih untuk mengoptimalkan pemanfaatan data.

2. METODE

Jenis penelitian yang dugunakan adalah dengan model eksperimen. Penelitian ini berfokus menggunakan metode klasifikasi dengan Agoritma *K-Nearest Neighbors*

(KNN) dan *Naïve Bayes* pada PT XYZ dengan data yang digunakan adalah data penjualan tiket kereta api tujuan malang tahun 2022, 2023, 2024 dengan data sebanyak 15.000 data dan dibagi menjadi 80% (12.000) sebagai *training* dan 20% (3.000) sebagai *testing* untuk digunakan dalam menentukan klasifikasi tren permintaan Tiket Kereta Api. Berikut memperlihatkan proporsi pembagian data yang digunakan dalam proses klasifikasi menggunakan algoritma *K-Nearest Neighbors* dan *Naïve Bayes* pada Tabel 1.

Tabel 1 Contoh Data Set

ID Pelangan	Nomor Invoice	Nama Kereta	Jam Keberangkatan	Kelas Tiket	Pembayaran	Jenis Kelamin	Pekerjaan	Umur	Katagori	Bulan	Tipe
1	INV1214105166528	Gajayana Executive	18.40 WIB	Eksekutif	460000	Laki-laki	Pelajar	16	Mid Price	Agustus 2022	Low
2	INV1215276307507	Majapahit	18.30 WIB	Ekonomi	230000	Perempuan	Wirausaha	57	Low Price	Desember 2022	Low
3	INV1212175858888	Brawijaya	15.40 WIB	Eksekutif	460000	Laki-laki	Pelajar	16	Mid Price	Juni 2022	Low
4	INV1216285880855	Jayabaya Economy	16.45 WIB	Ekonomi	260000	Laki-laki	Karyawan Swasta	36	Low Price	Maret 2022	Low
5	INV1217817716275	Gajayana Luxury Executive	18.40 WIB	Eksekutif	1050000	Laki-laki	Karyawan Swasta	34	High Price	Mei 2022	High
6	INV1218077605564	Jayabaya Executive	16.45 WIB	Eksekutif	460000	Perempuan	Karyawan Swasta	31	Mid Price	Februari 2022	Low
7	INV1212004711781	Jayabaya Economy	16.45 WIB	Ekonomi	260000	Laki-laki	Karyawan Swasta	57	Low Price	Juli 2022	Low
8	INV1215036340524	Jayabaya Executive	16.45 WIB	Eksekutif	460000	Laki-laki	Lainnya	22	Mid Price	Maret 2022	Low
9	INV1215338558210	Brawijaya	15.40 WIB	Eksekutif	460000	Laki-laki	Pegawai Negeri	26	Mid Price	April 2022	Low
10	INV1214452875276	Gajayana Executive	18.40 WIB	Eksekutif	460000	Perempuan	Pegawai Negeri	50	Mid Price	April 2022	Low
11	INV1217044867400	Gajayana Luxury Executive	18.40 WIB	Eksekutif	1050000	Laki-laki	Pegawai Negeri	53	High Price	September 2022	High
12	INV1212026068165	Gajayana Luxury Executive	18.40 WIB	Eksekutif	1050000	Laki-laki	Mahasiswa	22	High Price	Januari 2022	Low
13	INV1213031676551	Majapahit	18.30 WIB	Ekonomi	230000	Laki-laki	Wirausaha	35	Low Price	Mei 2022	Low
14	INV1214407756745	Matarmaja	10.20 WIB	Ekonomi	180000	Laki-laki	Pelajar	18	Low Price	Mei 2022	Low
15	INV1211782860663	Majapahit	18.30 WIB	Ekonomi	230000	Perempuan	Pelajar	14	Low Price	Mei 2022	Low
16	INV1213508458254	Majapahit	18.30 WIB	Ekonomi	230000	Laki-laki	Karyawan Swasta	48	Low Price	Oktober 2022	High
17	INV1214084040842	Gajayana Luxury Executive	18.40 WIB	Eksekutif	1050000	Perempuan	Karyawan Swasta	42	High Price	Juni 2022	Low
18	INV1216240366402	Brawijaya	15.40 WIB	Eksekutif	460000	Perempuan	Pelajar	13	Mid Price	Desember 2022	Low
19	INV1210812488758	Jayabaya Economy	16.45 WIB	Ekonomi	260000	Perempuan	Karyawan Swasta	37	Low Price	Juli 2022	Low
20	INV1215082232482	Gajayana Luxury Executive	18.40 WIB	Eksekutif	1050000	Laki-laki	Karyawan Swasta	35	High Price	September 2022	High
21	INV1213266857847	Jayabaya Executive	16.45 WIB	Eksekutif	460000	Perempuan	Pelajar	13	Mid Price	Agustus 2022	High
22	INV1216604041530	Jayabaya Economy	16.45 WIB	Ekonomi	260000	Laki-laki	Pegawai Negeri	25	Low Price	Agustus 2022	High
23	INV1213788250765	Jayabaya Executive	16.45 WIB	Eksekutif	460000	Perempuan	Lainnya	46	Mid Price	September 2022	Low
24	INV1216271650828	Gajayana Executive	18.40 WIB	Eksekutif	460000	Perempuan	Wirausaha	31	Mid Price	Oktober 2022	Low
25	INV1212685375417	Gajayana Luxury Executive	18.40 WIB	Eksekutif	1050000	Perempuan	Pegawai Negeri	45	High Price	Oktober 2022	High
26	INV1213415313324	Gajayana Luxury Executive	18.40 WIB	Eksekutif	1050000	Perempuan	Wirausaha	47	High Price	Agustus 2022	High
27	INV1217204567325	Majapahit	18.30 WIB	Ekonomi	230000	Perempuan	Lainnya	64	Low Price	Desember 2022	Low
28	INV1211532568612	Gajayana Executive	18.40 WIB	Eksekutif	460000	Perempuan	Pegawai Negeri	44	Mid Price	November 2022	Low
29	INV1213451243760	Majapahit	18.30 WIB	Ekonomi	230000	Perempuan	Karyawan Swasta	34	Low Price	Agustus 2022	High
30	INV1213578336520	Brawijaya	15.40 WIB	Eksekutif	460000	Perempuan	Pelajar	17	Mid Price	Desember 2022	Low
31	INV1217400874102	Jayabaya Economy	16.45 WIB	Ekonomi	260000	Laki-laki	Wirausaha	26	Low Price	Juni 2022	Low
32	INV1215053518004	Gajayana Luxury Executive	18.40 WIB	Eksekutif	1050000	Laki-laki	Karyawan Swasta	27	High Price	Desember 2022	High

Pada tahap pertama penelitian ini prediksi dilakukan berdasarkan data-data yang sudah dilakukan pembersihan dan pengklasifikasian sehingga mendapatkan atribut yang lebih sederhana dibandingkan dengan data aslinya. Data yang digunakan berupa data penjualan tiket kereta api tahun 2022, 2023 & 2024 dengan banyak data 15.000 data. Jadi data yang akan diolah telah memiliki variable tujuan yaitu banyak atau tidaknya permintaan tiket kereta api Hal ini dimaksudkan agar dapat diketahui nilai akurasi dari hasil prediksi berdasarkan penerapan dan tiga algoritma data mining yang digunakan. Hasil dari pengumpulan data yang akan digunakan dalam penelitian, Namun tidak semua data dapat digunakan karena perlu dilakukan *preprocessing* data atau pengolahan data awal untuk mendapatkan data yang baik Adapun rincian 12 atribut yang digunakan

Setelah data terkumpul, dilakukan proses pengujian terhadap model klasifikasi menggunakan dataset yang terdiri atas 15.000 baris data transaksi tiket kereta api. Dataset tersebut telah melewati tahapan pembersihan data (data *cleaning*) secara menyeluruh untuk menjamin kualitas dan keabsahan data yang digunakan dalam proses

analisis. Tahapan pembersihan ini mencakup penghapusan data duplikat, penanganan terhadap nilai yang hilang (*missing values*), pengoreksian kesalahan penulisan atau format, serta standarisasi data untuk memastikan keseragaman nilai pada setiap atribut.

Setelah data dipastikan bersih dan layak digunakan, dilakukan proses pembagian data (data *splitting*) untuk keperluan pelatihan (t*raining*) dan pengujian (testing). Pembagian ini menggunakan proporsi standar yaitu 80% data (sebanyak 12.000 baris) digunakan sebagai data latih**,** dan 20% data (sebanyak 3.000 baris) digunakan sebagai data uji**.** Tujuan dari pembagian ini adalah untuk melatih model pada sebagian besar data yang tersedia dan kemudian menguji kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya.

Pengujian dilakukan untuk mengevaluasi dan membandingkan performa dua algoritma klasifikasi, yakni *K-Nearest Neighbors* dan *Naïve Bayes*. Kedua algoritma tersebut dipilih karena memiliki karakteristik dan pendekatan yang berbeda dalam menyelesaikan permasalahan klasifikasi, sehingga diharapkan mampu memberikan wawasan yang komprehensif terkait efektivitas masing-masing metode dalam konteks data transaksi pembayaran tiket kereta api. Evaluasi kinerja model dilakukan dengan mengukur tiga metrik utama, yaitu

1. *Accuracy*

   Accuracy merupakan rasio prediksi benar (*positif* dan *negative*) dengan keseluruahan data rumus *accuracy* adalah

   $$Accuracy = \frac{True\ Positif + True\ Negatif}{\text{Total Data}} \times 100\% \quad (1)$$

2. *Precision*

   merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif. Rumus precision adalah

   $$Precision = \frac{True\ Positif}{True\ Positif + False\ Positif} \times 100\%. \quad (2)$$

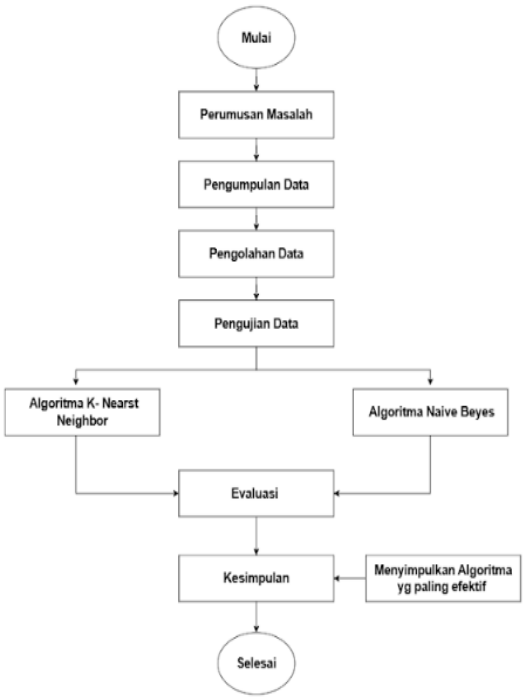
3. *Recall*

   Recall merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif. Rumus *recall* adalah :

   $$Recall = \frac{True\ Positif}{True\ Positif + False\ Negatif} \times 100\% \quad (3)$$

2.1 Alur Penelitian

Perumusan tujuan yang tepat akan membantu peneliti dalam menentukan langkah-langkah strategis yang sistematis dalam menjawab rumusan masalah. Dengan demikian, apabila tujuan penelitian ditetapkan secara jelas dan terstruktur, maka pelaksanaan penelitian, termasuk proses pengumpulan data, analisis, hingga pemecahan masalah, akan dapat dilakukan secara efektif, efisien, dan terfokus.



Gambar 1 Alur Penelitian

Dari Gambar 1 diatas dapat dideskripsikan alur penelitian bahwa pada tahap awal akan dilakukan pengumpulan data yang dibutuhkan dalam penelitian ini, kemudian akan dilakukan pengolahan data dengan menggunakan aplikasi pengolahan data agar lebih mudah diproses pada saat pengujian data. Setelah dilakukan pengolahan data kemudian data tersebut dilakukan pengujian data dengan menggunakan dua algoritma yang telah dipilih diantaranya Agoritma *K-Nearest Neighbors* (KNN) dan *Naïve Bayes*. Pada tahap evaluasi, jika hasil mengalami ketidak sesuaian dengan yang diharapkan maka program dievaluasi Kembali dan dilakukan literasi Kembali pada tahap pengujian, pengumpulan data, maupun pengolahan data berdasarkan hasil evaluasi mana yang perlu dikaji ulang.

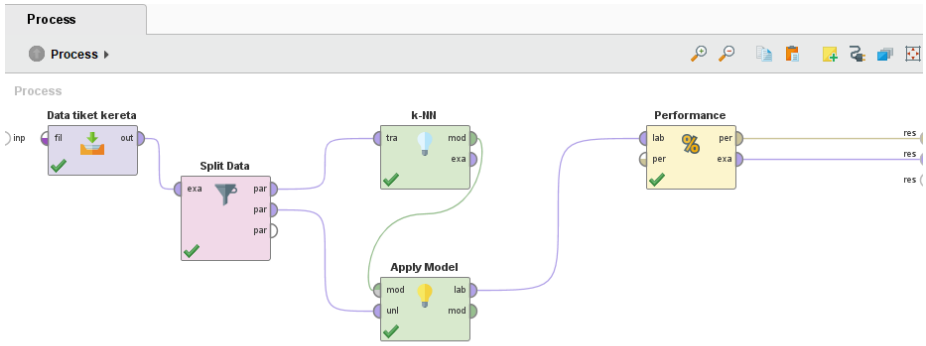
Dari implementasi algoritma tersebut, akan didapatkan hasil kesimpulan berdasarkan hasil evaluasi untuk memperoleh akurasi yang lebih tinggi dari pengujian tersebut dan akan diimplementasikan pada aplikasi yang akan dibuat.

3. HASIL DAN PEMBAHASAN

Pada penelitian ini, yang dilakukan pada PT XYZ dengan menggunakan metode klasifikasi digunakan untuk menguji performa model terhadap dataset yang tersedia. Proses pengujian dilakukan dengan membagi dataset menjadi dua bagian, yaitu, dari jumlah data yang dilakukan pengujian sebanyak 80% untuk *training* dan 20% sebagai *testing.* Hasil evaluasi model klasifikasi dalam penelitian ini dilakukan dengan membandingkan dua algoritma *machine learning* yang digunakan, yaitu algoritma *K-Nearest Neighbor* (KNN) dan *Naïve Bayes*. Evaluasi dilakukan berdasarkan metrik performa utama, yaitu *accuracy*, *presision*, dan *recall*, guna mengukur efektivitas dan efisiensi model dalam melakukan klasifikasi terhadap data yang telah dibagi menjadi data pelatihan (*training*) dan pengujian (*testing*). Berikut ini adalah hasil evaluasi model dua algoritma klasifikasi yang digunakan untuk pengujian adalah algotima *K-Nearest Neighbor* (KNN) dan *Naïve Bayes* hasil dari eveluasi yang sudah dilakukan.

3.1 Evaluasi Model *Predictions K-Nearest Neighbor*

Pada Gambar 2 dibawah ini dijelaskan bahwa evaluasi model prediksi dilakukan dengan tujuan untuk mengukur kinerja dan kemampuan model dalam melakukan klasifikasi terhadap permintaan tiket kereta api. Proses evaluasi ini mencakup pengujian terhadap tingkat *Accuracy*, *Presision*, dan *Recall* yang diperoleh dari hasil prediksi, sehingga dapat diketahui sejauh mana model mampu mengelompokkan data dengan benar dan andal berdasarkan karakteristik transaksi yang tersedia.



Gambar 2 Evaluasi Kinerja Model *Prediction* dengan Algoritma *K-Nearest Neighbors*

Evaluasi kinerja model prediction dengan menggunakan algoritma *K-Nearest Neighbors* dilakukan melalui beberapa tahapan berikut:

A. Data Tiket Kereta Api dalam *Format Excel*

Data yang digunakan dalam penelitian ini berupa data transaksi tiket kereta api yang disimpan dalam format *Microsoft Excel* (.xlsx)**.** Data ini mencakup informasi seperti ID pelanggan, Nomor Invoice, Nama Kereta, Jam Berangkatan, Kelas Tiket, Pembayaran, Jenis Kelamin, Pekerjaan, Umur, Katagori, Bulan,Tipe.

Data tersebut diimpor ke dalam *RapidMiner* untuk dilakukan proses analisis lebih lanjut. Format *Excel* dipilih karena kompatibel dan mudah diproses di dalam *platform RapidMiner*, baik untuk *preprocessing* maupun pemodelan data.

B. Split Data: Pembagian Data 80% dan 20%

Langkah selanjutnya dalam proses analisis adalah melakukan pembagian data (data splitting)**.** Data transaksi dibagi menjadi dua bagian:

1. 80% data digunakan sebagai data latih (*training* set)
2. 20% data digunakan sebagai data uji (testing set)

Tujuan dari pembagian ini adalah agar model yang dibangun dapat dilatih terlebih dahulu menggunakan data training, dan kemudian diuji kemampuannya terhadap data testing yang belum pernah dilihat sebelumnya, sehingga evaluasi model lebih objektif.

C. *Model K-Nearest Neighbors* (KNN)

Setelah proses pembagian data selesai, langkah berikutnya adalah membangun model klasifikasi menggunakan algoritma *K-Nearest Neighbors***.** Dalam penelitian ini, nilai parameter k ditetapkan sebesar 3, yang berarti setiap data uji akan diklasifikasikan berdasarkan mayoritas dari 3 tetangga terdekat dalam data latih. Model KNN digunakan karena sifatnya yang sederhana namun efektif dalam mengklasifikasikan data berdasarkan kemiripan antar fitur.

D. *Apply Model*

Tahapan ini adalah proses menerapkan model KNN yang telah dibangun terhadap data uji. Dengan menggunakan operator "*Apply Model*" di *RapidMiner*, model yang telah dilatih akan digunakan untuk memprediksi label (kelas) dari data testing. Langkah ini bertujuan untuk mengetahui sejauh mana model yang telah dibangun mampu mengklasifikasikan data baru secara akurat dan sesuai dengan kategori yang seharusnya.

E. *Performance: Recall*, *Precision*, dan *Accuracy*

Tahapan akhir adalah evaluasi kinerja model dengan menghitung tiga metrik utama, yaitu:

1. *Accuracy* (Akurasi): Mengukur seberapa banyak prediksi yang benar dibandingkan dengan keseluruhan jumlah data uji.
2. *Precision* (Presisi): Mengukur ketepatan model dalam memprediksi kelas tertentu (positif).
3. *Recall* (Daya Ingat): Mengukur kemampuan model dalam menemukan semua data yang termasuk dalam kelas tertentu (positif).

Pada Tabel 2 ini menyajikan hasil evaluasi performa model klasifikasi menggunakan algoritma *K-Nearest Neighbor* (K-NN) berdasarkan metrik evaluasi seperti *Accuracy*, *Precision*, dan *Recall*. Pengujian dilakukan terhadap data yang telah dibagi menjadi data latih dan data uji, dengan nilai parameter *k* yang telah ditentukan

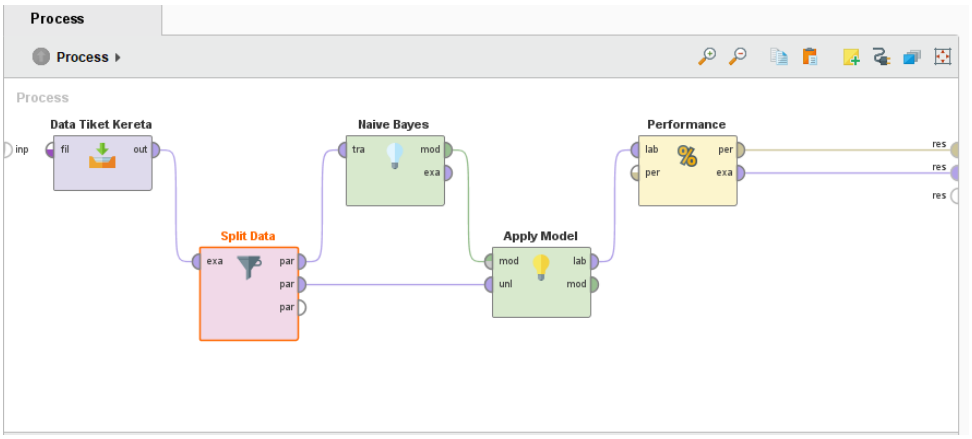
Table 2 Perbandingan Hasil Evaluasi Model *Predictions K-Nearest Neighbor*

	K-Nearest Neighbor

No	Tahun	2022	2023	2024
1	*Accuracy*	62.90 %	56.50 %	70.90 %
2	*Precision*	62.36 %	56.52 %	66.30 %
3	*Recall*	62.17 %	56.53 %	66.71 %

3.2 Evaluasi Model *Predictions Naïve Bayes*

Pada Gambar 3 dibawah ini dijelaskan bahwa evaluasi model prediksi dilakukan dengan tujuan untuk mengukur kinerja dan kemampuan model dalam melakukan klasifikasi terhadap permintaan tiket kereta api. Proses evaluasi ini mencakup pengujian terhadap tingkat *accuracy*, *presision*, dan *recall* yang diperoleh dari hasil prediksi, sehingga dapat diketahui sejauh mana model mampu mengelompokkan data dengan benar dan andal berdasarkan karakteristik transaksi yang tersedia.



Gambar 3 Model *Prediction* dengan Algoritma *Naïve Bayes*

A. Data Tiket Kereta Api dalam *Format Excel*

Data yang digunakan dalam penelitian ini berupa data transaksi tiket kereta api yang disimpan dalam format *Microsoft Excel* (.xlsx). Data ini mencakup informasi seperti, ID pelanggan, Nomor Invoice, Nama Kereta, Jam Berangkatan, Kelas Tiket, Pembayaran, Jenis Kelamin, Pekerjaan, Umur, Katagori, Bulan, Tipe.

Data tersebut diimpor ke dalam RapidMiner untuk dilakukan proses analisis lebih lanjut. Format *Excel* dipilih karena kompatibel dan mudah diproses di dalam *platform RapidMiner*, baik untuk preprocessing maupun pemodelan data.

B. Split Data: Pembagian Data 80% dan 20%

Langkah selanjutnya dalam proses analisis adalah melakukan pembagian data (data *splitting*). Data transaksi dibagi menjadi dua bagian:

1. 80% data digunakan sebagai data latih (*training* set)
2. 20% data digunakan sebagai data uji (testing set)

Tujuan dari pembagian ini adalah agar model yang dibangun dapat dilatih terlebih dahulu menggunakan data training, dan kemudian diuji kemampuannya terhadap data testing yang belum pernah dilihat sebelumnya, sehingga evaluasi model lebih objektif.

C. Model *Naïve Bayes*

Setelah proses pembagian data selesai, langkah selanjutnya adalah membangun model klasifikasi menggunakan algoritma *Naïve Bayes*. Dalam penelitian ini, model *Naïve Bayes* dikembangkan dengan prinsip bahwa setiap fitur dalam data bersifat independen satu sama lain, sehingga probabilitas gabungan dapat dihitung sebagai hasil perkalian dari probabilitas masing-masing fitur. Model ini digunakan karena kesederhanaannya dalam implementasi dan kemampuannya untuk memberikan klasifikasi yang efektif, terutama dalam menghadapi data dengan dimensi tinggi dan jumlah variabel yang banyak.

D. *Apply Model*

Tahapan ini adalah proses menerapkan model KNN yang telah dibangun terhadap data uji. Dengan menggunakan operator "*Apply Model*" di *RapidMiner*, model yang telah dilatih akan digunakan untuk memprediksi label (kelas) dari data testing. Langkah ini bertujuan untuk mengetahui sejauh mana model yang telah dibangun mampu mengklasifikasikan data baru secara akurat dan sesuai dengan kategori yang seharusnya.

E. *Performance*: *Recall*, *Precision*, dan *Accuracy*

Tahapan akhir adalah evaluasi kinerja model dengan menghitung tiga metrik utama, yaitu:

1. *Accuracy* (Akurasi): Mengukur seberapa banyak prediksi yang benar dibandingkan dengan keseluruhan jumlah data uji.

2. *Precision* (Presisi): Mengukur ketepatan model dalam memprediksi kelas tertentu (positif).
3. *Recall* (Daya Ingat): Mengukur kemampuan model dalam menemukan semua data yang termasuk dalam kelas tertentu (positif).

Pada Tabel 3 ini menyajikan hasil evaluasi performa model klasifikasi menggunakan algoritma *Naïve Bayes* berdasarkan metrik evaluasi utama yaitu *Accuracy*, *Precision*, dan *Recall*. Pengujian dilakukan terhadap data uji setelah model dibangun menggunakan data latih yang telah melalui proses pra-pemrosesan dan pembagian data.

Tabel 3 Perbandingan Hasil Evaluasi Model *Predictions Naïve Bayes*

		Naïve Bayes		
No	**Tahun**	**2022**	**2023**	**2024**
1	*Accuracy*	91.90 %	99.10 %	93.60 %
2	*Precision*	91.63 %	99.07 %	91.47 %
3	*Recall*	91.92 %	99.14 %	95.36 %

4. KESIMPULAN

Berdasarkan Hasil penelitian ini menunjukkan bahwa penerapan klasifikasi data penjualan tiket tahun 2022, 2023 & 2024 dengan menggunakan *algoritma K-Nearest Neighbor* dan *Naïve Bayes* Pada PT XYZ penulis meyimpulkan sebagai berikut: Algoritma *Naïve Bayes* adalah pilihan terbaik untuk dataset ini berdasarkan akurasi yang diperoleh pada kedua tahap evaluasi. Dengan nilai akurasi 99.10% pada model *prediction*. Algoritma *Naïve Bayes* secara signifikan lebih unggul dibandingkan dengan Algoritma *K-Nearest Neighbor* (KNN) dalam hal prediksi dan generalisasi data. Dari data tersebut, dapat disimpulkan bahwa algoritma *Naïve Bayes* secara konsisten memberikan performa yang lebih baik dibandingkan dengan algoritma *K-Nearest Neighbor* (KNN) dalam kasus ini. *accuracy* yang tinggi pada model *prediction* menunjukkan bahwa Algoritma *Naïve Bayes* dalam melakukan prediksi dan generalisasi terhadap data yang diberikan. Hasil evaluasi *Confusion Matrix* dengan model *prediction*, Algoritma *Naïve Bayes* adalah pilihan terbaik untuk dataset ini, dengan model *prediction* memiliki akurasi sebesar 99.10%. *Accuracy* mengukur seberapa banyak prediksi yang benar dari total prediksi yang dilakukan memiliki *Precision*

sebesar 99.07%. Presisi mengukur seberapa tepat model dalam memprediksi positif. Model *prediction* memiliki *recall* sebesar 99.14%.

5. DAFTAR PUSTAKA

[1] Alghifari, F. and Juardi, D. (2021) 'Penerapan Data Mining Pada Penjualan Makanan Dan Minuman Menggunakan Metode Algoritma Naïve Bayes', *Jurnal Ilmiah Informatika*, 9(02), pp. 75–81. Available at: https://doi.org/10.33884/jif.v9i02.3755.

[2] Amanda Pratiwi, Ananto Tri Sasongko and K. Pramudito, D. (2023) 'Analisis Prediksi Gilingan Plastik Terlaris Menggunakan Algoritma K-Nearest Neighbor Di Cv Menembus Batas', *Jurnal Informatika Teknologi dan Sains (Jinteks)*, 5(3), pp. 437–445. Available at: https://doi.org/10.51401/jinteks.v5i3.3323.

[3] Apriyadi, A., Lubis, M.R. and Damanik, B.E. (2022) 'Penerapan Algoritma C5.0 Dalam Menentukan Tingkat Pemahaman Mahasiswa Terhadap Pembelajaran Daring', *Komputa : Jurnal Ilmiah Komputer dan Informatika*, 11(1), pp. 11–20. Available at: https://doi.org/10.34010/komputa.v11i1.7386.

[4] Arta, I.K.J., Indrawan, G. and Rasben Dantes, G. (2019) 'Data Mining Rekomemdasi Calon Mahasiswa Berprestasi di STMIK Denpasar Menggunakan Metode Technique For Other Reference By Similarity to Ideal Solution', *Jurnal Ilmu Komputer Indonesia (JIKI)*, 4(1), pp. 11–21.

[5] Azizah, Q., Masriah, M. and Atmojo, W.T. (2022) 'Perancangan Data Warehouse Sistem Penerimaan Siswa Baru Menggunakan Online Analytical Processing (OLAP) di TK IT Mutiara', *Dirgamaya: Jurnal Manajemen dan Sistem Informasi*, 2(2), pp. 35–47. Available at: https://doi.org/10.35969/dirgamaya.v2i2.273.

[6] Diana, Y. *et al.* (2023) 'Analisa Penjualan Menggunakan Algoritma K-Medoids Untuk Mengoptimalkan Penjualan Barang', *JOISIE Journal Of Information System And Informatics Engineering*, 7(1), pp. 97–103.

[7] Eka Saputri, Danilla Oktaviyana Nurlyta Lestariningsih, E. (2023) 'Implementasi Data Mining Pada Penjualan Sepatu Menggunakan Algoritma Apriori (Kasus Toko Sepatu 3Stripesid)', *Jurnal Algoritma*, 4(1), pp. 667–676.

[8] Ferdyandi, M., Setiawan, N.Y. and Abdurrachman Bachtiar, F. (2022) 'Prediksi

Potensi Penjualan Makanan Beku Berdasarkan Ulasan Pengguna Shopee Menggunakan Metode Decision Tree Algoritma C4.5 Dan Random Forest (Studi Kasus Dapur Lilis)', *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 6(2), pp. 588–596. Available at: http://j-ptiik.ub.ac.id.

[9] Fitrianti, I., Voutama, A. and Umaidah, Y. (2023) 'Clustering Film Populer Pada Aplikasi Netflix Dengan Menggunakan Algoritma K-Means Dan Metode CRISP-DM Clustering Popular Movies on Netflix App Using K-Means Algorithm and CRISP-DM Method', *Jtsi*, 4(2), pp. 301–311.

[10] Ghita, F. and Trisminingsih, R. (2021) 'Pengujian Data Warehouse SOLAP untuk Komoditas Pertanian Indonesia Data Warehouse Testing in SOLAP for Indonesia Agricultural Commodities', *Jurnal Ilmu Komputer Agri-Informatika*, 8(1), pp. 42–56. Available at: http://solap.apps.cs.ipb.ac.id.

[11] Hijrah, Maulidar and Adria (2022) 'Analisis RapidMiner Dan Weka Dalam MemprediksiKualitas Kinerja Karyawan Menggunakan MetodeAlgoritma C4.5', *Http://Jurnal.Mdp.Ac.Id*, 9(2), pp. 1655–1665.

[12] Ismiyana Putri, D. and Yudhi Putra, M. (2023) 'Komparasi Algoritma Dalam Memprediksi Perubahan Harga Saham Goto Menggunakan Rapidminer', *Jurnal Khatulistiwa Informatika*, 11(1), pp. 14–20. Available at: https://doi.org/10.31294/jki.v11i1.16153.

[13] Manalu, D.A. and Gunadi, G. (2022) 'Implementasi Metode Data Mining K-Means Clustering Terhadap Data Pembayaran Transaksi Menggunakan Bahasa Pemrograman Python Pada Cv Digital Dimensi', *Infotech: Journal of Technology Information*, 8(1), pp. 43–54. Available at: https://doi.org/10.37365/jti.v8i1.131.

[14] Mardiani, E. *et al.* (2023) 'Komparasi Metode Knn, Naive Bayes, Decision Tree, Ensemble, Linear Regression Terhadap Analisis Performa Pelajar Sma', *Innovative: Journal Of …*, 3(2), pp. 13880–13892. Available at: http://j-innovative.org/index.php/Innovative/article/view/1949%0Ahttp://j-innovative.org/index.php/Innovative/article/download/1949/1468.

[15] Mukrodin, M., Taufiq, R. and Ermi, D.S.R. (2023) 'Data Mining Clustering Data Obat-Obatan Menggunakan Algoritma K-Means Pada Rsu an Ni'Mah Wangon', *JIKA (Jurnal Informatika)*, 7(2), p. 165. Available at: https://doi.org/10.31000/jika.v7i2.7553.

[16] Naldy, E.T. and Andri, A. (2021) 'Penerapan Data Mining Untuk Analisis Daftar Pembelian Konsumen Dengan Menggunakan Algoritma Apriori Pada Transaksi Penjualan Toko Bangunan MDN', *Jurnal Nasional Ilmu Komputer*, 2(2), pp. 89–101. Available at: https://doi.org/10.47747/jurnalnik.v2i2.525.

[17] Purwanto, J. and Renny, R. (2021) 'Perancangan Data Warehouse Rumah Sakit Berbasis Online Analytical Processing (OLAP)', *Jurnal Teknologi Informasi dan Ilmu Komputer*, 8(5), pp. 1077–1088. Available at: https://doi.org/10.25126/jtiik.2021854232.

[18] Rafi Nahjan, M., Nono Heryana and Apriade Voutama (2023) 'Implementasi Rapidminer Dengan Metode Clustering K-Means Untuk Analisa Penjualan Pada Toko Oj Cell', *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(1), pp. 101–104. Available at: https://doi.org/10.36040/jati.v7i1.6094.

[19] Ramadhantya, A.S. (2024) 'Penggunaan Rapidminer Untuk Memprediksi Kelulusan Mahasiswa Dengan Algorithm Naive Bayes', 10(1), pp. 52–60.

[20] Rizal, R., Martanto, M. and Arie Wijaya, Y. (2022) 'Analisa Dataset Software Defined Network Intrusion Menggunakan Algoritma Deep Learning H2O', *JATI (Jurnal Mahasiswa Teknik Informatika)*, 6(2), pp. 747–757. Available at: https://doi.org/10.36040/jati.v6i2.5724.

[21] Rizkiawan, A. and Wahyudi, T. (2023) 'Implementasi Data Mining Untuk Memprediksi Member Baru Menggunakan Linear Regression Pada Pt. Gsi', *Jurnal Tekinkom (Teknik Informasi dan Komputer)*, 6(1), pp. 118–126. Available at: https://doi.org/10.37600/tekinkom.v6i1.707.

[22] Romli, I. (2021) 'Penerapan Data Mining Menggunakan Algoritma K-Means Untuk Klasifikasi Penyakit Ispa', *Indonesian Journal of Business Intelligence (IJUBI)*, 4(1), p. 10. Available at: https://doi.org/10.21927/ijubi.v4i1.1727.

[23] Rosika, H. *et al.* (2024) 'Pakaian Menggunakan Metode K-Means Pada era digital perkembangan teknologi semakin berkembang khusunya pengelolaan data penjualan menjadi semakin krusial bagi bisnis untuk memahami perilaku konsumen , mengedentifikasi beberapa kelompok atau cluster memili', 5, pp. 221–231.

[24] Rusdah, R. and Bregastantyo, B.A. (2023) 'Model Prognosis Masa Pengobatan Pasien Tuberkulosis Dengan Metode C4.5', *Jurnal Teknologi Informasi dan Ilmu*

Komputer, 10(6), pp. 1197–1204. Available at: https://doi.org/10.25126/jtiik.1067393.

[25] Saputra, D.B. *et al.* (2024) 'Penerapan Model Crisp-Dm Pada Prediksi Nasabah Kredit', 7(2021), pp. 240–247.

[26] Virza, V.P., Tri Pranot, G. and Eko Putra, F. (2023) 'Klasifikasi Kebutuhan Sparepart Dengan Algoritma K-Nearest Neighbor Untuk Meningkatkan Penjualan Sparepart', *Bulletin of Information Technology (BIT)*, 4(3), pp. 287–293. Available at: https://doi.org/10.47065/bit.v4i3.729.

[27] Winantu, A. and Khatimah, C. (2023) 'Perbandingan Metode Klasifikasi Naive Bayes Dan K-Nearest Neighbor Dalam Memprediksi Prestasi Siswa', *INTEK : Jurnal Informatika dan Teknologi Informasi*, 6(1), pp. 58–64. Available at: https://doi.org/10.37729/intek.v6i1.3006.

[28] Yoliadi, D.N. (2023) 'Data Mining Dalam Analisis Tingkat Penjualan Barang Elektronik Menggunakan Algoritma K-Means', *Insearch: Information System Research Journal*, 3(01). Available at: https://doi.org/10.15548/isrj.v3i01.5829.

[29] Yudiana, Y., Yulia Agustina, A. and Nur Khofifah, dan (2023) 'Prediksi Customer Churn Menggunakan Metode CRISP-DM Pada Industri Telekomunikasi Sebagai Implementasi Mempertahankan Pelanggan', *Indonesian Journal of Islamic Economics and Business*, 8(1), pp. 01–20. Available at: http://e-journal.lp2m.uinjambi.ac.id/ojp/index.php/ijoieb.

[30] Zahra, F., Ridla, M.A. and Azise, N. (2024) 'Implementasi Data Mining Menggunakan Algoritma Apriori Dalam Menentukan Persediaan Barang (Studi Kasus : Toko Sinar Harahap)', *JUSTIFY : Jurnal Sistem Informasi Ibrahimy*, 3(1), pp. 55–65. Available at: https://doi.org/10.35316/justify.v3i1.5335.