

APPLICATION OF NAÏVE BAYES CLASSIFIER ALGORITHM IN DETERMINING THE LEVEL OF CUSTOMER SATISFACTION WITH RUMBAI POST OFFICE SERVICES

Anisa Halifa^{*1}, Rice Novita²

^{1,2}Information Systems, Faculty of Science and Technology, UIN Sultan Syarif Kasim Riau, Indonesia
Email: ¹11950321537@students.uin-suska.ac.id, ²rice.novita@uin-suska.ac.id

(Article received: May 18, 2023; Revision: June 04, 2023; published: Desember 23, 2023)

Abstract

Technological advances make service and delivery of goods grow rapidly. Coupled with people's changing shopping habits by shopping online, shipping companies are increasingly needed. PT. POS Indonesia is the first expedition company in Indonesia. Currently PT. POS Indonesia has opened many POS Office branches in every region in Indonesia, one of which is the Rumbai POS Office located in Pekanbaru City. To continue to maintain the company while competing with other expeditions, the Rumbai POS Office must continue to maintain its customers by improving the quality of service. Survey analysis can be done to determine the extent of customer satisfaction with the services provided. To find out the level of customer satisfaction, you can use the classification method. Naive Bayes is a popular and effective machine learning algorithm for classification problems. The study used datasets sourced from the results of questionnaire distribution to customers of the Rumbai POS Office. The questionnaire used 14 indicators derived from the Community Satisfaction Index set by the Ministry of Agriculture in 2004. The classification resulted in a Satisfied class of 16 data with a percentage of 84.2% and a Dissatisfied class of 3 data with a percentage of 15.8%, it can be concluded that the service at the Rumbai POS Office is good. From the classification results, it is proven that the Naive Bayes algorithm is able to predict well the level of customer satisfaction with an accuracy value of 94.74%, precision of 100%, and recall of 94.12%. The results of this research can later be used as information for the Rumbai POS Office to be able to improve service quality.

Keywords: *Classification, Customer Satisfaction, Data Mining, Naive Bayes.*

PENERAPAN ALGORITMA NAÏVE BAYES CLASSIFIER DALAM MENENTUKAN TINGKAT KEPUASAN PELANGGAN TERHADAP PELAYANAN KANTOR POS RUMBAL

Abstrak

Kemajuan teknologi membuat jasa pelayanan dan pengiriman barang berkembang pesat. Ditambah dengan kebiasaan berbelanja masyarakat yang berubah dengan berbelanja *online* membuat perusahaan ekspedisi semakin dibutuhkan. PT. POS Indonesia merupakan perusahaan ekspedisi pertama yang ada di Indonesia. Saat ini PT. POS Indonesia telah membuka banyak cabang Kantor POS yang ada pada setiap daerah di Indonesia, salah satunya Kantor POS Rumbai yang berada di Kota Pekanbaru. Untuk terus mempertahankan perusahaan selama bersaing dengan ekspedisi lain, Kantor POS Rumbai harus tetap mempertahankan pelanggannya dengan meningkatkan kualitas pelayanan. Analisis survei dapat dilakukan untuk mengetahui sejauh mana kepuasan pelanggan terhadap pelayanan yang diberikan. Untuk mengetahui tingkat kepuasan pelanggan dapat menggunakan metode klasifikasi. Naive Bayes merupakan algoritma yang populer dan efektif pada *machine learning* dalam masalah klasifikasi. Penelitian menggunakan *dataset* yang besumber dari hasil penyebaran kuesioner kepada pelanggan Kantor POS Rumbai. Kuesioner menggunakan 14 indikator yang berasal dari Indeks Kepuasan Masyarakat yang telah ditetapkan oleh KepMenPan pada tahun 2004. Klasifikasi menghasilkan kelas Puas sebanyak 16 data dengan perserntase 84,2% dan kelas Tidak Puas sebanyak 3 data dengan persentase 15,8%, dapat disimpulkan bahwa pelayanan pada Kantor POS Rumbai sudah baik. Dari hasil klasifikasi terbukti bahwa algoritma Naive Bayes mampu memprediksi dengan baik tingkat kepuasan pelanggan dengan nilai akurasi sebesar 94,74%, presisi sebesar 100%, dan recall sebesar 94,12%. Hasil dari penelitian ini nantinya dapat dijadikan informasi bagi pihak Kantor POS Rumbai untuk dapat meningkatkan kualitas pelayanan.

Kata kunci: *Data Mining, Kepuasan Pelanggan, Klasifikasi, Naive Bayes.*

1. PENDAHULUAN

Perkembangan jasa pelayanan dan pengiriman terus meningkat dari tahun ke tahun, sehingga menjadi perhatian masyarakat luas. Ditambah lagi dengan maraknya bisnis *online shop* yang bermunculan menyebabkan jasa pengiriman semakin diperlukan. Terdapat beberapa perusahaan ekspedisi di Indonesia seperti, POS Indonesia, TIKI, JNE, J&T, dan masih banyak lagi. PT. POS Indonesia merupakan perusahaan ekspedisi tertua yang ada di Indonesia, dan termasuk bagian dari Badan Usaha Milik Negara (BUMN). Pada dasarnya terdapat 2 tujuan utama yang dimiliki oleh BUMN yakni, memberikan pelayanan publik (*public service*) yang sebaik-baiknya bagi masyarakat, dan tujuan berikutnya mendapatkan keuntungan atau laba dari hasil pemberian layanan publik kepada masyarakat [1]. PT. POS Indonesia telah membuka banyak cabang Kantor POS di seluruh Indonesia, salah satunya Kantor POS Rumbai yang berada di Kota Pekanbaru. Selain jasa pengiriman barang, Kantor POS Rumbai juga melayani berbagai macam pembayaran dan pembelian materai. Kantor POS Rumbai harus tetap mempertahankan pelanggannya dengan meningkatkan kualitas pelayanan. Jika pelanggan menerima layanan yang berkualitas baik, maka dapat menimbulkan tingkat kepuasan pelanggan yang tinggi [2]. Wijaya dan Girsang mengatakan bahwa kepuasan pelanggan dianggap sangat signifikan bagi suatu perusahaan karena dapat menghasilkan peningkatan profit pada perusahaan, serta memungkinkan perusahaan untuk mengidentifikasi pelanggan setia maupun tidak [3]. Menyediakan pelayanan yang berkualitas tinggi dapat menjadi sarana untuk menarik perhatian pelanggan dan memastikan bahwa perusahaan dapat terus berkembang [4]. Wood mengatakan bahwa keberhasilan perusahaan tergantung pada kemampuannya untuk memberikan kepuasan pelanggan, yang merupakan siklus proses yang saling terkait [5].

Kepuasan pelanggan merupakan suatu keadaan dimana terpenuhinya harapan, keinginan, dan kebutuhan konsumen [3]. Kepuasan pelanggan mencakup ketidaksamaan antara pentingnya tingkat layanan terhadap hasil yang dapat dirasakan [6]. Kepuasan pelanggan juga dapat dikatakan sebagai tingkat dimana anggapan terhadap jasa sesuai dengan harapan pada pelanggan [7]. Masyarakat akan mencari, memilih, dan membandingkan kualitas pelayanan yang terbaik dari setiap perusahaan ekspedisi. Aspek yang perlu diperhatikan dalam peningkatan pada pelayanan meliputi keramahan petugas, kebersihan tempat, fasilitas yang memadai, dan sebagainya. Menawarkan layanan dengan kualitas terbaik akan memastikan bahwa pelanggan puas, ini merupakan teknik yang dapat digunakan untuk bersaing [8]. Perusahaan akan berhasil

mendapatkan pelanggan dalam jumlah yang banyak apabila perusahaan tersebut dinilai mampu memberikan kepuasan bagi pelanggan. Untuk terus mempertahankan perusahaan, Kantor POS Rumbai perlu mengetahui sejauh mana tingkat kepuasan pelanggan saat ini, sehingga dapat memberikan acuan untuk meningkatkan kualitas pelayanan di Kantor POS Rumbai.

Untuk mengetahui kepuasan pelanggan ada beberapa cara, salah satunya dengan menggunakan teknik *data mining*. *Data Mining* merupakan sebuah proses untuk meng-ekstraksi, mengidentifikasi informasi, dan pengetahuan yang ada dalam *database* besar, dengan metode matematika, statistik, kecerdasan buatan, dan pembelajaran mesin [9]. Dalam *machine learning* terdapat beberapa algoritma dalam mengklasifikasi tingkat kepuasan salah satunya Naive Bayes. Naive Bayes adalah algoritma klasifikasi dengan menggunakan probabilitas dan statistik yang didasarkan pada teorema Bayes dalam memprediksi peluang asumsi yang kuat [10]. Naive Bayes mengasumsikan seluruh atribut yang saling tidak ketergantungan (*independen*) yang diberikan oleh nilai pada variabel kelas [11]. Naive Bayes mempunyai keunggulan yaitu minimnya data latih yang digunakan untuk menentukan estimasi parameter yang dibutuhkan dalam proses klasifikasi, serta berperan jauh lebih baik dalam situasi yang nyata dan kompleks [12].

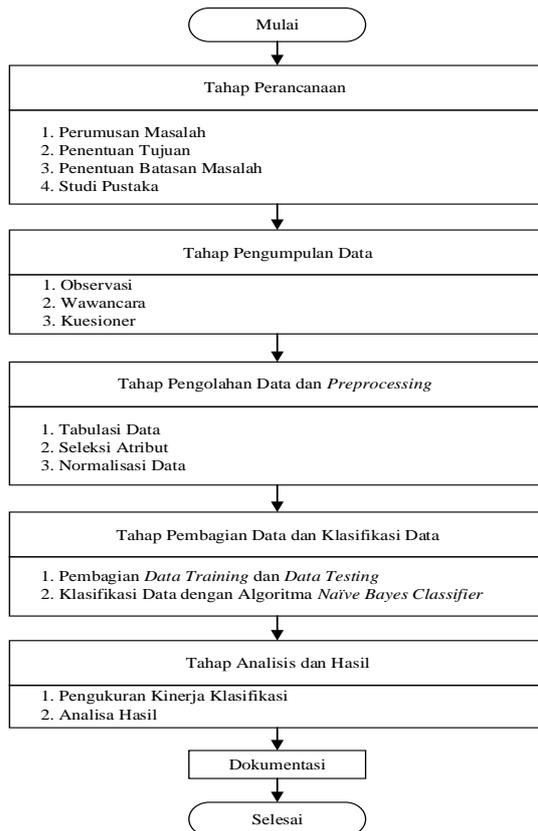
Beberapa penelitian mengenai klasifikasi kepuasan menggunakan Naive Bayes. Pertama adalah penelitian terkait kepuasan mahasiswa selama pembelajaran daring. Dari *dataset* yang berjumlah 110 menghasilkan nilai akurasi sebesar 100%, nilai presisi 100%, dan nilai recall 100% [13]. Kedua adalah penelitian terkait kepuasan nasabah. Penelitian menggunakan 25 data uji, dan menghasilkan kelas Puas sejumlah 20 data dan Tidak Puas sejumlah 5 data, dengan akurasi sebesar 88% [14]. Ketiga adalah penelitian terkait perbandingan algoritma C4.5 dengan Naive Bayes dalam prediksi kepuasan. Hasil akhir penelitian menunjukkan bahwa prediksi yang dihasilkan algoritma Naive Bayes lebih baik dibandingkan algoritma C4.5, dengan akurasi sebesar 96,89%, presisi 95,50%, dan recall 95% [15].

Berdasarkan uraian di atas maka penelitian ini menggunakan teknik *data mining* dengan algoritma Naive Bayes dalam menentukan tingkat kepuasan pelanggan terhadap pelayanan yang diberikan oleh Kantor POS Rumbai. Data yang akan digunakan berasal dari penyebaran kuesioner secara *offline* kepada pelanggan Kantor POS Rumbai, dengan menggunakan 14 indikator yang berasal dari Indeks Kepuasan Masyarakat yang telah ditetapkan oleh KepMenPAN pada tahun 2004. Dikarenakan populasi pelanggan Kantor POS Rumbai yang tidak teridentifikasi, maka pengambilan jumlah sampel menggunakan rumus *Lemeshow*. Pengolahan data

menggunakan *Microsoft Excel* dan *RapidMiner*. Hasil akhir dari penelitian ini diharapkan dapat menjadi sebuah informasi yang berguna mengenai akurasi kepuasan pelanggan terhadap pelayanan Kantor POS Rumbai.

2. METODE PENELITIAN

Kerangka metodologi dari seluruh tahapan pada penelitian ini dapat dilihat pada gambar 1.



Gambar 1. Alur Penelitian

2.1. Perencanaan

Pada tahap perencanaan terdapat 4 proses yang dilakukan. Pertama adalah perumusan masalah, bertujuan untuk membatasi pencarian dari isu penelitian. Kedua adalah penentuan tujuan, guna untuk mendapatkan target yang ingin dicapai pada penelitian. Ketiga adalah penentuan batasan masalah, bertujuan agar ruang lingkup dalam pembahasan tidak melenceng dari penelitian. Keempat adalah studi pustaka, bertujuan untuk mendapatkan referensi yang sesuai dan mendukung penelitian.

2.2. Pengumpulan Data

Tahapan awal dalam pengumpulan data yaitu observasi dan wawancara, yang bertujuan mengamati secara langsung dan mengumpulkan informasi dari Kantor POS Rumbai yang dapat digunakan dalam penelitian.

Data dikumpulkan melalui hasil penyebaran kuesioner kepada pelanggan Kantor POS Rumbai.

Kuesioner terdiri dari 14 butir pertanyaan yang diambil dari Unsur Indeks Kepuasan Masyarakat oleh KepMenPAN pada tahun 2004. Skala pada kuesioner penelitian menggunakan skala *Likert*, dengan menggunakan model 4 pilihan (skala empat), dan tidak menggunakan jawaban tengah (netral) dengan alasan [16] :

1. Memiliki arti ganda,
2. Jawaban netral tidak menentukan pendapat responden apakah tidak setuju atau setuju,
3. Kecenderungan responden memilih jawaban netral.

2.2.1. Populasi dan Sampel Penelitian

Populasi merupakan objek ataupun subjek dalam wilayah generalisasi yang memiliki karakteristik dan kualitas tersendiri untuk dipelajari oleh peneliti sehingga dapat menghasilkan kesimpulan [17]. Populasi yang digunakan pada penelitian ini adalah semua pelanggan yang datang ke Kantor Pos Rumbai.

Sampel merupakan bagian yang mewakili populasi yang memiliki jumlah dan karakteristik [18]. Penelitian ini menggunakan teknik *non-probability sampling* dengan *Accidental Sampling* sebagai jenis pengambilan sampel. Alasan menggunakan teknik *non-probability sampling* dikarenakan peluang dari setiap populasi yang dijadikan sampel tidak diketahui, hal ini disebabkan oleh tidak terdistribusinya total populasi. Sementara jenis *Accidental Sampling* adalah teknik penentuan sampel berdasarkan kebetulan, pelanggan bertemu dengan peneliti secara kebetulan dan dapat dijadikan sampel.

Untuk menentukan jumlah sampel penelitian jika total dari populasi tidak diketahui secara tepat dapat menggunakan rumus *Lemeshow* [19].

$$n = \frac{z^2 \cdot P(1-P)}{d^2} \quad (1)$$

Keterangan:

n = Total keseluruhan sampel

p = Estimasi maksimum = 50% = 0.5

z = Nilai standar = 1.96

d = alpha atau *sampling error* = 10% = 0.1

$$n = \frac{1,96^2 \cdot 0,5(1 - 0,5)}{0,1^2}$$

$$n = \frac{3,8416 \cdot 0,5 \cdot 0,5}{0,1^2}$$

$$n = \frac{0,9604}{0,1^2}$$

$$n = 96,04 = 97$$

Berdasarkan hasil perhitungan menggunakan rumus *Lemeshow*, maka didapatkan total sampel yang akan digunakan adalah 97 responden.

2.3. Pengolahan dan Preprocessing Data

Data yang telah terkumpul kemudian disusun kedalam bentuk tabel pada *Microsoft Excel*, menyeleksi atribut yang akan digunakan, dan melakukan normalisasi data, guna untuk menghindari terjadinya data yang tidak konsisten.

2.4. Pembagian Data dan Klasifikasi

Dataset keseluruhan dibagi menjadi 80% sebagai *data training* dan 20% sebagai *data testing*. Kemudian dilakukan klasifikasi dengan algoritma Naive Bayes. Tahapan klasifikasi dilakukan secara manual menggunakan *Microsoft Excel* dan diuji menggunakan *RapidMiner*.

2.5. Naïve Bayes Classifier

Naive Bayes merupakan salah satu algoritma klasifikasi yang sederhana dan berbasis probabilistik yang menghitung probabilitas berdasarkan kombinasi nilai dan frekuensi dari dataset yang digunakan [20]. Metode Naive Bayes menggunakan bagian dari ilmu matematika yang disebut dengan hipotesis kemungkinan dalam melacak kemungkinan terbaik dari pesanan potensial, dengan memeriksa pengulangan setiap karakteristik pada *data training* [21]. Karakteristik utama dari Naive Bayes yaitu memiliki asumsi yang kuat (naif) terhadap independensi pada setiap kejadian atau kondisi [22]. Kelebihan menggunakan Naive Bayes dalam proses klasifikasi adalah membutuhkan *data training* dengan jumlah yang lebih sedikit dalam menentukan estimasi parameter [23]. Dalam melakukan klasifikasi Naive Bayes dapat menggunakan persamaan berikut [24]:

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \quad (2)$$

Dimana:

X = Data yang akan diklasifikasi dengan kelas yang belum diketahui

H = Hipotesis data (kelas spesifik)

$P(H|X)$ = Posteriori probabilitas atau Probabilitas pada hipotesis H berdasar kondisi X

$P(H)$ = Prior probabilitas atau Probabilitas hipotesis H

$P(X|H)$ = Probabilitas X berdasarkan kondisi pada hipotesis H

$P(X)$ = Probabilitas X

Dalam proses klasifikasi Naive Bayes membutuhkan sejumlah petunjuk untuk menentukan kelas yang tepat untuk sampel yang dianalisis.

$$Posterior = \frac{prior \times likelihood}{evidence} \quad (3)$$

Pada satu sampel, setiap kelas selalu mempunyai nilai *evidence* yang tetap. Dalam proses

klasifikasi sampel, untuk menentukan kelas yang tepat dilakukan perbandingan nilai *posterior* terhadap nilai *posterior* kelas lainnya.

2.6. Data Mining

Data Mining merupakan proses penemuan pola dan penambangan data menjadi suatu informasi yang mempunyai nilai [25]. *Gatner Group* mendefinisikan *data mining* semacam metode statistik dan matematika, dengan menciptakan ikatan yang berarti, identifikasi pola serta kecenderungan dengan mengekstrak sejumlah besar informasi yang tersimpan pada penyimpanan [26]. *Data Mining* dibagi menjadi dua kategori utama [27]:

1. *Descriptive Mining*, proses untuk menentukan karakteristik penting dari kata yang terdapat dalam satu *database*,
2. *Predictive*, proses menentukan pola dengan menggunakan beberapa variabel lain dimasa depan dalam sebuah data.

Dalam *data mining* terdapat beberapa tahapan sebagai berikut [28] :

1. Tahap Seleksi Data (*Data Selection*),
2. Tahap Pembersihan Data (*Data Cleaning*),
3. Tahap Integrasi Data (*Data Integration*),
4. Tahap Transformasi Data (*Data Transformation*),
5. Tahap Penambangan Data (*Data Mining*),
6. Tahap Evaluasi Pola (*Pattern Evaluation*)
7. Tahap Presentasi Pengetahuan (*Knowledge Presentation*).

2.7. Confusion Matrix

Confusion Matrix merupakan teknik untuk menilai kemampuan model klasifikasi yang dibentuk dalam mengidentifikasi data dengan kategori kelas yang berbeda [29]. Mengukur kinerja model klasifikasi dilakukan dengan mencari nilai akurasi. Akurasi adalah tingkat ketepatan antara nilai aktual dengan nilai prediksi [30]. Pada tabel 1 merupakan gambaran model *Confusion Matrix* [29].

Tabel 1. *Confusion Matrix*

		Kelas Prediksi	
		Yes	No
Aktual Kelas	Yes	TP	FN
	No	FP	TN

Keterangan :

Kelas Aktual (yes) : kelas sebenarnya dengan nilai yes;

Kelas Aktual (no) : kelas sebenarnya dengan nilai no;

Kelas Prediksi (yes) : prediksi kelas dengan nilai yes;

Kelas Prediksi (no) : prediksi kelas dengan nilai no;

True Positive (TP) : jumlah kelas ‘yes’ yang diprediksi sebagai ‘yes’;

False Negative (FN) : jumlah kelas ‘yes’ yang diprediksi sebagai ‘no’;
 False Positive (FP) : jumlah kelas ‘no’ yang diprediksi sebagai ‘yes’;
 True Negative (TN) : jumlah kelas ‘no’ yang diprediksi sebagai ‘no’.

Untuk menghitung tingkat akurasi, presisi, dan recall dapat menggunakan persamaan berikut [31]:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (4)$$

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (5)$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (6)$$

2.8. RapidMiner

RapidMiner adalah software yang umumnya digunakan untuk mesin penambangan dan analisis data, serta terintegrasi dengan berbagai bahasa pemrograman sehingga dapat digunakan secara mudah [32]. RapidMiner dapat bekerja dalam banyak sistem operasi dikarenakan RapidMiner ditulis menggunakan bahasa pemrograman Java.

3. HASIL DAN PEMBAHASAN

3.1. Data Penelitian

Penelitian ini menggunakan data training sebanyak 78 data dengan nilai Puas sejumlah 65 data dan Tidak Puas sejumlah 13 data, sedangkan data testing yang digunakan berjumlah 19 data. Adapun data testing pada penelitian ini dapat dilihat pada tabel 2.

Tabel 2. Data Testing

R	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	K
79	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	?
80	1.00	0.50	1.00	1.00	1.00	0.50	1.00	1.00	1.00	1.00	0.50	0.50	1.00	1.00	?
81	0.50	1.00	1.00	0.50	1.00	0.50	0.50	1.00	0.50	0.50	1.00	0.50	1.00	0.50	?
82	0.50	0.50	0.50	0.50	1.00	1.00	0.50	1.00	1.00	0.50	0.50	0.50	0.50	0.50	?
83	1.00	0.50	1.00	0.50	0.50	1.00	1.00	1.00	0.50	1.00	1.00	1.00	0.50	0.50	?
84	0.50	0.50	0.50	0.50	0.50	0.50	0.00	0.00	0.50	0.50	0.00	0.50	0.50	0.50	?
85	0.50	1.00	0.00	0.00	0.50	0.50	0.50	0.50	1.00	1.00	1.00	0.50	0.50	1.00	?
86	0.50	0.50	0.50	0.50	0.50	0.50	0.50	1.00	0.50	0.50	1.00	0.50	0.50	0.50	?
87	1.00	1.00	0.50	1.00	1.00	1.00	1.00	0.50	1.00	1.00	1.00	1.00	1.00	1.00	?
88	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	?
...
97	0.50	1.00	0.50	0.50	1.00	0.50	1.00	0.50	0.50	1.00	0.50	1.00	0.50	1.00	?

3.2. Perhitungan Menggunakan Metode Naïve Bayes

1) Menghitung Nilai Probabilitas Kelas
 Kelas Puas berjumlah 65 data dan kelas Tidak Puas berjumlah 13 data. Sehingga dilakukan perhitungan sebagai berikut.

$$P(H|P) = \frac{65}{78} = 0.83333$$

$$P(H|TP) = \frac{13}{78} = 0.16667$$

2) Menghitung Nilai Probabilitas Atribut
 Selanjutnya mencari nilai Probabilitas masing-masing atribut P1 sampai dengan P14. Perhitungan probabilitas atribut P1.

$$P(1,00 | P) = \frac{35}{65} = 0.53846$$

$$P(1,00 | TP) = \frac{2}{13} = 0.15385$$

$$P(0,50 | P) = \frac{13}{29} = 0.44615$$

$$P(0,50 | TP) = \frac{6}{13} = 0.46154$$

$$P(0,00 | P) = \frac{1}{65} = 0.01538$$

$$P(0,00 | TP) = \frac{5}{13} = 0.38462$$

$$P(0,00 | P) = \frac{0}{65} = 0.00000$$

$$P(0,00 | TP) = \frac{0}{13} = 0.00000$$

Tabel 3. Perhitungan Probabilitas Atribut P1

P01	Puas	Tidak Puas	Puas	Tidak Puas
SB (1.00)	35	2	0.53846	0.15385
B (0.50)	29	6	0.44615	0.46154
TB (0.00)	1	5	0.01538	0.38462
STB (0.000)	0	0	0.00000	0.00000
Total	65	13	1.00000	1.00000

Probabilitas Atribut P1 pada kategori Puas dengan skala SB menghasilkan probabilitas 0,53846, B menghasilkan probabilitas 0,44615, TB menghasilkan probabilitas 0,01538, dan STB menghasilkan probabilitas 0. Dengan begitu maka jumlah probabilitas kategori Puas adalah 1. Sedangkan pada kategori Tidak Puas dengan skala SB menghasilkan probabilitas 0,15385, B menghasilkan probabilitas 0,46154, TB menghasilkan probabilitas 0,38462, dan STB menghasilkan probabilitas 0. Sehingga jumlah probabilitas kategori Tidak Puas adalah 1. Demikian juga selanjutnya dilakukan perhitungan yang sama pada atribut P2 sampai dengan atribut P14.

3) Menghitung Nilai Probabilitas Data Testing

Data testing yang berjumlah 19 data akan dicari nilai Probabilitasnya masing-masing, sehingga dapat diketahui responden 79 sampai dengan 97 termasuk dalam klasifikasi "Puas (P)" atau "Tidak Puas (TP)". Perhitungan Probabilitas pada data testing dapat dilihat sebagai berikut:

$$\begin{aligned}
 P(R79|P) &= P(P1|P) \times P(P2|P) \times P(P3|P) \times \\
 &P(P4|P) \times P(P5|P) \times P(P6|P) \times \\
 &P(P7|P) \times P(P8|P) \times P(P9|P) \times \\
 &P(P10|P) \times P(P11|P) \times P(P12|P) \times \\
 &P(P13|P) \times P(P14|P) \\
 &= 0,4 \times 0,6 \times 0,5 \times 0,4 \times 0,5 \times 0,5 \times 0,5 \\
 &\times 0,5 \times 0,3 \times 0,5 \times 0,5 \times 0,5 \times 0,5 \times \\
 &0,5 \\
 &= 0.0000265
 \end{aligned}$$

$$\begin{aligned}
 P(R79|TP) &= P(P1|TP) \times P(P2|TP) \times \\
 &P(P3|TP) \times P(P4|TP) \times \\
 &P(P5|TP) \times P(P6|TP) \times \\
 &P(P7|TP) \times P(P8|TP) \times \\
 &P(P9|TP) \times P(P10|TP) \times \\
 &P(P11|TP) \times P(P12|TP) \times \\
 &P(P13|TP) \times P(P14|TP) \\
 &= 0,5 \times 0,4 \times 0,5 \times 0,5 \times 0,7 \times 0,4 \times \\
 &0,2 \times 0,2 \times 0,7 \times 0,2 \times 0,5 \times 0,2 \times \\
 &0,5 \times 0,7 \\
 &= 0.0000007
 \end{aligned}$$

Demikian juga selanjutnya dilakukan perhitungan yang sama pada *data testing* R80 sampai dengan R97.

- 4) Menghitung Hasil Probabilitas Akhir
 Setelah mendapatkan nilai Puas dan Tidak Puas pada masing masing *data testing*, selanjutnya menghitung probabilitas akhir dari setiap *data testing*. *Data Testing* R79.

$$\begin{aligned}
 P(X | P) &= 0.0000265 \times 0.83333 \\
 &= 0.0000220 \\
 P(X | TP) &= 0.0000007 \times 0.16667 \\
 &= 0.0000001
 \end{aligned}$$

Demikian juga selanjutnya dilakukan perhitungan yang sama pada *data testing* R80 sampai dengan 97.

- 5) Membandingkan Hasil Probabilitas Akhir
 Jika nilai akhir Puas pada lebih besar dibanding nilai Tidak Puas, maka termasuk kedalam klasifikasi kelas "Puas". Sebaliknya jika nilai akhir Tidak Puas lebih besar dibanding nilai Puas, maka termasuk kedalam klasifikasi "Tidak Puas". Hasil perbandingan probabilitas akhir pada setiap *data testing* ditunjukkan pada tabel 4.

Tabel 4. Hasil Perbandingan Probabilitas Akhir

Data Testing	Puas	Tidak Puas	Klasifikasi
R79	0.0000220	0.0000001	Puas
R80	0.0000680	0.0000000	Puas
R81	0.0000235	0.0000000	Puas
R82	0.0000553	0.0000000	Puas
R83	0.0000193	0.0000000	Puas
R84	0.0000000	0.0000029	Tidak Puas

R85	0.0000000	0.0000043	Tidak Puas
R86	0.0000228	0.0000000	Puas
R87	0.0000512	0.0000000	Puas
R88	0.0000220	0.0000001	Puas
R89	0.0000141	0.0000000	Puas
R90	0.0000530	0.0000000	Puas
R91	0.0000729	0.0000000	Puas
R92	0.0000220	0.0000001	Puas
R93	0.0000272	0.0000000	Puas
R94	0.0000000	0.0000072	Tidak Puas
R95	0.0000166	0.0000000	Puas
R96	0.0000235	0.0000000	Puas
R97	0.0000106	0.0000000	Puas

Hasil perhitungan manual terhadap 19 *data testing* diperoleh 3 data yang termasuk dalam prediksi kelas "Tidak Puas" dengan persentase 15,8%, dan 16 data yang termasuk dalam prediksi kelas "Puas" dengan persentase 84,2%.

3.3. Pengukuran Kinerja Klasifikasi

Tabel 5 menunjukkan hasil *Confusion Matrix* yang diperoleh melalui proses klasifikasi pada data testing.

Tabel 5. *Confusion Matrix*

	Klasifikasi	Prediksi	
		Puas	Tidak Puas
Aktual	Puas	16	1
	Tidak Puas	0	2

Berdasarkan *Confusion Matrix* diperoleh 16 data sebagai *true positive* (TP), 2 data sebagai *true negative* (TN), 0 data sebagai *false positive* (FP), dan 1 data sebagai *false negative* (FN).

$$\begin{aligned}
 Accuracy &= \frac{16+2}{16+1+0+2} \times 100\% \\
 &= 94,74\%
 \end{aligned}$$

$$\begin{aligned}
 Precision &= \frac{16}{16+0} \times 100\% \\
 &= 100\%
 \end{aligned}$$

$$\begin{aligned}
 Recall &= \frac{16}{16+1} \times 100\% \\
 &= 94,12\%
 \end{aligned}$$

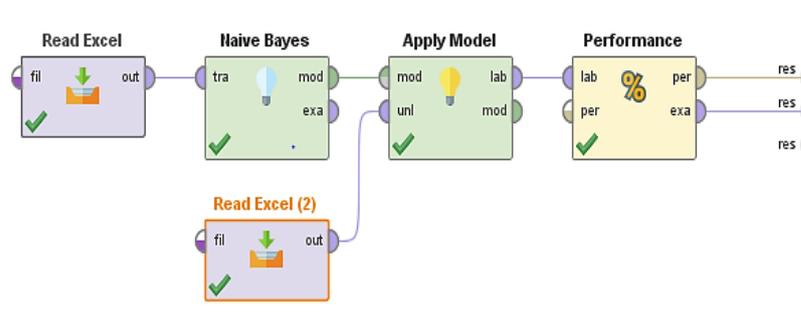
Maka hasil dari pengujian klasifikasi menggunakan Naive Bayes dengan perhitungan manual pada data sampel, didapatkan nilai *accuracy* 94,74%, *precision* 100%, dan *recall* 94,12%.

3.3. Pengujian Pada Tools RapidMiner

Pengujian dengan menggunakan *RapidMiner* perlu dilakukan untuk mengetahui kecocokan antara hasil perhitungan manual dengan hasil pada *tools*.

- 1) Pemodelan Pada *RapidMiner*

Pemodelan *data training* dan *data testing* dapat dilihat pada Gambar 2.



Gambar 2. Pemodelan Data Training dan Data Testing

Gambar 2 menunjukkan tampilan pemodelan pada RapidMiner. Masukkan *data training* dan *testing* pada operator *Read Excel*. Selanjutnya masukkan operator *Naive Bayes*, diikuti dengan operator *Apply Model*, dan juga operator *Performance*. Kemudian sambungkan setiap operator tersebut. Setelah semua operator terhubung, selanjutnya dapat menjalankan proses klasifikasi.

2) Hasil Klasifikasi

Hasil klasifikasi *data testing* dapat dilihat pada Gambar 3. Klasifikasi menggunakan *data testing* sejumlah 19 data menghasilkan kelas Puas dengan prediksi Puas sebanyak 16 data, kelas Puas dengan prediksi Tidak Puas sebanyak 1 data, dan kelas Tidak Puas dengan prediksi Tidak Puas sebanyak 2 data.

Row No.	CLASS	prediction(C...	confidence(...	confidence(...	No	P1	P2	P3	P
4	Puas	Puas	1.000	0.000	R82	0.5	0.5	0.5	0
5	Puas	Puas	1.000	0.000	R83	1	0.5	1	0
6	Tidak Puas	Tidak Puas	0.000	1.000	R84	0.5	0.5	0.5	0
7	Puas	Tidak Puas	0.000	1.000	R85	0.5	0.5	0	0
8	Puas	Puas	1.000	0.000	R86	0.5	0.5	0.5	0
9	Puas	Puas	1	0	R87	1	1	0.5	1
10	Puas	Puas	0.977	0.023	R88	0.5	0.5	0.5	0
11	Puas	Puas	1.000	0.000	R89	0.5	1	0.5	1
12	Puas	Puas	1	0	R90	1	1	1	1
13	Puas	Puas	1.000	0.000	R91	0.5	0.5	0.5	1
14	Puas	Puas	0.977	0.023	R92	0.5	0.5	0.5	0
15	Puas	Puas	1.000	0.000	R93	1	0.5	1	1
16	Tidak Puas	Tidak Puas	0.000	1.000	R94	0.5	0.5	0.5	0

Gambar 3. Hasil Klasifikasi Data Testing

3) Hasil Pengukuran

Hasil pengukuran kinerja algoritma Naive Bayes dapat dilihat pada Gambar 4.

accuracy: 94,74%

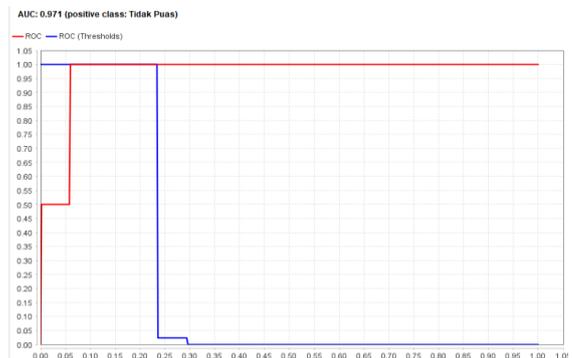
	true Puas	true Tidak Puas	class precision
pred. Puas	16	0	100.00%
pred. Tidak Puas	1	2	66.67%
class recall	94.12%	100.00%	

Gambar 4. Hasil Pengukuran

Hasil pengukuran pada *RapidMiner* menghasilkan *accuracy* dengan nilai 94,74%, *precision* dengan nilai 100%, dan *recall* dengan nilai 94,12%.

4) Nilai AUC

Nilai AUC pada kurva ROC dapat dilihat pada Gambar 5.



Gambar 5. Kurva ROC

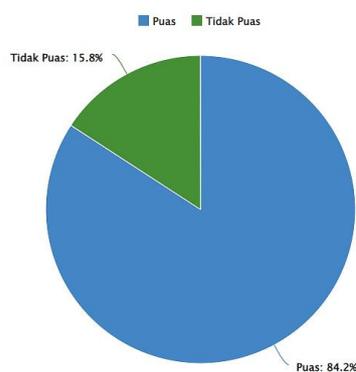
Dari pengujian diperoleh nilai AUC berdasarkan kurva ROC sebesar 0,971 dimana hampir mendekati 1, sehingga tergolong *Excellent Classification*.

4. DISKUSI

Beberapa penelitian terdahulu terkait klasifikasi kepuasan menggunakan algoritma Naive Bayes. Pertama penelitian terkait kepuasan pelanggan

terhadap pelayanan di Dinas Tenaga Kerja Kabupaten Karawang. Penelitian membandingkan algoritma Naive Bayes dan C4.5 dalam pengolahan data. Hasil akhir menunjukkan Naive Bayes lebih unggul dibandingkan C4.5 karena menghasilkan nilai prediksi yang lebih tinggi yaitu 96,89% [15]. Penelitian kedua terkait prediksi kepuasan pelayanan perekaman E-KTP. Peneliti tersebut juga melakukan perbandingan dengan dua algoritma yaitu *Decision Tree* dan Naive Bayes. Penelitian menghasilkan algoritma Naive Bayes sebagai akurasi yang tertinggi yaitu 91,70% sedangkan *Decision Tree* sebesar 65,90% [33]. Penelitian ketiga terkait prediksi tingkat kepuasan mahasiswa terhadap pelayanan akademis dengan algoritma Naive Bayes. Penelitian menggunakan indikator dari lima dimensi *servqual*, dan menghasilkan akurasi sebesar 96,71% [29].

Berdasarkan dari ketiga penelitian terbukti bahwa Naive Bayes efektif diimplementasikan pada kasus klasifikasi tingkat kepuasan dengan hasil akurasi yang cukup tinggi. Oleh karena itu penulis ingin menggunakan algoritma Naive Bayes untuk melakukan klasifikasi tingkat kepuasan pelanggan terhadap pelayanan Kantor POS Rumbai. Penelitian menjadikan 14 Unsur Indeks Kepuasan Masyarakat dari KepMenPan sebagai indikator dalam kuesioner. Hasil klasifikasi dapat dilihat pada gambar 6.



Gambar 6. Diagram Hasil Klasifikasi

Hasil klasifikasi pada algoritma Naive Bayes dengan menggunakan 19 *data testing* menghasilkan klasifikasi kelas Puas sebesar 84,2% dan kelas Tidak Puas sebesar 15,8%. Kelas Puas yang sebelumnya Puas sebanyak 16 data, kelas Tidak Puas yang sebelumnya Tidak Puas terdiri dari 2 data, dan kelas Tidak Puas yang sebelumnya Puas terdiri dari 1 data. Pengujian pada *RapidMiner* menghasilkan klasifikasi yang sama dengan nilai *accuracy* 94,74%, nilai *precision* 100%, dan nilai *recall* 94,12%.

5. KESIMPULAN

Melalui hasil penelitian yang telah dilakukan, diperoleh kesimpulan bahwa algoritma Naive Bayes dapat di implementasikan dengan baik untuk klasifikasi tingkat kepuasan pelanggan terhadap pelayanan Kantor POS Rumbai (Puas atau Tidak Puas). Hasil klasifikasi dengan 59 data training dan

19 data testing memperoleh 3 data sebagai kelas Tidak Puas dan 16 data sebagai kelas Puas. Persentase pada kelas Puas sebesar 84,2% dan kelas Tidak Puas sebesar 15,8%. Dapat disimpulkan bahwa pelanggan Kantor POS Rumbai sudah puas dengan pelayanan yang diberikan. Pengujian pada *RapidMiner* memprediksi hasil klasifikasi dengan nilai *accuracy* sebesar 94,74%, *precision* 100%, *recall* 94,12%, dan nilai AUC 0.971, mendekati angka 1, sehingga tergolong *Excellent Classification*. Hasil dari penelitian nantinya sangat berguna bagi pihak Kantor POS Rumbai sebagai acuan dalam memperhatikan kualitas pelayanan.

Adapun saran yang ingin disampaikan untuk penelitian selanjutnya adalah sebagai berikut.

Penelitian dapat menggunakan *dataset* lebih banyak lagi, dikarenakan semakin banyak data yang digunakan maka semakin baik klasifikasi yang dihasilkan. Selain itu, pemilihan atribut sangat penting dalam penelitian.

Penelitian dapat dikembangkan dengan menambah beberapa algoritma klasifikasi lain sebagai pembanding. Sehingga dapat menghasilkan prediksi yang terbaik.

DAFTAR PUSTAKA

- [1] B. Sutikno, "Kualitas Pelayanan PT . Pos Indonesia (PERSERO) Cabang," *J. Prof.*, vol. 9, no. 1, pp. 125–132, 2022.
- [2] A. F. Anita Nurjanah, Nurul Aulia, Errissya Rasywir, "Analisis Tingkat Kepuasan Konsumen Terhadap Layanan Ekspedisi J & T Express Dengan Metode Algoritma Naive Bayes," *J. Manaj. Teknol. dan Sist. Inf.*, vol. 1, no. April, pp. 78–88, 2022.
- [3] R. Alfiani, Y. Umaidah, J. H. Ronggo Waluyo, T. Jambe Timur, and J. Barat, "Cara sitasi: Alfiani R, Umaidah Y. 2022. Prediksi Kepuasan Customer Terhadap Performa Terapis Baby Massage Dengan Algoritma Naive Bayes," *Bina Insa. ICT J.*, vol. 9, no. 1, pp. 83–92, 2022.
- [4] M. F. Afif, Saryadi, and R. E. Werdani, "Pengaruh Kualitas Pelayanan, Kualitas Produk, Harga Terhadap Kepuasan Konsumen," *J. Adm. Bisnis*, vol. 11, no. 2, pp. 370–380, 2022.
- [5] F. Nurlita, "Analisis Kepuasan Pelanggan Terhadap Kualitas Pelayanan PDAM Tirta Pengabuan Kuala Tungkal," *Ekon. Syariah*, vol. 2, no. 2, pp. 1–15, 2019.
- [6] A. K. Febrian, Y. H. Chrisnanto, D. Pupita, N. Sabrina, and J. Achmad Yani, "Studi Komparasi Metode Klasifikasi K-Nearest Neighbor dan Naive Bayes dalam Mengidentifikasi Kepuasan Pelanggan Terhadap Produk," *Semin. Nas. Tek. Elektro, Sist. Informasi, dan Tek. Inform.*, p. 333, 2022.

- [7] F. Ariani and A. Taufik, "Perbandingan Metode Klasifikasi Data Mining untuk Prediksi Tingkat Kepuasan Pelanggan Telkomsel Prabayar," *SATIN - Sains dan Teknol. Inf.*, vol. 6, no. 2, pp. 46–55, 2020.
- [8] M. Jufriyanto, "Analisis Tingkat Kepuasan Konsumen Pada Kualitas Pelayanan Kedai Kopi Shelter," *Matrik*, vol. 20, no. 2, p. 79, 2020.
- [9] A. D. Amanda, A. P. Windarto, and H. Qurniawan, "Analisis Kepuasan Konsumen terhadap Pelayanan Store Ms Glow Menggunakan Metode Naïve Bayes," *Hello World J. Ilmu Komput.*, vol. 1, no. 3, pp. 130–144, 2022.
- [10] A. N. Yuliarina, "Comparison of Prediction Analysis of GoFood Service Users Using the KNN & Naive Bayes Algorithm With RapidMiner Software," *J. Tek. Inform.*, vol. 3, no. 4, pp. 847–856, 2022.
- [11] D. A. Pratiwi, R. M. Awangga, and M. Y. H. Setyawan, *Seleksi Calon Kelulusan Tepat Waktu Mahasiswa Teknik Informatika Menggunakan Metode Naive Bayes*. Bandung: Kreatif Industri Nusantara, 2020.
- [12] W. I. Rahayu, C. Prianto, and E. A. Novia, "Perbandingan Algoritma K-Means Dan Naïve Bayes Untuk Memprediksi Prioritas Pembayaran Tagihan Rumah Sakit Berdasarkan Tingkat Kepentingan Pada Pt. Pertamina (Persero)," *J. Tek. Inform.*, vol. 13, no. 2, pp. 1–8, 2021.
- [13] A. R. Damanik, S. Sumijan, and G. W. Nurcahyo, "Prediksi Tingkat Kepuasan dalam Pembelajaran Daring Menggunakan Algoritma Naïve Bayes," *J. Sistim Inf. dan Teknol.*, vol. 3, pp. 88–94, 2021.
- [14] M. G. Sadewo, A. P. Windarto, and I. S. Damanik, "Algoritma Naïve Bayes Dalam Memprediksi Kepuasan Nasabah," *Pros. Semin. Nas. Ris. Inf. Sci.*, vol. 1, no. September, p. 318, 2019.
- [15] Y. Umaidah and U. Enri, "Prediction of Public Service Satisfaction Using C4 . 5 and Naïve Bayes Algorithm," *J. PILAR Nusa Mandiri*, vol. 17, no. 2, pp. 143–148, 2021.
- [16] S. Azwar, *Sikap Manusia Teori dan Pengukurannya*. Yogyakarta: Pustaka Belajar, 1997.
- [17] Sugiyono, *Metode Penelitian Kuantitatif, Kualitatif dan R&D*. Bandung: Alfabeta, 2019.
- [18] Sugiyono, *Metode Penelitian Bisnis*. Bandung: Alfabeta, 2013.
- [19] S. Riyanto and A. A. Hatmawan, *Metode Riset Penelitian Kuantitatif Penelitian Di Bidang Manajemen, Teknik, Pendidikan Dan Eksperimen*. Sleman: Deepublish, 2020.
- [20] L. Siliyani, Iqbal Agis Junizar, Uyu Nuraeni, Edi Tohidi, and Irfan Ali, "Penerapan Algoritma Naive Bayes Untuk Mengetahui Kepuasan Mahasiswa Terhadap Layanan Administrasi Keuangan," *KOPERTIP J. Ilm. Manaj. Inform. dan Komput.*, vol. 4, no. 3, pp. 72–79, 2020.
- [21] E. Martantoh and N. Yanih, "Implementasi Metode Naïve Bayes Untuk Klasifikasi Karakteristik Kepribadian Siswa Di Sekolah MTS Darussa'adah Menggunakan Php Mysql," *J. Teknol. Sist. Inf.*, vol. 3, no. 2, pp. 166–175, 2022.
- [22] R. Annisa, "Klasifikasi Kepuasan Pasien BPJS Kesehatan Terhadap Pelayanan Klinik Menggunakan Algoritma Naive Bayes Classifier," Universitas Islam Negeri Sultan Syarif Kasim, 2020.
- [23] M. K. Hartono and H. Hendry, "Prediction of Baby Birth Rate Using Naïve Bayes Classification Algorithm in Randau Village," *J. Tek. Inform.*, vol. 3, no. 4, pp. 863–868, 2022.
- [24] S. Ramadani, N. Z. S. Ayu, N. Nurhayati, F. Azzahra, and A. P. Windarto, "Analisis Data Mining Naive Bayes Klasifikasi Pada Kelayakan Penerima PKH," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 4, no. 1, pp. 374–381, 2020.
- [25] A. Anas and R. Hartha Delima, "Penerapan Data Mining Untuk Menganalisis Data Mahasiswa Baru STIE-GK Muara Bulian," *J. Ilm. Media Sisfo*, vol. 15, no. 2, pp. 119–128, 2021.
- [26] D. T. Larose and C. T. Larose, *Discovering Knowledge in Data: an Introduction to Data Mining*, Second., vol. 100, no. 472. Canada: John Wiley & Sons, 2005.
- [27] Noviyanto, "Penerapan Data Mining dalam Mengelompokkan Jumlah Kematian," *J. Inform. dan Komput.*, vol. 22, no. 2, pp. 183–188, 2020.
- [28] A. Supoyo, "Biaglala Informatika Analisis Data Mining Untuk Memprediksi Lama Perawatan Pasien Covid-19 Di DIY," *Biaglala Inform.*, vol. 10, no. 1, p. 2022, 2022.
- [29] M. Siddik, Y. Desnelita, and Gustientiedina, "Penerapan Naïve Bayes untuk Memprediksi Tingkat Kepuasan Mahasiswa Terhadap Pelayanan Akademis," *J. Infomedia*, vol. 2, no. 4, pp. 89–93, 2019.
- [30] Suparyanto, "Klasifikasi Kepuasan Layanan Akademik Di STMIK El Rahma Menggunakan Metode Algoritma Naive Bayes," *Fahma*, vol. 20, no. 2, pp. 100–111, 2022.

- [31] M. A. Maulana, A. Triayudi, and I. D. Sholihati, "Accuracy Analysis of Community Satisfaction in Population Administration Services Using the C4.5 Algorithm and Naïve Bayes Method," *J. Mantik*, vol. 4, no. 1, pp. 850–860, 2020.
- [32] V. R. Prasetyo, H. Lazuardi, A. A. Mulyono, and C. Lauw, "Penerapan Aplikasi RapidMiner Untuk Prediksi Nilai Tukar Rupiah Terhadap US Dollar Dengan Metode Linear Regression," *J. Nas. Teknol. dan Sist. Inf.*, vol. 7, no. 1, pp. 8–17, 2021.
- [33] T. H. Apandi and C. A. Sugianto, "Algoritma Naive Bayes untuk Prediksi Kepuasan Pelayanan Perekaman e-KTP," *JUITA J. Inform.*, vol. 7, no. 2, p. 125, 2019..