

Penerapan Data Mining Dengan Algoritma *Naive Bayes Classifier* Dalam Memprediksi Pembelian Cat

Fitriana Harahap¹, Nidia Enjelita Saragih², Elida Tuti Siregar³, Husin Sariangseh⁴

¹Universitas Potensi Utama, Jl. Yos Sudarso KM 6,5 No.3A Tanjung Mulia, Medan 20241, Indonesia

²Universitas Potensi Utama, Jl. Yos Sudarso KM 6,5 No.3A Tanjung Mulia, Medan 20241, Indonesia

³Universitas Potensi Utama, Jl. Yos Sudarso KM 6,5 No.3A Tanjung Mulia, Medan 20241, Indonesia

INFORMASIARTIKEL

Sejarah Artikel:

DiterimaRedaksi: 29 April 2021

RevisiAkhir: 09 Mei 2021

DiterbitkanOnline: 25 Mei 2021

KATA KUNCI

Data Mining

Purchase of paint

Naive Bayes.

KORESPONDENSI

E-mail: nidia.1924@gmail.com

A B S T R A C T

Companies need several types of communication technology that can predict customer purchase interest, the goal is that the company can properly consider product sales and determine the company's paint product supply. So far, the decision of the Home Smart sales manager has been made by looking at the closeness of the supplier relationship and how many sponsors are funding the company. So that sometimes the product cannot compete with other companies. The Naive Bayes classifier algorithm is one of the algorithms included in the classification technology. The application of the Naive Bayes method is expected to predict paint purchases from suppliers. From 60 paint purchase data tested with the Naive Bayes method, the results reached 80% of the accuracy of the predictions. Of the 60 tested paint purchase data, 48 paint purchase data were successfully classified correctly.

1. PENDAHULUAN

Persaingan dalam dunia bisnis mendorong para pebisnis untuk terus melakukan terobosan demi kelangsungan usahanya. Saat ini Home Smart Medan telah menyediakan berbagai macam jenis pelapis dinding dengan berbagai macam merk, tipe dan harga, namun belum diketahui minat konsumen terhadap pelapis dinding. Manajer pembelian membuat keputusan untuk membeli cat melalui pemasok dengan melihat kedekatan hubungan pemasok dan besar biaya sponsor yang diberikan untuk perusahaan. Terkadang harga, jenis, dan merk cat tidak menjadi prioritas perusahaan sehingga produk cat kalah bersaing dengan perusahaan lain. Pembelian produk cat yang kurang efektif menyebabkan produk, menjadi kurang menarik bagi pelanggan. Tentu saja, ini akan membuat konsumen beralih ke perusahaan serupa yang menggunakan pelapis dinding yang menurut mereka paling mereka butuhkan. Dalam hal ini, untuk menentukan cat yang dibeli oleh pelanggan, metode *nave bayes* digunakan untuk klasifikasi data mining untuk memprediksi volume pembelian cat.

Metode klasifikasi merupakan proses menemukan fungsi konsep atau model atau kategori data yang menggambarkan tujuan Fitriana Harahap

dalam memperkirakan kelas objek yang labelnya tidak diketahui. Model itu sendiri dapat berupa aturan "jika-maka", pohon keputusan, atau rumus matematika [1][2].

Algoritma pengklasifikasi *Naive Bayes* merupakan proses klasifikasi probabilistik sederhana berdasarkan teorema Bayes (atau hukum Bayes) yang memiliki independensi yang kuat, dengan kata lain dalam nave Bayesian, model yang digunakan adalah model fitur independen. Dalam Bayesian (terutama *Naive Bayes*), kuatnya independensi elemen mengacu pada elemen dalam data, yang tidak ada hubungannya dengan apakah ada elemen lain dalam data yang sama. Keuntungan dari klasifikasi adalah bahwa hanya sejumlah kecil data pelatihan yang diperlukan untuk memperkirakan parameter (*mean* dan *varians variabel*) yang diperlukan untuk klasifikasi[3].

Aplikasi yang mengklasifikasikan dalam bentuk algoritma pengklasifikasi *nave bayes* diharapkan dapat memprediksi atau mendeteksi lebih awal minat beli calon pelanggan berdasarkan transaksi masa lalu, sehingga perusahaan dapat lebih memahami cat yang diminati pelanggan mereka.

Dalam penelitian ini dokumen yang digunakan adalah jurnal-jurnal berikut Penelitian yang dilakukan [4] menyimpulkan

bahwa berdasarkan data keluarga sebagai data latih, metode naive bayes berhasil mengklasifikasikan 47 dari 60 data uji. data. Dengan demikian metode Naive Bayesian berhasil menentukan prediksi konsumsi listrik pada rumah tangga dengan akurasi sebesar 78.3333%.

Penelitian yang dilakukan [5] Hasil analisis sentimen pelanggan toko online A, B, C, D, E adalah: variabel harga (VH) sebesar 0,00000023, variabel produk (VP) 0,0000049, variabel transaksi kemudahan penggunaan (VKT) Adalah 0,0000048, keamanan transaksi variabel (VKMT) adalah 0,0000038, dan kepercayaan variabel (VK) adalah 0,000015.

Penelitian yang dilakukan [6], kesimpulan dari penelitian adalah hasil tes yang menunjukkan bahwa faktor yang memiliki pengaruh paling besar dalam menentukan klasifikasi prestasi akademik mahasiswa antara lain indeks prestasi kumulatif (IPK), indeks prestasi semester pertama (IPK), IPK semester keempat dan jenis kelamin. Karena itu, faktor-faktor tersebut layak menjadi bahan evaluasi pengelolaan perguruan tinggi. Nilai akurasi, recall, dan akurasi yang diperoleh dari algoritma NBC masing-masing adalah 83%, 50%, dan 70% dengan pengujian data siswa kelas 2005-2009.

Berdasarkan analisis penelitian terdahulu Algoritma *Naive Bayes* dari teknik pengklasifikasian dapat membantu dalam memprediksi pembelian cat dari suplier berdasarkan data penjualan cat pada perusahaan.

2. TINJAUAN PUSTAKA

2.1 Data Mining

Data mining adalah proses analisis yang bertujuan untuk memeriksa data dalam jumlah besar untuk menemukan pengetahuan tersembunyi yang berharga dan konsisten. Tujuan dari data mining adalah untuk menemukan tren atau pola yang dibutuhkan dalam database besar untuk membantu pengambilan keputusan di masa depan [7][8][9].

2.2. Naive Bayes Classifier

Klasifikasi *Naive Bayes* pertama kali diusulkan oleh ilmuwan Inggris Thomas Bayes menggunakan metode probabilitas dan statistik. Algoritma ini menghasilkan teknik prediksi berbasis probabilitas sederhana dengan menggunakan teorema Bayes (atau Bayes). Hal ini berarti, pada *Naive Bayes*, model yang digunakan merupakan “model fitur independen”. Dalam Bayesian (terutama *Naive Bayes*), independensi kuat dari suatu elemen berarti bahwa elemen dalam data tidak ada hubungannya dengan apakah ada elemen lain dalam data yang sama [10][11].

Adapun bentuk umum dari Teorema Bayes dapat dilihat pada persamaan berikut:

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

X = Data dengan class yang belum diketahui

H = Hipotesis data X merupakan suatu class spesifik

$P(H|X)$ = Probabilitas hipotesis H berdasarkan kondisi x (posteriori prob.)

$P(H)$ = Probabilitas hipotesis H (prior prob.)

$P(X|H)$ = Probabilitas X berdasarkan kondisi tersebut

$P(X)$ = Probabilitas dari X

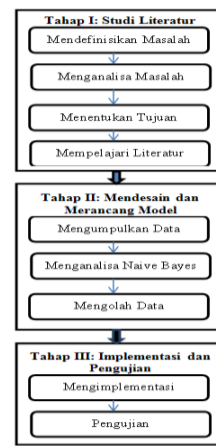
2.3. Klasifikasi (*classification*)

Klasifikasi didefinisikan sebagai fungsi pembelajaran yang dapat memetakan (mengkategorikan) elemen data (item) menjadi salah satu dari beberapa kategori yang telah ditentukan. Contoh lain dari klasifikasi bisnis dan penelitian adalah [1]: Satu. Tentukan apakah transaksi kartu kredit itu curang. b. Diperkirakan apakah pengajuan KPR nasabah baik atau buruk. C. Diagnosis penyakit pada pasien untuk memahami penyakit yang mana.

3. METODOLOGI

3.1 Kerangka kerja Penelitian

Pembahasan mengenai tahapan-tahapan yang harus dilakukan dalam menyelesaikan masalah tercakup dalam bagian kerangka kerja penelitian. Kerangka kerja penelitian ini dapat dilihat pada gambar 3:



Gambar 3. Kerangka Kerja Penelitian

1. Mendefinisikan Masalah
ini merupakan tahap pertama dari kerangka kerja, dimana sistem akan dipelajari dan direview untuk mengamati sistem secara lebih mendalam, serta menggali masalah yang muncul pada sistem tersebut. Tahap mendefinisikan masalah merupakan langkah awal dalam menemukan rumusan pertanyaan penelitian.
2. Menganalisa Masalah
Kemudian masalah yang ditemukan pada tahapan sebelumnya, akan dianalisis. Tahapan ini bertujuan untuk memahami masalah yang telah ditentukan sebelumnya. Tahapan analisa masalah diharapkan membantu peneliti dalam memahami masalah dengan lebih baik.
3. Menentukan Tujuan
Langkah selanjutnya setelah analisa masalah yaitu menentukan tujuan dari penelitian. Dalam hal ini, tujuan yang menjadi sasaran adalah untuk menyelesaikan permasalahan yang ada.
4. Mempelajari Literatur
Tahapan ini dilakukan dalam rangka melengkapi kosakata aturan, teori pendukung pemecahan masalah, serta konsep yang digunakan dalam penelitian ini. Tahapan penelitian ini bertujuan untuk mengumpulkan semua data yang dibutuhkan untuk penelitian.

3.2. Tahapan Mendesain dan Merancang Model Sistem

1. Kumpulkan data Kumpulkan data untuk melatih dan menguji *Naive Bayes*. Banyaknya data yang berhasil dikumpulkan, menentukan keberhasilan penyelesaian masalah tersebut.
2. Analisis dengan algoritma *Naive Bayes*. Data yang terkumpul kemudian akan dianalisis, untuk menyesuaikan pengolahan data dengan menggunakan metode *Naive Bayes*.
3. Mengolah data. Data yang telah dianalisis akan diolah dengan menggunakan metode *Naive Bayes*.

3.3. Tahapan Implementasi

1. Implementasi. Menurut pemrosesan data, tahap realisasi adalah tentang bagaimana mengimplementasikan pemrosesan data di alat.
2. Pengujian. Tahap ini akan menjadi penentu apakah perangkat lunak yang telah dikembangkan berhasil memenuhi tujuan yang diharapkan.

Alur dari metode *Naive Bayes* dapat dilihat pada gambar 4:



Gambar 4. Alur Metode Naive Bayes

4. HASIL DAN PEMBAHASAN

4.1 Penerapan Metode Naive Bayes

Naive Bayes didasarkan pada asumsi yang disederhanakan bahwa nilai atribut tidak bergantung secara kondisional satu sama lain pada nilai keluaran tertentu. Dengan kata lain, mengingat nilai output, probabilitas pengamatan kolektif merupakan produk dari probabilitas individu. Adapun yang menjadi keuntungan dari penggunaan metode *Naive Bayes* adalah bahwa metode ini hanya memerlukan sedikit data pelatihan untuk menentukan estimasi parameter yang diperlukan dalam proses klasifikasi. Dalam metode *naive bayes*, string data konstan dibedakan dari data numerik kontinu. Perbedaan ini terlihat saat menentukan nilai probabilitas setiap kondisi, baik kondisi dengan nilai data string maupun kondisi dengan nilai data numerik. Penerapan metode *naive bayes* adalah sebagai berikut.

4.2. Data Training

Naive Bayes didasarkan pada asumsi, dalam penentuan data yang akan dianalisis menggunakan metode *naive bayes*, tahapan pertama yang harus dikerjakan adalah membaca data latih. Adapun data latih yang digunakan dapat dilihat pada Tabel 1 berikut.

Tabel 1. Data Training

No	Kualitas	Animo	Harga	Kompetisi	Kooperasi	Beli
----	----------	-------	-------	-----------	-----------	------

1	Super	Tinggi	Mahal	Tinggi	Sangat Baik	Ya
2	Super	Tinggi	Mahal	Tinggi	Baik	Ya
3	Super	Tinggi	Mahal	Tinggi	Kurang Baik	Tidak
4	Super	Tinggi	Mahal	Sedang	Sangat Baik	Ya
5	Super	Tinggi	Mahal	Sedang	Baik	Ya
6	Super	Tinggi	Mahal	Sedang	Kurang Baik	Tidak
7	Super	Tinggi	Mahal	Rendah	Sangat Baik	Ya
8	Super	Tinggi	Mahal	Rendah	Baik	Ya
9	Super	Tinggi	Mahal	Rendah	Kurang Baik	Tidak
10	Super	Tinggi	Normal	Tinggi	Sangat Baik	Ya
11	Super	Tinggi	Normal	Tinggi	Baik	Ya
12	Super	Tinggi	Normal	Tinggi	Kurang Baik	Tidak
13	Super	Tinggi	Normal	Sedang	Sangat Baik	Ya
14	Super	Tinggi	Normal	Sedang	Baik	Ya
15	Super	Tinggi	Normal	Sedang	Kurang Baik	Ya
60	Super	Rendah	Mahal	Sedang	Kurang Baik	Tidak

4.3 Probabilitas Kriteria Kualitas

Berdasarkan data pembelian cat pada tabel di atas, diketahui bahwa jumlah data latih adalah sebanyak 60 data. Dari 60 data tersebut, terdapat 15 data cat dengan kualitas super dan pembelian cat ya, 7 data cat dengan kualitas super dan pembelian tidak, 5 data cat dengan kualitas medium dan pembelian ya. 6 data cat dengan kualitas medium dan pembelian tidak. 14 data cat dengan kualitas standart dan pembelian ya, 13 data cat dengan kualitas standart dan pembelian tidak.

Tabel 2. Probabilitas Kriteria Kualitas

Animo	Jumlah Kejadian "Pembelian Cat"		Probabilitas	
	Ya	Tidak	Ya	Tidak
Tinggi	27	18	0.79	0.69
Sedang	4	3	0.12	0.12
Rendah	3	5	0.09	0.19
Jumlah	34	26	0.57	0.43

4.4 Probabilitas Kriteria Animo

Berdasarkan data pembelian cat pada tabel data training(data latih), diketahui bahwa jumlah data latih adalah sebanyak 60 data. Dari 60 data tersebut terdapat 27 data cat dengan Animo Tinggi dan pembelian cat ya, 18 data cat dengan Animo Tinggi dan pembelian tidak, 4 data cat dengan Animo Sedang dan pembelian ya. 3 data cat dengan Animo Sedang dan pembelian tidak. 3 data cat dengan Animo Rendah dan pembelian ya, 5 data cat dengan Animo Rendah dan pembelian tidak.

Tabel 3. Probabilitas Kriteria Animo

Animo	Jumlah Kejadian "Pembelian Cat"		Probabilitas	
	Ya	Tidak	Ya	Tidak
Tinggi	27	18	0.79	0.69
Sedang	4	3	0.12	0.12
Rendah	3	5	0.09	0.19
Jumlah	34	26	0.57	0.43

4.5 Probabilitas Kriteria Harga

Berdasarkan data pembelian cat pada tabel data training (data latih) di atas, diketahui bahwa jumlah data latih adalah sebanyak 60 data. Dari 60 data tersebut terdapat 16 data cat dengan harga mahal dan pembelian cat ya, 14 data cat dengan harga mahal dan pembelian tidak, 12 data cat dengan harga normal dan pembelian ya. 9 data cat dengan harga normal dan pembelian tidak. 6 data cat dengan harga terjangkau dan pembelian ya, 3 data cat dengan harga terjangkau dan pembelian tidak.

Tabel 4. Probabilitas Kriteria Harga

Harga	Jumlah Kejadian "Pembelian Cat"		Probabilitas	
	Ya	Tidak	Ya	Tidak
Mahal	16	14	0.47	0.54
Normal	12	9	0.35	0.35
Terjangkau	6	3	0.18	0.12
Jumlah	34	26	0.57	0.43

4.6 Probabilitas Kriteria Kompetisi

Berdasarkan data pembelian cat pada tabel data training (data latih) di atas, diketahui bahwa jumlah data latih (data training) adalah sebanyak 60 data. Dari 60 data tersebut terdapat 14 data cat dengan kompetisi tinggi dan pembelian cat ya, 12 data cat dengan kompetisi tinggi dan pembelian tidak, 15 data cat dengan kompetisi sedang dan pembelian ya. 8 data cat dengan kompetisi sedang dan pembelian tidak. 5 data cat dengan kompetisi rendah dan pembelian ya, 6 data cat dengan kompetisi

	Jumlah Kejadian "Pembelian Cat"		Probabilitas	
	Ya	Tidak	Ya	Tidak
Tinggi	14	14	0.41	0.54
Sedang	15	9	0.44	0.35
Rendah	5	3	0.15	0.12
Jumlah	34	26	0.57	0.43

Tabel 5. Probabilitas Kriteria Kompetisi

4.7 Probabilitas Kriteria Kooperasi

Berdasarkan data pembelian cat pada tabel data training (data latih) di atas, diketahui bahwa jumlah data latih (data training) adalah sebanyak 60 data. Dari 60 data tersebut terdapat 14 data cat dengan Kooperasi sangat baik dan pembelian cat ya, 7 data cat dengan Kooperasi sangat baik dan pembelian tidak, 18 data cat dengan Kooperasi baik dan pembelian ya. 3 data cat

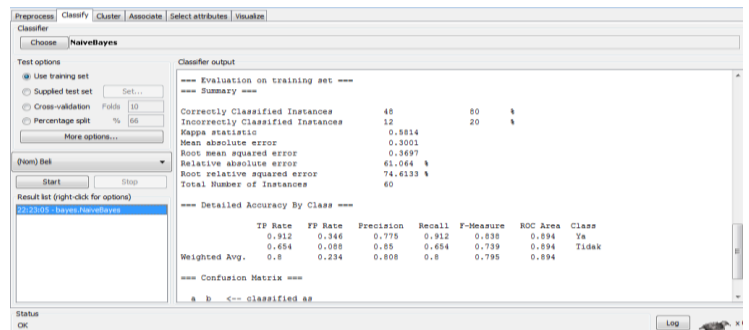
dengan Kooperasi baik dan pembelian tidak. 2 data cat dengan Koopeasi kurang baik dan pembelian ya, 16 data cat dengan Kooperasi kurang baik dan pembelian tidak.

Tabel 6. Probabilitas Kriteria Kooperasi

Kooperasi	Jumlah Kejadian "Pembelian Cat"		Probabilitas	
	Ya	Tidak	Ya	Tidak
Sangat Baik	14	7	0.41	0.27
Baik	18	3	0.53	0.12
Kurang Baik	2	16	0.06	0.62
Jumlah	34	26	0.57	0.43

4.8. Pengujian Metode Naïve Bayes

Berdasarkan nilai probabilitas akan dilakukan pengujian sebanyak 60 data untuk kemudian dicari penyelesaiannya dengan menggunakan tools weka sehingga dihasilkan hasil klasifikasi pembelian cat seperti pada gambar 3:



Gambar 3. Hasil Klasifikasi Metode Naive Bayes

Oleh karena itu, berdasarkan gambar di atas diketahui bahwa persentase untuk Correctly Classified Instance adalah sebesar 80%. Sedangkan persentase untuk Incorrectly Classified Instance adalah 20%. Di mana dari 60 data pembelian cat, ada sebanyak 48 data pembelian cat yang berhasil diklasifikasikan dengan benar dan sebanyak 12 data pembelian cat yang tidak berhasil diklasifikasikan dengan benar.

5. KESIMPULAN DAN SARAN

Berdasarkan penelitian tentang memprediksi pembelian cat dapat disimpulkan beberapa kesimpulan yaitu, Metode Naive Bayes menggunakan data training dalam menghasilkan probabilitas masing-masing kriteria untuk class yang berbeda. Karena itu nilai-nilai probabilitas dari kriteria tersebut dapat dioptimalkan agar bisa digunakan dalam memprediksi pembelian cat berdasarkan proses klasifikasi oleh metode Naive Bayes itu sendiri.

Berdasarkan data pembelian cat yang dijadikan data training, metode Naive Bayes berhasil mengklasifikasikan 48 data dari 60 data yang diuji. Sehingga metode Naive Bayes berhasil memprediksi pembelian cat dengan persentase keakuratan sebesar 80 %.

Adapun saran yang dapat diberikan mengenai penelitian tentang memprediksi pembelian cat oleh konsumen yaitu pengujian lebih baik dilakukan menggunakan metode lain untuk menentukan metode yang lebih akurat untuk memprediksi pembelian cat.

DAFTAR PUSTAKA

[1] D. Nofriansyah, K. Erwansyah, and M. Ramadhan, “Penerapan Data Mining dengan Algoritma Naive Bayes Classifier untuk Mengetahui Minat Beli Pelanggan terhadap Kartu Internet XL (Studi Kasus di CV. Sumber Utama Telekomunikasi),” *J. Saintikom*, vol. 15, no. 2, pp. 81–92, 2016.

[2] A. S. Mubarak, F. T. Informasi, U. Hasyim, and T. Jombang, “Deteksi Minat Beli Pelanggan Terhadap Produk Paket Internet Menggunakan Algoritma Naive Bayes,” vol. 03, pp. 58–63, 2019.

[3] T. I. Andini, W. Witanti, and F. Renaldi, “Prediksi Potensi Pemasaran Produk Baru dengan Metode Naive Bayes Classifier dan Regresi Linear,” *Semin. Nas. Apl. Teknol. Inf.*, pp. 27–32, 2016.

[4] A. Saleh, “Implementasi Metode Klasifikasi Naive Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga,” *Creat. Inf. Technol. J.*, vol. 2, no. 3, pp. 207–217, 2015.

[5] P. D. Atika, “IMPLEMENTASI ALGORITMA NAIVE BAYES CLASSIFIER UNTUK ANALISIS SENTIMEN CUSTOMER PADA TOKO ONLINE,” vol. 12, no. 4, pp. 303–314, 2019, doi: 10.30998/faktorexacta.v12i4.5224.

[6] M. S. Mustafa, M. R. Ramadhan, and A. P. Thenata, “Implementasi Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier,” *Creat. Inf. Technol. J.*, vol. 4, no. 2, p. 151, 2018, doi: 10.24076/citec.2017v4i2.106.

[7] D. F. Pramesti, M. T. Furqon, and C. Dewi, “Implementasi Metode K-Medoids Clustering Untuk Pengelompokan Data Potensi Kebakaran Hutan / Lahan Berdasarkan Persebaran Titik Panas (Hotspot),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 9, pp. 723–732, 2017.

[8] S. Suprianto, “Implementasi Algoritma Naive Bayes Untuk Menentukan Lokasi Strategis Dalam Membuka Usaha Menengah Ke Bawah di Kota Medan (Studi Kasus: Disperindag Kota Medan),” *J. Sist. Komput. dan Inform.*, vol. 1, no. 2, p. 125, 2020, doi: 10.30865/json.v1i2.1939.

[9] R. R. Husin Sariangsah*, Wanayumini, “Penentuan Kelas Menggunakan Algoritma K Medoids Untuk Clustering Siswa Tunagrahita,” *J. Media*, vol. 5, pp. 83–89, 2021, doi: 10.30865/mib.v5i1.2547.

[10] Y. S. Nugroho, “Klasifikasi dan Klastering Penjurusan Siswa SMA Negeri 3 Boyolali,” *Khazanah Inform. J. Ilmu Komput. dan Inform.*, vol. 1, no. 1, p. 1, 2015, doi: 10.23917/khif.v1i1.1175.

[11] H. Annur, “Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes,” *Ilk. J. Ilm.*, vol. 10, no. 2, pp. 160–165, 2018, doi: 10.33096/ilkom.v10i2.303.160-165.

BIODATA PENULIS



Fitriana Harahap
Dosen Universitas Potensi Utama, Jl. Yos Sudarso KM 6,5 No.3A Tanjung Mulia, Medan 20241, Indonesia



Nidia Enjelita Saragih
Dosen Universitas Potensi Utama, Jl. Yos Sudarso KM 6,5 No.3A Tanjung Mulia, Medan 20241, Indonesia



Elida Tuti Siregar
Dosen Universitas Potensi Utama, Jl. Yos Sudarso KM 6,5 No.3A Tanjung Mulia, Medan 20241, Indonesia



Husin Sariangsah
Mahasiswa Program Studi Magister Ilmu Komputer Universitas Potensi Utama, Jl. Yos Sudarso KM 6,5 No.3A Tanjung Mulia, Medan 20241, Indonesia