

Penerapan Algoritma Naïve Bayes Untuk Klasifikasi Data Pengisian ATM

Application of Naïve Bayes Algorithm for ATM Filling Data Classification

Dahri Yani Hakim Tanjung^{a,1}, Khairul Ummi^{b,2}

Universitas Potensi Utama, Jl. K.L Yos Sudarso km 6,5 no. 3A, Medan, 20241, Indonesia^a
notashapire@gmail.com¹, ummi12gibmie@gmail.com²

ABSTRAK

Uang merupakan alat pembayaran yang sah untuk memenuhi kebutuhan ekonomi manusia, seiring dengan perkembangan jaman dan teknologi muncul transaksi non tunai dalam sistem pembayaran. Hal tersebut mendorong pihak perbankan menyediakan jasa penyedia uang dalam bentuk electronic delivery channels berupa ATM. Fungsi ATM sendiri memberi kemudahan kepada konsumen dalam bertransaksi baik tunai maupun non tunai. Data pengisian ATM dari berbagai Bank yang transaksi perhari cukup banyak dapat diolah dengan menggunakan data mining. Data mining mampu mengolah data dengan jumlah yang sangat banyak salah satunya menggunakan fungsi klasifikasi sehingga menghasilkan pengetahuan baru. Data mining dengan fungsi klasifikasi diharapkan mampu dalam pengolahan data sehingga dapat memonitoring terhadap data pengisian ATM. Penerapan metode Naïve bayes digunakan untuk klasifikasi status pengisian ATM yaitu Isi atau Tidak Isi, dimana proses pengujiannya menggunakan bantuan tools Rapid Miner yang bertujuan untuk mengukur tingkat akurasi dari metode Naïve Bayes yang digunakan. Dataset pengisian Mesin ATM yang digunakan dalam penelitian ini yaitu 55 record dengan 6 attribut berupa persentase CIB, Lokasi, Status ATM, Restock, CashOut dan EndCash. Dari penelitian ini, Algoritma Naïve Bayes mampu dalam menganalisa terhadap data pengisian ATM dengan berhasil mengklasifikasikan 50 data dari 55 data yang diuji dengan nilai probabilitas sebesar 0,004 untuk klasifikasi Isi dan persentase keakuratan sebesar 90,91%, untuk class precision Isi yaitu 94,74% untuk precision tidak isi 88,89%, sedangkan untuk class recall Isi yaitu 81,82% dan class recall tidak isi 96,97%. Hasil tersebut, membuktikan bahwa metode Naïve Bayes dapat digunakan untuk membantu dalam prediksi pengisian ATM, karena menghasilkan tingkat keakuratan yang cukup tinggi

Kata Kunci : Data Mining, Klasifikasi, Naïve Bayes, Pengisian ATM

ABSTRACT

Money is a legal means of payment to meet human economic needs, along with the times and technology, non-cash transactions appear in the payment system. This encourages banks to provide money supply services in the form of electronic delivery channels in the form of ATMs. The ATM function itself makes it easy for consumers to transact both cash and non-cash. ATM filling data from various banks that have quite a lot of transactions per day can be processed using data mining. Data mining is able to process very large amounts of data, one of which uses a classification function to produce new knowledge. Data mining with a classification function is expected to be able to process data so that it can monitor ATM charging data. The application of the Nave Bayes method is used to classify ATM charging status, namely Filled or Not Filled, where the testing process uses the help of Rapid Miner tools which aims to measure the level of accuracy of the Naïve Bayes method used. The ATM machine filling dataset used in this study are 55 records with 6 attributes in the form of CIB percentage, Location, ATM Status, Restock, CashOut and EndCash. From this study, the Naïve Bayes Algorithm is able to analyze ATM charging data by successfully classifying 50 data from 55 data tested with a probability value of 0.004 for Content classification and an accuracy percentage of 90.91%, for class precision Content, which is 94.74%. For precision, the content is not 88.89%, while for the recall class the content is 81.82% and the recall class is not filled in 96.97%. These results prove that the Naïve Bayes method can be used to help predict ATM filling, because it produces a fairly high level of accuracy

Keywords : Data Mining, Classification, Naïve Bayes, ATM filling

Info Artikel :

Disubmit: 11 April 2022

Direview: 20 April 2022

Diterima : 23 Juni 2022

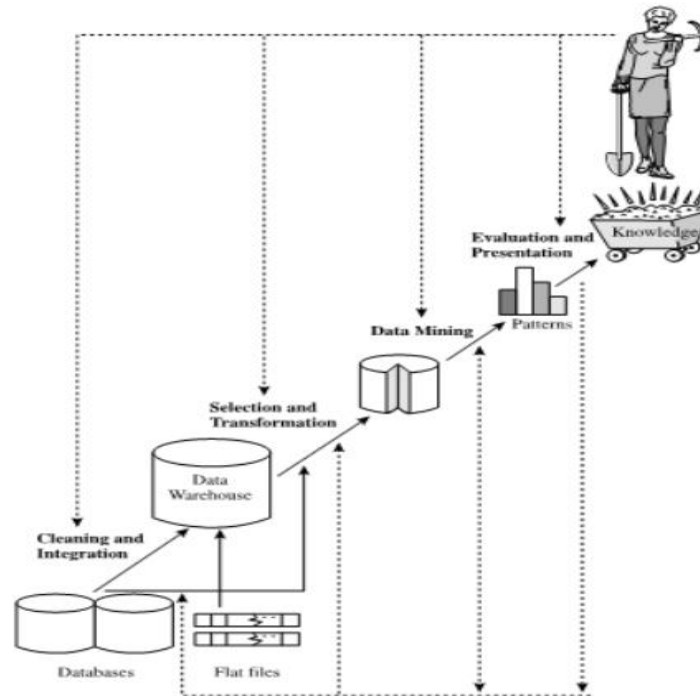
1. PENDAHULUAN

Uang merupakan alat pembayaran yang sah untuk memenuhi kebutuhan ekonomi manusia, seiring dengan perkembangan jaman dan teknologi muncul transaksi non tunai dalam sistem pembayaran. Hal tersebut mendorong pihak perbankan menyediakan jasa penyedia uang dalam bentuk electronic delivery channels berupa ATM. Fungsi ATM sendiri memberi kemudahan kepada konsumen dalam bertransaksi baik tunai maupun non tunai. Seiring dengan meningkatnya transaksi yang dilakukan menggunakan ATM maka kinerja dari mesin tersebut harus tetap optimal agar dapat terus melayani kebutuhan masyarakat dalam bertransaksi baik tunai maupun non tunai. Berdasarkan hal tersebut dibutuhkan monitoring dalam pengisian uang pada mesin ATM sehingga dapat memantau persediaan uang yang ada didalam ATM dan mengurangi terjadinya downtime yang dapat terjadi kapan saja.

Data pengisian ATM dari berbagai Bank yang transaksi perhari cukup banyak dapat diolah dengan menggunakan data mining. Data mining mampu mengolah data dengan jumlah yang sangat banyak salah satunya menggunakan fungsi klasifikasi sehingga menghasilkan pengetahuan baru. Data mining dengan fungsi klasifikasi diharapkan mampu dalam pemahaman dan pengolahan data sehingga dapat memonitoring terhadap data pengisian ATM.

A. Data Mining

Data mining merupakan ekstraksi implisit, dari sebuah data yang kompleks yang tidak diketahui, yang berpotensi berguna dan menghasilkan informasi baru atau pengetahuan baru. Ide dari data mining adalah untuk merancang sebuah software yang dapat menyaring database dengan membentuk pola secara otomatis. Pola yang unik akan di generalisasikan untuk membuat ramalan akurat untuk data di masa depan. Data Mining digunakan untuk mengekstraksi informasi dari data mentah dalam sebuah database sehingga menghasilkan sebuah informasi yang diekspresikan dalam bentuk yang dapat dipahami dan dapat digunakan untuk berbagai tujuan seperti klasifikasi pengisian ATM. Data mining sering didefinisikan sebagai penemuan informasi tersembunyi dalam database. Data mining menyediakan metodologi dan teknologi untuk mengidentifikasi informasi data yang berguna untuk pengambilan keputusan[1][2]. Dan data mining merupakan bagian dari KDD (*Knowledge Discovery from Data*) [3], dimana proses KDD dapat dilihat pada gambar 1 berikut.



Gambar 1. Proses KDD

Pada gambar 1 diatas dapat diuraikan tahap-tahap dari proses KDD, yaitu:

- Cleaning (Pembersihan data), merupakan proses menghilangkan/pembersihan data yang tidak konsisten (inconsistent)
- Integration, merupakan proses penggabungan data dari databse lain kedalam satu database baru
- Selection, merupakan proses pemilihan data yang digunakan dalam database, hanya memilih data yang akan digunakan untuk dianalisis
- Transformation, merupakan proses pengklasifikasian data kedalam bentuk/format yang sesuai dan bergantung pada jenis data atau pola informasi yang akan dicari
- Data Mining, merupakan proses utama/bagian paling penting dalam mencari pola dengan menggunakan metode untuk menemukan informasi yang tersembunyi dari sebuah data.
- Pattern Evaluation, merupakan proses evaluasi/identifikasi terhadap pola yang ditemukan
- Knowledge Presentation, merupakan proses penyajian dari pengetahuan baru dengan menggunakan metode yang diperoleh pengguna

B. Klasifikasi

Klasifikasi merupakan metode Supervised learning (pembelajaran terawasi) yang dipergunakan untuk memperkirakan kelas asal suatu objek yang labelnya tidak diketahui, metode yang mencoba menemukan hubungan antara atribut masukan dan atribut target. Tujuan klasifikasi untuk meningkatkan akurasi hasil yang diperoleh dari data[4]. Klasifikasi untuk prediksi pada data mining merupakan dua bentuk tugas analisis data yang digunakan untuk mengekstrak model yang memvisualisasikan kelas data atau tren dalam memprediksi data masa depan [5][6]. Proses klasifikasi memiliki dua tahap; pertama adalah proses pembelajaran di mana set data pelatihan berada dianalisis dengan algoritma klasifikasi. Model yang dipelajari atau pengklasifikasi disajikan dalam bentuk klasifikasi aturan atau pola. Tahap kedua adalah penggunaan model untuk klasifikasi, dan kumpulan data uji digunakan untuk memperkirakan keakuratan aturan klasifikasi[7]. Metode Klasifikasi telah banyak digunakan dalam berbagai bidang dalam kehidupan sehari-hari dan untuk penelitian

selanjutnya sedang dilakukan pengembangan dari algoritma yang berbeda dengan keahlian dari berbagai bidang ilmu[8]. Klasifikasi memiliki tujuan untuk memprediksi sejumlah data kompleks yang tidak diketahui untuk setiap kasus dalam data secara akurat. Misalnya, dalam mengidentifikasi pengisian uang ATM menggunakan model klasifikasi, yang dimulai dengan mengumpulkan data dengan label kelas yang diketahui kemudian diprediksi dan mungkin dikembangkan berdasarkan data yang diamati untuk proses pengisian selama periode waktu tertentu[9].

C. Naïve Bayes Classifier

Naïve Bayes merupakan metode klasifikasi dari data mining dengan probabilitas dan statistik yang dikemukakan oleh Thomas Bayes seorang ilmuwan British [10]. Algoritma naïve bayes mengasumsikan bahwa semua variabel bebas adalah nilai variabel kelas. Metode ini hanya membutuhkan jumlah data pelatihan untuk menentukan perkiraan parameter yang diperlukan dalam Proses klasifikasi. Algoritma Naïve Bayes sering bekerja jauh lebih baik dalam situasi dunia nyata yang paling kompleks dari yang diharapkan. Di samping itu, algoritma Naïve Bayes tidak membutuhkan banyak data latih untuk menentukan estimasi parameter proses klasifikasi.[11][12] Teorema Bayes merupakan rumus matematika yang digunakan untuk menentukan peluang bersyarat yang dapat dilihat pada persamaan 1

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (1)$$

Dalam persamaan 1, dimana variable X adalah data dengan Class yang belum diketahui, variabel H merupakan hipotesis data kelas khusus, $P(H|X)$ merupakan probabilitas hipotesis H berdasarkan kondisi X (probabilitas posterior), $P(H)$ merupakan probabilitas Hipotesis H (probabilitas prior), $P(X|H)$ merupakan Probabilitas X didasarkan pada kondisi hipotesis H, dan $P(X)$ merupakan Probabilitas X.

Teorema Bayes kemudian dikembangkan mengingat berlakunya hukum probabilitas total, menjadi:

$$P(H|X) = \frac{P(X|H)}{\sum_{i=1}^n P(H_i|X)} \cdot P(H) \quad (2)$$

Keterangan

i : 1,2,3, ..., n jumlah data Hipotesis (prior probabilitas)

dimana : $H_1 \cup H_2 \cup H_3 \dots \cup H_n = S$

S : Probabilitas total H

Untuk menjabarkan Teorema Bayes, perlu diketahui bahwa proses pembagian terstruktur membutuhkan petunjuk-petunjuk sebagai penentu kelas mana yang sesuai bagi sampel yg dianalisis tersebut. Oleh sebab itu, metode Naive Bayes disesuaikan sebagai berikut

$$P(C|F_1, \dots, F_n) = \frac{P(F_1, \dots, F_n|C)}{P(F_1, \dots, F_n)} \cdot P(C) \quad (3)$$

Di mana :

C : merupakan representasi dari kelas

$F_1 \dots F_n$: merupakan representasi karakteristik petunjuk untuk melakukan klasifikasi.

Oleh Karena itu, dapat dituliskan rumus sederhana sebagai berikut:

$$Posterior = \frac{Prior * Likelihood}{Evidence} \quad (4)$$

Nilai Evidence tidak berubah pada setiap kelas. Nilai posterior akan dibandingkan nantinya dengan nilai posterior yang terdapat pada kelas lain sebagai penentu kelas apa yang akan diklasifikasikan. Untuk lebih lanjut rumus dari naïve bayes dapat dijabarkan dengan (C,F1, ... , Fn) menggunakan persamaan perkalian sebagai berikut:

$$P(C|F) = P(F_1|C).P(F_2|C).P(F_3|C) \dots P(F_n|C).P(C) \quad (4)$$

Pada persamaan 4 di atas merupakan rumus yang akan digunakan dalam proses klasifikasi dari model teorema Naïve Bayes. Dan selanjutnya proses klasifikasi untuk data yang berlanjut(kontinyu) dapat menggunakan rumus Densitas Gauss sebagai berikut:

$$P(X_i = x_i | Y = y_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}}} e^{-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2}} \quad (5)$$

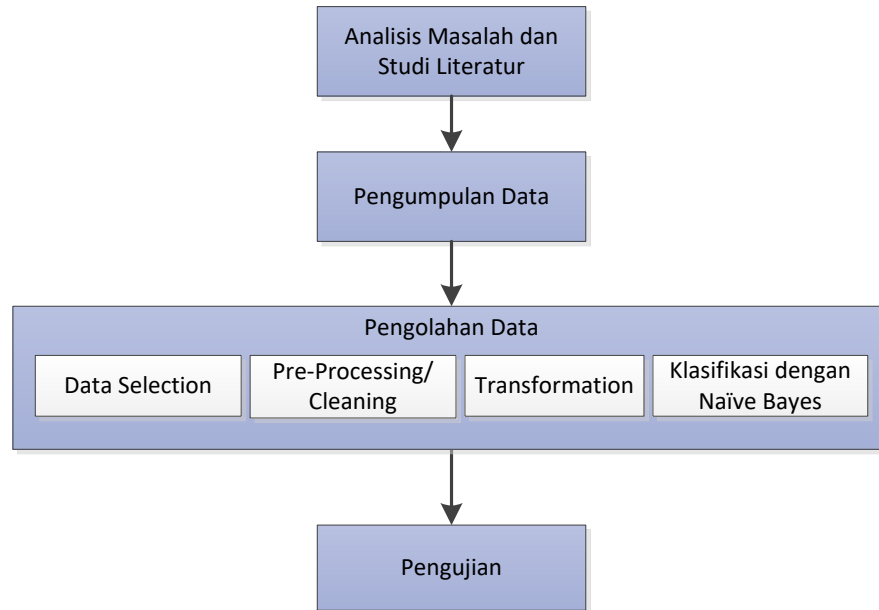
Dimana, variable P merupakan peluang, variable Xi merupakan atribut ke-I, variable xi merupakan nilai pada atribut ke-I, variable Y merupakan kelas yang akan dicari, variable yi merupakan subclass yang akan dicari, μ (mean) merupakan nilai rata-rata dari seluruh atribut dan σ (standar deviasi) merupakan variasi dari seluruh atribut

D. Rapid Miner Tools

Rapid Miner Studio adalah alur kerja desain visual yang digunakan untuk analisis prediktif. Yang menggabungkan pembelajaran mesin dan pengetahuan big data untuk analisis yang lebih baik.. IDE (Integrated Development Environment) memiliki repository, Operator, Views, Global Search, Parameter, help panel, Fungsional Port, dan Process panel dengan tools biasa seperti file, edit[13] Rapid Miner digunakan untuk membuat model dan rencana dengan seperangkat alat dan fungsi yang kuat sehingga pengguna memiliki kemungkinan untuk melepaskan atau mengekstrak statistik dan informasi penting. Pengguna sepenuhnya diizinkan untuk memanfaatkan kemampuan eksplorasi data dan statistik deskriptif seperti grafik dan visualisasi[14].

2. METODE

Pada bab ini akan menjelaskan tentang metode penelitian yang dipakai untuk menjelaskan langkah-langkah dalam penelitian ini. Adapun metode penelitian tersebut dapat digambarkan melalui kerangka kerja penelitian yang dapat dilihat pada gambar berikut:



Gambar 2. Kerangka Kerja Penelitian

Dari kerangka kerja penelitian yang terdapat pada gambar 2 diatas, dapat diuraikan sebagai berikut:

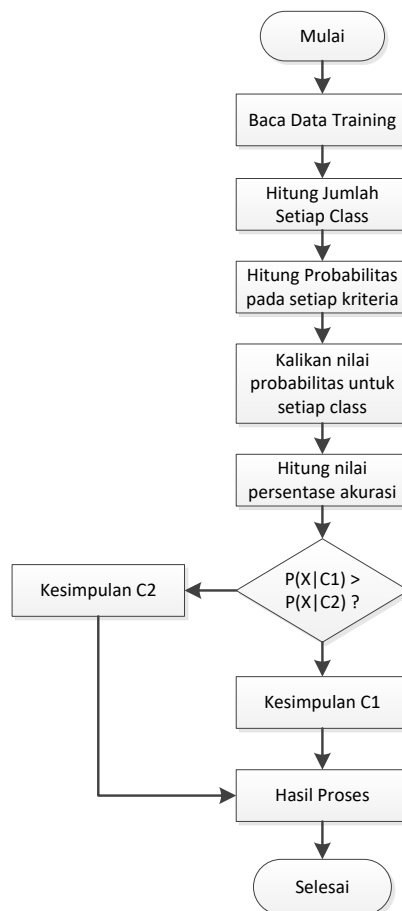
- 1) **Analisis Masalah dan Studi Literatur**
 Pada tahap ini merupakan langkah awal yang digunakan dalam merumuskan masalah. Selanjutnya dilakukan analisa dari permasalahan yang ada untuk mengetahui bagaimana cara menyelesaikan masalah tersebut. Kemudian langkah selanjutnya adalah mempelajari teori-teori yang berkaitan dengan penelitian dimana teori-teori dapat diperoleh dari berbagai macam literature baik nasional maupun internasional.
- 2) **Pengumpulan Data**
 Tahap ini dilakukan untuk mendapatkan data yang berkaitan dengan penelitian dilakukan diskusi dan pengambilan data contoh
- 3) **Pengolahan Data**
 Pada tahap pengolahan data mining, ada beberapa proses yang dilakukan antara lain Data Selection, Preprocessing/Cleaning, Transformation, dan Data mining dengan metode klasifikasi naïve bayes. Pada Data Selection, dilakukan seleksi data yang sesuai dan dibutuhkan dalam melakukan analisa. Selanjutnya Preprocessing/Cleaning dilakukan pembersihan data apabila terdapat data yang tidak sesuai, noise dan data yang tidak konsisten serta menghilangkan duplikasi data. Selanjutnya proses Transformation, dilakukan perubahan data sesuai dengan data yang dibutuhkan. Dan terakhir pada proses Data Mining, dimana pada penelitian ini metode klasifikasi data mining yang digunakan adalah naïve bayes, yang digunakan untuk proses analisa menggunakan dan mencari solusi dari permasalahan yang sudah diuraikan sebelumnya.
- 4) **Pengujian**
 Tahap akhir pada sistem ini adalah pengujian, dimana pengujian yang dilakukan terhadap dataset menggunakan bantuan tools Rapid Miner yang diharapkan dapat memberikan validasi dari algoritma naïve bayes yang digunakan.

3. HASIL DAN PEMBAHASAN

Pada bab ini akan dibahas pengolahan data dengan algoritma naïve bayes yang kemudian dilakukan pengujian sistem dengan tool Rapid Miner untuk melihat nilai akurasi dari algoritma yang digunakan.

A. Penerapan Algoritma Naïve Bayes

Untuk proses pengolahan data dibutuhkan dataset yang akan digunakan untuk dilakukan pengolahan dan melakukan analisa dari metode yang digunakan, dimana data tersebut dibagi menjadi dua bagian yaitu data latih dan juga data uji. Data tersebut digunakan untuk proses perhitungan sehingga menghasilkan informasi baru dan untuk menguji tingkat akurasi dari metode naïve bayes maka dibutuhkan alat bantu berupa tools yang akan digunakan untuk proses pengujian akurasi klasifikasi yaitu dengan bantuan software RapidMiner. Untuk melihat gambaran proses penerapan metode naïve bayes dapat dilihat pada flowchart dari gambar 3 berikut



Gambar 3. Flowchart Naïve Bayes

Dari diagram alir tersebut, proses perhitungan Algoritma Naive Bayes menggunakan data latih dan data testing sebagai data uji. Berikut merupakan data training yang dapat dilihat pada tabel 1.

Tabel 1. Data Training

ID ATM	CIB %	Lokasi	Status ATM	Restock	CashOut	EndCash	Status Pengisian
S1AJSM08AG	Kecil	Dekat	Baik	Sedang	Banyak	Sedikit	Isi
S1AKIM08AL	Besar	Sedang	Baik	Banyak	Banyak	Banyak	Tidak Isi
S1AKJH01OC	Kecil	Sedang	Baik	Sedang	Sedang	Sedang	Isi
S1AKJH01OD	Besar	Jauh	Baik	Banyak	Sedang	Banyak	Tidak Isi
S1AKJH06AC	Sedang	Jauh	Baik	Sedang	Banyak	Sedikit	Isi
S1AKJH08AJ	Besar	Jauh	Perbaikan	Sedang	Banyak	Sedang	Tidak Isi
....							
....							
....							
S1AUSU09YN	Sedang	Sedang	Baik	Sedang	Banyak	Sedikit	Isi
S1AUSU09YQ	Sedang	Jauh	Baik	Sedang	Banyak	Sedikit	Isi
S1AUSU11YL	Besar	Sedang	Baik	Sedang	Banyak	Sedang	Tidak Isi
S1AUSUA028	Besar	Dekat	Baik	Sedang	Sedang	Banyak	Tidak Isi

Dari data training yang terdapat pada tabel 1 diatas, akan dilakukan pengujian menggunakan data testing sebagai berikut, yaitu:

Tabel 2. Data Testing

ID ATM	CIB %	Lokasi	Status ATM	Restock	CashOut	EndCash	Status Pengisian
S1IUSUA017	Sedang	Sedang	Baik	Banyak	Banyak	Banyak	?

Adapun kategori atau kriteria yang digunakan dalam penelitian ini adalah:

1. Lokasi ATM, untuk variabel Lokasi ATM diklasifikasikan dalam 3 kategori yaitu, dekat, sedang dan jauh
2. Status ATM, untuk variable status ATM diklasifikasikan menjadi 2 kategori yaitu, Baik dan perbaikan
3. Restock (Saldo Awal), variable restock diklasifikasikan menjadi 3 kategori yaitu, Sedikit, Sedang dan Banyak
4. Cashout, varibel Cashout diklasifikasikan menjadi 3 kategori yaitu, Sedikit, Sedang dan Banyak
5. EndCash, varibael Endcash diklasifikasikan menjadi 3 kategori yaitu, Sedikit, Sedang dan Banyak
6. CIB (Saldo Akhir), variable CIB diklasifikasikan menjadi 3 kategori yaitu, Kecil, Sdang dan Besar

Untuk hasil klasifikasi dapat dilakukan perhitungan sebagai berikut:

Menghitung Probabilitas pada setiap kriteria

1. CIB
 $P(X1 = \text{Sedang} \mid \text{Klasifikasi} = \text{Isi})$
 $P(X1 = 10/22) = 0,45$
 $P(X1 = \text{Sedang} \mid \text{Klasifikasi} = \text{Tidak Isi})$
 $P(X1 = 2/33) = .0,06$
2. Lokasi
 $P(X2 = \text{Sedang} \mid \text{Klasifikasi} = \text{Isi})$
 $P(X2 = 7/22) = 0,32$
 $P(X2 = \text{Sedang} \mid \text{Klasifikasi} = \text{Tidak Isi})$
 $P(X2 = 13/33) = 0,39$
3. Status ATM
 $P(X3 = \text{Baik} \mid \text{Klasifikasi} = \text{Isi})$
 $P(X3 = 17/22) = 0,77$
 $P(X3 = \text{Baik} \mid \text{Klasifikasi} = \text{Tidak Isi})$
 $P(X3 = 25/33) = .0,76$
4. Restock
 $P(X4 = \text{Banyak} \mid \text{Klasifikasi} = \text{Isi})$
 $P(X4 = 5/22) = 0,23$
 $P(X4 = \text{Banyak} \mid \text{Klasifikasi} = \text{Tidak Isi})$
 $P(X4 = 6/33) = .0,18$
5. CashOut
 $P(X5 = \text{Banyak} \mid \text{Klasifikasi} = \text{Isi})$
 $P(X5 = 18/22) = 0,82$
 $P(X5 = \text{Banyak} \mid \text{Klasifikasi} = \text{Tidak Isi})$
 $P(X5 = 14/33) = .0,42$
6. EndCash
 $P(X6 = \text{Banyak} \mid \text{Klasifikasi} = \text{Isi})$
 $P(X6 = 4/22) = 0,18$
 $P(X6 = \text{Banyak} \mid \text{Klasifikasi} = \text{Tidak Isi})$
 $P(X6 = 12/33) = .0,36$

Menghitung nilai probabilitas setiap class

1. $P(C1) = \frac{\text{Classstatuspengisian} = \text{Isi}}{\text{Jumlah data}}$
 $P(\text{Isi}) = 22/55 = 0,4$
2. $P(C2) = \frac{\text{Classstatuspengisian} = \text{Tidak Isi}}{\text{Jumlah data}}$
 $P(\text{Tidak Isi}) = 33/55 = 0,6$

Menghitung semua nilai dari setiap klasifikasi Isi dan Tidak Isi

1. $P(X \mid \text{Klasifikasi} = \text{Isi})$
 $= 0,45 * 0,32 * 0,77 * 0,23 * 0,82 * 0,18$

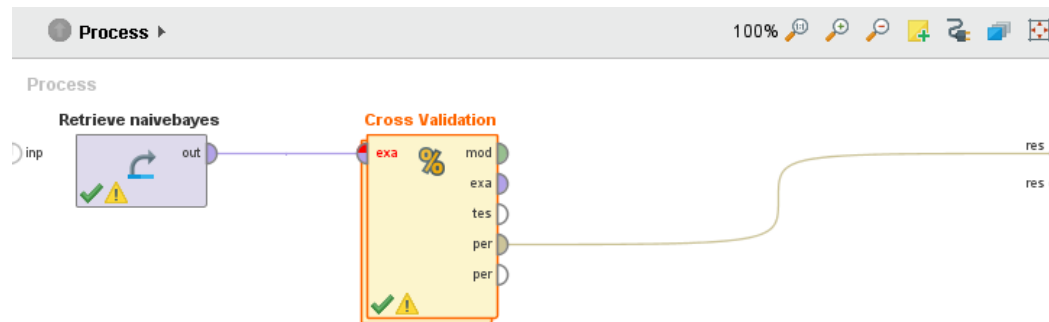
$$= 0,004$$

$$\begin{aligned} 2. \quad & P(X \mid \text{Klasifikasi} = \text{Tidak Isi}) \\ &= 0,06 * 0,39 * 0,76 * 0,18 * 0,42 * 0,36 \\ &= 0,001 \end{aligned}$$

Dari hasil perhitungan akhir dari masing-masing klasifikasi, maka probabilitas $P(X \mid \text{Klasifikasi} = \text{Isi})$ lebih tinggi dari $P(X \mid \text{Klasifikasi} = \text{Tidak Isi})$ yaitu 0,004 banding 0,001, sehingga dapat disimpulkan bahwa status pengisian ATM tersebut adalah Isi.

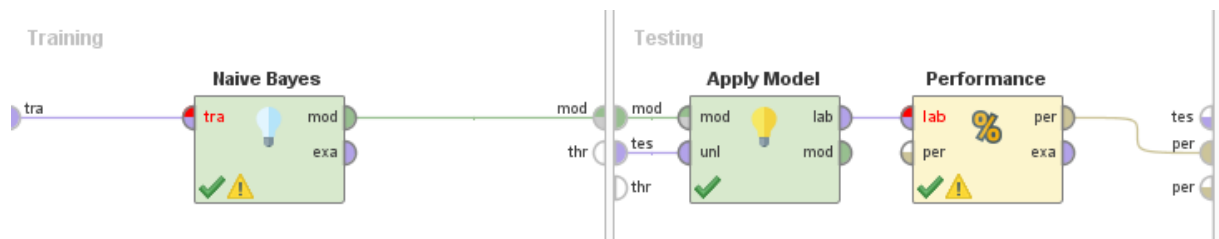
B. Pengujian Algoritma Naïve Bayes dengan RapidMiner

Pada tahap ini dilakukan pengujian menggunakan bantuan software Rapid Miner untuk mengetahui nilai akurasi, recall dan precision pada algoritma naïve bayes. Rapid Miner merupakan software yang digunakan sebagai ide pemecah masalah untuk melakukan analisa terhadap data mining, analisis prediksi dan text mining. Dengan menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik[15]. Untuk melakukan pengolahan data dengan Rapid Miner pada algoritma naïve bayes dapat dilihat pada gambar 4 berikut.



Gambar 4. Model Validasi Naïve Bayes

Data input pada gambar 4 diatas disimpan dalam format excel (.xls) dan ada juga set data pelatihan yang diberikan agar sistem dapat belajar dari data yang diberikan



Gambar 5. Pengujian Validasi Naïve Bayes

Pada gambar 5 merupakan uji validasi dari algoritma Naïve Bayes dalam rapidminer, dimana operator dan parameter yang digunakan akan berpengaruh terhadap akurasi dan model yang terbentuk.

accuracy: 90.67% +/- 12.65% (micro average: 90.91%)

	true Isi	true Tidak Isi	class precision
pred. Isi	18	1	94.74%
pred. Tidak Isi	4	32	88.89%
class recall	81.82%	96.97%	

Gambar 6. Performance Vector Algoritma Naïve Bayes

Dari hasil pengujian yang telah dilakukan dapat dilihat pada gambar 6 diatas, dimana nilai accuracy yaitu 90,91% class precision yaitu Isi 94,74%, tidak isi 88,89%, dan untuk class recall yaitu Isi 81,82% sedangkan tidak isi 96,97%.

Secara Umum precision, recall, dan accuracy dapat dirumuskan sebagai berikut:

- 1) Nilai precision dihasilkan dari jumlah data benar yang bernilai positif (TP) dibagi dengan jumlah data benar yang bernilai positif (TP) ditambah data salah yang bernilai positif (FP)
- 2) Nilai recall dihasilkan dari data benar yang bernilai positive (TP) dengan hasil penjumlahan dari data benar yang bernilai positif (TP) dijumlah dengan data salah yang bernilai negatif (FN).
- 3) Nilai akurasi dihasilkan dari jumlah TP + TN dibagi jumlah total data testing yang diuji

Untuk menghitung nilai recall dan precision pada klasifikasi Isi sebagai berikut :

1. Class Isi

$$Recall = \frac{TP}{TP + FN} * 100\% = \frac{18}{18 + 4} = 0,8182 = 81,82\%$$

$$Precision = \frac{TP}{TP + FP} * 100\% = \frac{18}{18 + 1} = 0,9474 = 94,74\%$$

2. Class Tidak Isi

$$Recall = \frac{TP}{TP + FN} * 100\% = \frac{32}{32 + 1} = 0,9697 = 96,97\%$$

$$Precision = \frac{TP}{TP + FP} * 100\% = \frac{32}{32 + 4} = 0,8889 = 88,89\%$$

3. Tingkat Akurasi

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Accuracy = \frac{18 + 32}{18 + 32 + 1 + 4}$$

$$Accuracy = \frac{50}{55} = 90,91\%$$

Berdasarkan hasil yang telah diuraikan diatas bahwa algortima naïve bayes telah mengklasifikasikan sebanyak 50 data yang berhasil diuji dengan benar dari 55 data diuji dengan persentase keakuratan sebesar 90,91% dan terdapat 5 data yang tidak berhasil diuji dengan benar

4. KESIMPULAN

Bahwa hasil akhir yang diperoleh menggunakan data uji sebanyak 55 data dengan bantuan software Rapid Miner dapat diketahui bahwa Berdasarkan hasil pengujian yang dilakukan menggunakan pengklasifikasian algoritma naive bayes

1. Algoritma Naïve Bayes mampu dalam menganalisa terhadap data pengisian ATM dengan berhasil mengklasifikasi 50 data dari 55 data yang diuji dengan nilai probabilitas sebesar 0,004 untuk klasifikasi Isi dan persentase keakuratan sebesar 90,91%
2. Nilai akurasi keseluruhan yaitu 90,91% dan class precision Isi yaitu 94,74% untuk precision tidak isi 88,89%, sedangkan untuk class recall Isi yaitu 81,82% dan class recall tidak isi 96,97%.

Adapun saran yang dapat diberikan dari penelitian ini sebagai acuan bagi peneliti selanjutnya adalah melakukan penelitian berikutnya dengan menggunakan algoritma klasifikasi yang berbeda dan bisa juga dilakukan perbandingan dari masing-masing metode klasifikasi tersebut, untuk dilihat keakurasian algoritma mana yang paling terbaik.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Universitas Potensi Utama yang telah banyak memberikan dukungan dalam penelitian ini.

REFERENSI

- [1] K. Ahmed and T. Jesmin, "Comparative Analysis of Data Mining Classification Algorithms in Type-2 Diabetes Prediction Data Using WEKA Approach," *Int. J. Sci. Eng.*, vol. 7, no. 2, pp. 155–160, 2014, doi: 10.12777/ijse.7.2.150-154.
- [2] A. Saleh, "Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga," *Creat. Inf. Technol. J.*, vol. 2, no. 3, pp. 207–217, 2015.
- [3] M. H. Rifqo, A. Wijaya, and J. Pseudocode, "IMPLEMENTASI ALGORITMA NAIVE BAYES DALAM PENENTUAN PEMBERIAN KREDIT," vol. IV, no. September, pp. 120–128, 2017.
- [4] S. Hendrian, "Algoritma Klasifikasi Data Mining Untuk Memprediksi Siswa Dalam Memperoleh Bantuan Dana Pendidikan," *Fakt. Exacta*, vol. 11, no. 3, pp. 266–274, 2018, doi: 10.30998/faktorexacta.v11i3.2777.
- [5] K. Adhatrao, A. Gaykar, A. Dhawan, R. Jha, and V. Honrao, "Predicting Students' Performance Using ID3 and C4.5 Classification Algorithms," *Int. J. Data Min. Knowl. Manag. Process*, vol. 3, no. 5, pp. 39–52, 2013, doi: 10.5121/ijdkp.2013.3504.
- [6] M. Yusa, E. Utami, and E. Luthfi. Taufiq, "Evaluasi Performa Algoritma Klasifikasi Decision Tree ID3, C4.5, dan CART Pada Dataset Readmisi Pasien Diabetes," *Infosys (Information Syst. J.)*, vol. 4, no. 1, pp. 23–34, 2016.
- [7] D. Singh, N. Choudhary, and J. Samota, "Analysis of Data Mining Classification with Decision tree Technique," *Glob. J. Comput. Sci. Technol.*, vol. 13, no. 13, 2013.
- [8] D. Y. H. Tanjung, "Analisis perbandingan algoritma id3 dan c4.5 terhadap data pengisian uang atm," *CSRID*, pp. 231–242, doi: <http://dx.doi.org/10.22303/csr.id.13.3a.2021.231-242>.
- [9] K. Sumathi, S. Kannan, and K. Nagarajan, "Data Mining: Analysis of student database using Classification Techniques," *Int. J. Comput. Appl.*, vol. 141, no. 8, pp. 22–27, 2016, doi: 10.5120/ijca2016909703.

-
- [10] B. Budiman, R. Nursyanti, R. Y. R. Alamsyah, and I. Akbar, "Data Mining Implementation Using Naïve Bayes Algorithm and Decision Tree J48 In Determining Concentration Selection," *Int. J. Quant. Res. Model.*, vol. 1, no. 3, pp. 123–134, 2020, doi: 10.46336/ijqrm.v1i3.72.
 - [11] M. F. A. Saputra, T. Widiyaningtyas, and A. P. Wibawa, "Illiteracy classification using K means-naïve bayes algorithm," *Int. J. Informatics Vis.*, vol. 2, no. 3, pp. 153–158, 2018, doi: 10.30630/joiv.2.3.129.
 - [12] F. Paquin, J. Rivnay, A. Salleo, N. Stingelin, and C. Silva, "Multi-phase semicrystalline microstructures drive exciton dissociation in neat plastic semiconductors," *J. Mater. Chem. C*, vol. 3, no. January 2019, pp. 10715–10722, 2015, doi: 10.1039/b000000x.
 - [13] B. Devipriya and Y. Kalpana, "Evaluation of sentiment data using classifier model in rapid miner tool," *Int. J. Eng. Adv. Technol.*, vol. 9, no. 1, pp. 2966–2972, 2019, doi: 10.35940/ijeat.A1323.109119.
 - [14] A. Naik and L. Samant, "Correlation Review of Classification Algorithm Using Data Mining Tool: WEKA, Rapidminer, Tanagra, Orange and Knime," *Procedia Comput. Sci.*, vol. 85, no. Cms, pp. 662–668, 2016, doi: 10.1016/j.procs.2016.05.251.
 - [15] N. Riyanah, "Penerapan Algoritma Naive Bayes Untuk Klasifikasi Penerima Bantuan Surat Keterangan Tidak Mampu (Implementation of Algorithms Naïve Bayes for Classification Recipients Help Letter Description Not Able)," vol. 2, no. 4, pp. 206–213, 2021.