

## COMPARISON OF PREDICTION ANALYSIS OF GOFOOD SERVICE USERS USING THE KNN & NAIVE BAYES ALGORITHM WITH RAPIDMINER SOFTWARE

Agista Nindy Yuliarina<sup>\*1</sup>, Hendry<sup>2</sup>

<sup>1,2</sup>Program Studi Teknik Informatika, Universitas Kristen Satya Wacana, Indonesia  
Email: <sup>1</sup>[672018171@student.uksw.edu](mailto:672018171@student.uksw.edu), <sup>2</sup>[hendry@uksw.edu](mailto:hendry@uksw.edu)

(Naskah masuk: 25 April 2022, Revisi : 09 Mei 2022, diterbitkan: 20 Agustus 2022)

### Abstract

*GoFood is a service provider that has a very important role in human life, especially in this growing era. Currently, many service providers are competing to meet the needs of users, including GoFood. However, not all service providers can meet and know the needs needed by users, because they focus on the services offered and only the quality of services provided. Therefore, survey analysis is needed to obtain customer satisfaction data that will be used to satisfy GoFood service users. The classification method uses the KNN and Naive Bayes algorithms, which are good algorithms for testing 1,000 records of GoFood user data that have been obtained previously. The test results using Cross Validation and T-Test show that the KNN algorithm is the best algorithm with 98.80% Accuracy and 100% Recall, while Naive Bayes obtains 94.10% Accuracy and 94.43% Recall.*

**Keywords:** Classification, GoFood, KNN, Naive Bayes.

## PERBANDINGAN ANALISIS PREDIKSI KEPUASAN PENGGUNA LAYANAN GOFOOD MENGGUNAKAN ALGORITMA KNN & NAIVE BAYES DENGAN SOFTWARE RAPIDMINER

### Abstrak

*GoFood merupakan penyedia layanan jasa yang memiliki peran sangat penting dalam kehidupan manusia, terutama pada era zaman yang semakin berkembang ini. Saat ini banyak penyedia layanan jasa yang berlomba-lomba untuk dapat memenuhi kebutuhan para pengguna, termasuk GoFood. Namun tidak semua penyedia layanan jasa dapat memenuhi dan mengetahui kebutuhan yang diperlukan oleh pengguna, dikarenakan mereka hanya fokus terhadap jasa yang ditawarkan dan mengesampingkan kualitas dari pelayanan yang diberikan. Oleh karena itu diperlukan analisis survei untuk memperoleh data kepuasan pelanggan yang akan digunakan untuk melakukan prediksi kepuasan pengguna layanan GoFood. Dengan menggunakan algoritma KNN dan Naive Bayes, dimana kedua algoritma ini merupakan algoritma yang baik untuk melakukan pengujian terhadap 1.000 record data pengguna GoFood yang telah diperoleh sebelumnya. Hasil pengujian dengan menggunakan Cross Validation dan T-Test menunjukkan bahwa algoritma KNN merupakan algoritma terbaik dengan Accuracy 98,80% dan Recall 100% sedangkan Naive Bayes memperoleh hasil Accuracy sebesar 94,10% dan Recall sebesar 94,43%.*

**Kata kunci:** GoFood, Klasifikasi, KNN, Naive Bayes.

### 1. PENDAHULUAN

Perkembangan teknologi pada era globalisasi saat ini sangatlah pesat dan signifikan, baik dalam lingkup sosial, pendidikan, dan terutama dalam lingkup ekonomi. Perkembangan ini dapat berlangsung dalam skala yang kecil maupun skala yang besar, oleh karena itu dalam hal ini semua pihak harus dapat mengikuti dan menyeimbangkan diri untuk mampu bertahan dan mampu bersaing, tak terkecuali *Gojek* sebagai penyedia layanan *GoFood* yang merupakan layanan pesan antar makanan.

Dengan banyaknya persaingan yang ketat ini, pelaku usaha harus mampu menghadapi dan

mengetahui kualitas pelayanan yang diberikan untuk kepuasan pelanggan. Menurut (Djaslim Saladin, 2003) kepuasan pelanggan adalah perasaan senang atau kecewa seseorang yang berasal dari perbandingan antara kesan yang dirasakan terhadap kinerja atau hasil dari suatu produk dan harapan yang di inginkan untuk produk tersebut. Kepuasan pelanggan sangat penting dalam bidang penjualan jasa yang ditawarkan, karena pelanggan akan mencari layanan jasa yang sesuai dengan yang mereka inginkan seperti kenyamanan, kepuasan, dan keamanan dalam menggunakan layanan tersebut.

*GoFood* merupakan layanan pesan antar makanan yang sangat membantu dan memudahkan para penggunanya. Terutama pada masa pandemi *Covid-19* ini, banyak orang yang kini menggunakan layanan pesan antar makan karena merasa takut dan cemas untuk keluar rumah. Untuk dapat mempertahankan para penggunanya perusahaan harus dapat memahami harapan dan keinginan dari para pengguna sehingga perusahaan dapat mengetahui seberapa besar tingkat kepuasan dari setiap penggunanya.

Layanan pesan antar makanan yang ditawarkan *GoFood* dirasa sangat membantu dan memudahkan para penggunanya. Dengan menggunakan layanan ini pengguna dapat memesan makanan yang diinginkan secara *online* tanpa harus melakukan kontak secara langsung. Terutama pada masa pandemi *Covid-19* ini, banyak orang yang kini menggunakan layanan pesan antar makan karena merasa takut dan cemas untuk keluar rumah. Banyaknya layanan pesan antar dan ketatnya persaingan membuat perusahaan di bidang layanan ini bersaing untuk menarik pengguna untuk menggunakan layanan yang ditawarkan. Untuk dapat mempertahankan para penggunanya perusahaan harus dapat memahami harapan dan keinginan dari para pengguna sehingga perusahaan dapat mengetahui seberapa besar tingkat kepuasan dari setiap penggunanya.

Untuk dapat memprediksi kepuasan pelanggan dibutuhkan metode prediksi yang baik dengan perhitungan yang cukup akurat, seperti metode KNN dan Naive Bayes. KNN atau *K-Nearest Neighbor* adalah sebuah metode klasifikasi terhadap sekumpulan data berdasarkan pembelajaran data. KNN termasuk dalam *supervised learning*, dimana hasil *query instance* yang baru diklasifikasikan berdasarkan mayoritas kedekatan jarak dari kategori yang ada dalam KNN. Naive Bayes merupakan sebuah metode klasifikasi yang digunakan untuk dalam memprediksi peluang yang berakar pada *teorema bayes* dengan asumsi yang kuat. Naive Bayes adalah metode klasifikasi untuk memprediksi peluang dengan menggunakan probabilitas dan statistik yang berakar pada *teorema bayes* dengan asumsi yang kuat. Sedangkan *Rapidminer* merupakan sebuah *tools* atau *software* yang akan digunakan untuk mengolah data dan melakukan analisis terhadap yang akan digunakan dalam mengolah data dan melakukan analisis terhadap *data mining*, *text mining*, serta dan analisis prediksi pada penelitian ini.

Saat ini *GoFood* merupakan salah satu layanan terbaik yang banyak digunakan. *GoFood* juga sudah memberikan tempat bagi pengguna untuk menilai kepuasan mereka terhadap pengemudi (*driver*) dan restoran (*restaurant*). Sehingga di sini penulis akan membuat survei penilaian dengan berdasarkan tiga kategori yaitu pelayanan, kenyamanan, dan keamanan.

Hasil dari survei tersebut akan digunakan sebagai dataset untuk melakukan pengujian dengan menggunakan *data mining*. *Data mining* disebut sebagai *Knowledge Discovery in Database (KDD)*, yaitu merupakan sebuah kegiatan pengumpulan, penggunaan data, histori untuk menemukan keteraturan pola atau hubungan yang disimpan dalam sebuah database yang besar [1]. Tahapan proses *data mining* terdiri dari beberapa tahapan seperti, *select data*, *preprocessing data*, *data transformation*, dan proses *data mining* (dengan algoritma KNN dan Naive Bayes). Penggunaan algoritma KNN dan Naive Bayes bertujuan untuk membandingkan persamaan dan perbedaan beberapa objek berdasarkan kerangka pemikiran yang telah dibuat, sehingga dapat ditemukan jawaban atas permasalahan yang terjadi (Hudson, 2007). Setelah melalui beberapa tahapan pengujian, hasil yang diperoleh akan dilakukan evaluasi dan validasi dengan tujuan untuk memperoleh hasil akhir yang terbaik berdasarkan dari tingkat *accuracy* dan *recall*.

## 2. METODE PENELITIAN

Pada penelitian ini, ada beberapa tahapan yang dilakukan seperti pada Gambar 1 dibawah ini. Dimana tahapan atau metodologi penelitian ini merupakan sebuah proses atau alur mengenai bagaimana cara dalam memperoleh data dan informasi sampai proses mengolah data untuk mencari dan mendapatkan data yang akurat, yang digunakan dalam penelitian ini.

### 2.1. Identifikasi Masalah

Dalam penelitian ini, tahapan pertama yang dilakukan adalah identifikasi masalah. Tahapan ini merupakan merupakan langkah awal dalam sebuah penelitian untuk menentukan fokus permasalahan penelitian dengan tujuan agar penelitian yang dilakukan dapat terukur.

Berdasarkan hasil identifikasi permasalahan yang telah dilakukan, dapat disimpulkan sebagai berikut:

1. Bagaimana cara untuk memprediksi kepuasan pengguna layanan *GoFood* dengan menggunakan algoritma KNN dan Naive Bayes berdasarkan hasil dari akurasi.
2. Bagaimana cara mengimplementasikan algoritma KNN dan Naive Bayes untuk mengetahui tingkat kepuasan pengguna layanan *GoFood*.

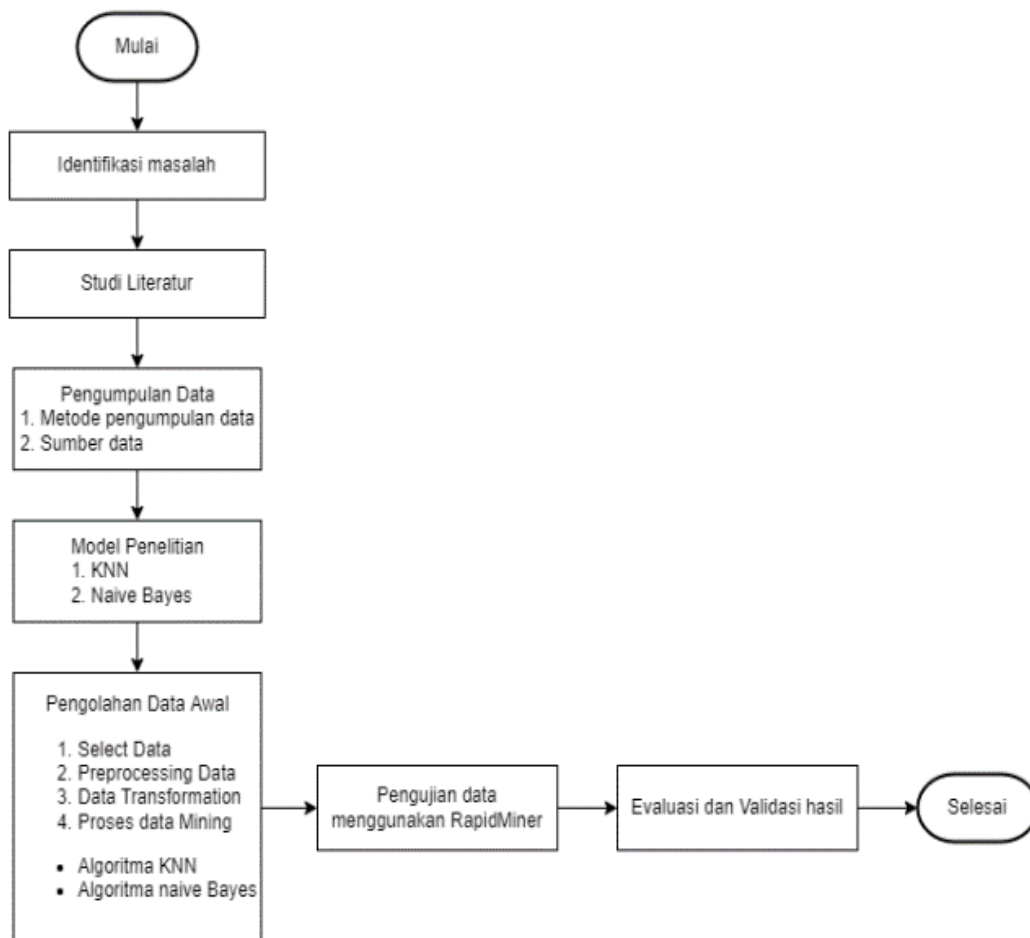
Bagaimana cara menentukan algoritma terbaik untuk melakukan prediksi kepuasan pengguna layanan *GoFood*.

### 2.2. Studi Literatur

Setelah dilakukan identifikasi masalah, tahap selanjutnya adalah penulis melakukan studi literatur untuk menyelesaikan permasalahan dengan mengumpulkan informasi dari berbagai sumber.

Sumber tersebut seperti jurnal, buku-buku, serta informasi lain yang berkaitan dengan topik dalam

penelitian ini yaitu seputar implementasi algoritma KNN dan Naive Bayes untuk melakukan prediksi.



Gambar 1. Flowchart Penelitian

## 2.3. Pengumpulan Data

### 2.3.1. Metode Pengumpulan Data

Dalam penelitian ini, data diperoleh dengan melakukan survei dengan menyebarkan kuesioner kepada para mahasiswa pengguna aplikasi Gojek terkhusus pada layanan GoFood untuk memperoleh data kepuasan pengguna layanan GoFood. Berdasarkan pengumpulan data yang telah dilakukan, berhasil memperoleh sebuah dataset yang valid yang akan digunakan untuk melakukan pengujian.

### 2.3.2. Sumber Data

Dalam penelitian ini sumber data yang digunakan merupakan data primer. Data primer sendiri merupakan data yang mengacu pada real time atau data yang terus berkembang, dimana dalam proses pengumpulan data ini diperoleh secara langsung. Untuk proses pengumpulan data ini, penulis melakukan survei dengan menggunakan kuesioner yang disebarikan kepada para mahasiswa yang menggunakan layanan GoFood.

## 2.4. Model penelitian

Dalam penelitian ini akan dilakukan pengujian dengan menggunakan metode klasifikasi data mining KNN dan Naive Bayes terhadap dataset pengguna layanan GoFood. Data yang tersedia akan diolah untuk kemudian dilihat tingkat akurasi dari setiap algoritma yang digunakan. Adapun hasil dari pengolahan data menggunakan algoritma KNN dan Naive Bayes akan menghasilkan model yang akan digunakan untuk pengujian dengan menggunakan Cross Validation dan T-Test.

### 2.4.1. Naive Bayes

Algoritma Naive Bayes merupakan algoritma klasifikasi yang dapat bekerja dengan sangat baik dibandingkan dengan model klasifikasi yang lainnya. Dibuktikan oleh (Nurhachita and Negara 2020) dalam jurnalnya yang berjudul “A Comparison Between Naive Bayes and The K-Means Clustering Algorithm for The Application of Data Mining on The Admission of New Students” yang menunjukkan bahwa Naive Bayes Classifier memiliki tingkat akurasi yang lebih

baik dibandingkan dengan model klasifikasi yang lainnya [2].

Naive Bayes merupakan algoritma klasifikasi menggunakan metode probabilitas dan statistik yang ditemukan oleh Thomas Bayes, untuk memprediksi peluang di masa depan berdasarkan pengalaman pada masa sebelumnya. Penggunaan algoritma Naive Bayes memiliki keuntungan, yaitu metode ini hanya membutuhkan data training (pelatihan data) yang kecil untuk menentukan parameter atau ukuran dalam proses klasifikasi serta dapat bekerja jauh lebih baik dan kompleks. Karena variabel independen yang digunakan hanya varian dari variabel kelas yang digunakan dalam menentukan klasifikasi, tidak keseluruhan dari matriks kovarians.

Berikut merupakan persamaan dari metode Naive Bayes:

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

Keterangan:

X : data dengan class yang belum diketahui

H : hipotesis data class spesifik

P(H) : probabilitas hipotesis H (prior probabilitas)

P(X) : probabilitas X

P(H|X) : probabilitas hipotesis H berdasarkan kondisi X (posteriori probabilitas)

P(X|H) : probabilitas X berdasarkan kondisi H

### 2.4.1. KNN

K-Nearest Neighbor adalah sebuah metode yang menerapkan algoritma supervised, dimana proses klasifikasi yang dilakukan berdasarkan mayoritas dari kategori K-NN yang diperoleh dari hasil uji sampel yang baru. Ketepatan dari algoritma K-NN ditentukan dari ada atau tidak adanya data yang tidak relevan serta dari bobot fitur terhadap klasifikasi.

Algoritma K-Nearest Neighbor memiliki tujuan untuk mengklasifikasikan atau mengelompokkan

objek data baru berdasarkan jarak data baru tersebut ke beberapa data terdekat dengan berdasarkan atribut dan sampel dari data training. Algoritma K-NN ini juga menggunakan Neighborhood Classification sebagai nilai prediksi dari sebuah instance baru.

Teknik mengolah data dengan algoritma K-NN adalah sebagai berikut [3]:

- Mulai input data training, label data training, k, data testing.
- Menghitung jarak setiap data testing ke data training.
- Menentukan k untuk data training dengan jarak terdekat dengan data.
- Melakukan testing.
- Memeriksa label dari k data.
- Menentukan label dengan frekuensi terbanyak.
- Memasukkan data testing ke kelas dengan frekuensi terbanyak.
- Label untuk semua data testing di dapat (berhenti)

Untuk menghitung jarak antara dua titik x dan y dapat menggunakan perhitungan jarak Euclidean, seperti berikut:

$$d(X_1, Y_2) = \sum_i \left| \frac{n1i}{n1} - \frac{n2i}{n2} \right| \quad (2)$$

Dimana  $X_1$ ,  $I = 1, 2$  merupakan atribut kategori dan  $n_{1i}$ ,  $n_{2i}$  mewakili frekuensi yang sesuai.

### 2.5. Pengolahan Data Awal

Tahap ini merupakan tahap awal dalam pengolahan data mining, dimana data yang telah diperoleh akan diolah ke dalam format yang dibutuhkan sesuai dengan kebutuhan, seperti pengelompokan data dan penentuan atribut data [4]. Dalam pengolahan data awal ini, ada beberapa tahapan yang akan dilakukan yaitu:

Name	Type	Missing	Statistics	Filter (34 / 34 attributes)
Label Saya puas dengan pelayanan y...	Nominal	0	Least: tidak setuju (6) Most: Setuju (512)	Values: Setuju (512), setuju (458), ... [2 more]
Nama Responden	Nominal	0	Least: yulia kristanti (1) Most: Putri (5)	Values: Putri (5), dinda (5), ... [1062 more]
Jenis kelamin	Binominal	0	Negative: Perempuan Positive: Laki-laki	Values: Perempuan (964), Laki-laki (147)
Usia Responden	Integer	0	Min: 15 Max: 38	Average: 20.893
Domisili (saat ini)	Nominal	0	Least: yogyakarta (1) Most: Jakarta (57)	Values: Jakarta (57), Bandung (41), ... [367 more]
Aplikasi GoJek (layanan GoFoo...	Nominal	0	Least: Tidak Setuju (3) Most: Sangat Setuju (826)	Values: Sangat Setuju (826), Setuju (276), ... [2 more]
Layanan GoFood mudah untuk ...	Nominal	0	Least: Tidak Setuju (4) Most: Sangat Setuju (779)	Values: Sangat Setuju (779), Setuju (323), ... [2 more]

✓ <b>Prosedur pelayanan GoFood di...</b>	Nominal	0	Least Sangat Tidak Setuju (3)	Most Sangat Setuju (579)	Values Sangat Setuju (579), Setuju (516), ...[2 more]
✓ <b>GoFood memberikan pelayana...</b>	Nominal	0	Least Sangat Tidak Setuju (3)	Most Setuju (574)	Values Setuju (574), Sangat Setuju (522), ...[2 more]
✓ <b>Driver GoFood baik dan ramah ...</b>	Nominal	0	Least Sangat Tidak Setuju (2)	Most Setuju (703)	Values Setuju (703), Sangat Setuju (376), ...[2 more]
✓ <b>Selalu ada driver GoFood yang...</b>	Nominal	0	Least Sangat Tidak Setuju (7)	Most Setuju (509)	Values Setuju (509), Sangat Setuju (402), ...[2 more]
✓ <b>Driver GoFood terdidik dan ma...</b>	Nominal	0	Least Sangat Tidak Setuju (2)	Most Setuju (656)	Values Setuju (656), Sangat Setuju (391), ...[2 more]
✓ <b>Driver GoFood cepat tanggap t...</b>	Nominal	0	Least Sangat Tidak Setuju (2)	Most Setuju (595)	Values Setuju (595), Sangat Setuju (482), ...[2 more]
✓ <b>Driver GoFood mampu mengan...</b>	Nominal	0	Least Sangat Tidak Setuju (4)	Most Setuju (671)	Values Setuju (671), Sangat Setuju (372), ...[2 more]
✓ <b>Layanan GoFood (pesan antar ...</b>	Nominal	0	Least Sangat Tidak Setuju (3)	Most Setuju (547)	Values Setuju (547), Sangat Setuju (546), ...[2 more]
✓ <b>Keamanan makanan yang dipe...</b>	Nominal	0	Least Sangat Tidak Setuju (2)	Most Setuju (576)	Values Setuju (576), Sangat Setuju (515), ...[2 more]
✓ <b>Kerahasiaan data nomor HP ko...</b>	Nominal	0	Least Sangat Tidak Setuju (3)	Most Sangat Setuju (553)	Values Sangat Setuju (553), Setuju (505), ...[2 more]
✓ <b>Kenyamanan dan keamanan tet...</b>	Nominal	0	Least Sangat Tidak Setuju (2)	Most Setuju (578)	Values Setuju (578), Sangat Setuju (521), ...[2 more]
✓ <b>Saya sangat senang dengan pe...</b>	Nominal	0	Least Sangat Tidak Setuju (2)	Most Sangat Setuju (561)	Values Sangat Setuju (561), Setuju (536), ...[2 more]
✓ <b>Saya akan merekomendasikan l...</b>	Nominal	0	Least Sangat Tidak Setuju (3)	Most Setuju (580)	Values Setuju (580), Sangat Setuju (498), ...[2 more]
✓ <b>Saya berbagi pengalaman dal...</b>	Nominal	0	Least Sangat Tidak Setuju (34)	Most Setuju (556)	Values Setuju (556), Tidak Setuju (348), ...[2 more]
✓ <b>Saya berbagi informasi dalam p...</b>	Nominal	0	Least Sangat Tidak Setuju (31)	Most Setuju (560)	Values Setuju (560), Tidak Setuju (349), ...[2 more]
✓ <b>Terdapat promo menarik pada l...</b>	Nominal	0	Least Sangat Tidak Setuju (5)	Most Setuju (580)	Values Setuju (580), Sangat Setuju (466), ...[2 more]
✓ <b>Pembayaran transaksi sesuai d...</b>	Nominal	0	Least Sangat Tidak Setuju (3)	Most Setuju (565)	Values Setuju (565), Sangat Setuju (523), ...[2 more]
✓ <b>Jasa pengantaran makanan ata...</b>	Nominal	0	Least Sangat Tidak Setuju (2)	Most Sangat Setuju (556)	Values Sangat Setuju (556), Setuju (525), ...[2 more]
✓ <b>Pengguna layanan GoFood pua...</b>	Nominal	0	Least Sangat Tidak Setuju (3)	Most Setuju (627)	Values Setuju (627), Sangat Setuju (444), ...[2 more]
✓ <b>Layanan GoFood memberi kem...</b>	Nominal	0	Least Sangat Tidak Setuju (1)	Most Sangat Setuju (680)	Values Sangat Setuju (680), Setuju (427), ...[2 more]
✓ <b>Driver GoFood menguasai jala...</b>	Nominal	0	Least Sangat Tidak Setuju (9)	Most Setuju (628)	Values Setuju (628), Sangat Setuju (332), ...[2 more]
✓ <b>GoFood merupakan layanan pe...</b>	Nominal	0	Least Sangat Tidak Setuju (3)	Most Sangat Setuju (755)	Values Sangat Setuju (755), Setuju (336), ...[2 more]
✓ <b>GoFood memiliki kredibilitas p...</b>	Nominal	0	Least Sangat Tidak Setuju (2)	Most Sangat Setuju (579)	Values Sangat Setuju (579), Setuju (522), ...[2 more]
✓ <b>GoFood sering memberikan pr...</b>	Nominal	0	Least Sangat Tidak Setuju (9)	Most Sangat Setuju (497)	Values Sangat Setuju (497), Setuju (484), ...[2 more]
✓ <b>Saya percaya informasi seputar...</b>	Nominal	0	Least Sangat Tidak Setuju (3)	Most Setuju (709)	Values Setuju (709), Sangat Setuju (319), ...[2 more]
✓ <b>Ketika saya tidak puas dengan ...</b>	Nominal	0	Least Sangat Setuju (52)	Most Tidak Setuju (616)	Values Tidak Setuju (616), Sangat Tidak Setuju (261)
✓ <b>Saya akan merekomendasikan ...</b>	Nominal	0	Least Sangat Tidak Setuju (13)	Most Setuju (606)	Values Setuju (606), Sangat Setuju (371), ...[2 more]

Gambar 2. Hasil Pengolahan Data Awal

### 2.5.1. Select Data

Pada tahap select data ini akan dilakukan seleksi atau pemilihan variabel data yang akan dianalisis, karena tidak semua data akan digunakan untuk pengujian dan hanya data tertentu saja yang akan digunakan sesuai dengan kebutuhan dan tujuan penulisan pada penelitian ini.

### 2.5.2. Preprocessing data

Dalam tahap ini, dataset yang akan digunakan akan dilakukan pembersihan data terlebih dahulu menggunakan operator Replace Missing Value yang tersedia di RapidMiner, sehingga menghasilkan data yang bebas dari missing value. Missing value akan dideteksi dengan operator ini dengan membandingkan data dengan nilai rata-rata. Namun, operator Replace Missing Value tidak dapat memperoleh data dari label kosong, sehingga disini memerlukan penggunaan operator tambahan yaitu Filter Example dengan label is not missing.

Gambar 2 diatas ini merupakan hasil dari tahapan pengolahan data awal, dimana pada kolom Missing nilai yang tercantum adalah 0 untuk setiap atribut data, hal ini menandakan bahwa tidak terdapat data yang mengalami missing value.

### 2.5.3. Data Transformation

Data Transformation adalah tahap mengubah data menjadi bentuk yang sesuai untuk selanjutnya diproses dalam data mining, dimana data sebelumnya telah melewati tahap preprocessing atau pembersihan data. Sehingga diperoleh data yang siap untuk dilakukan pengujian dengan menggunakan algoritma KNN dan Naive bayes.

Dibawah ini merupakan data yang akan digunakan dalam melakukan pengujian, yang terdiri dari 10 atribut dengan 1.000 sampel yang telah diisi oleh responden dan 1 label dengan tipe data binominal (tipe data yang terdiri dua suku kata). Label yang digunakan tersebut terdapat pada atribut "Kepuasan Pengguna" dengan 98% pengguna merasa puas dan sisanya merasa tidak puas.

Tabel 1. Data Atribut Penelitian

Atribut	Keterangan	Status
Nama Responden	Nama Mahasiswa	Atribut
Jenis kelamin	Laki-laki atau Perempuan	Atribut
Usia Responden	Usia Mahasiswa	Atribut
Domisili	Tempat tinggal responden	Atribut
Kualitas Layanan	Tingkat layanan yang berkaitan dengan harapan dan kebutuhan pengguna layanan <i>GoFood</i> Dengan skala penilaian (1-4)	Atribut
Citra Merek	Persepsi yang muncul di benak konsumen atau pengguna layanan ketika mengingat suatu merek dari produk tertentu, yaitu	Atribut

	produk layanan <i>GoFood</i> yang diberikan oleh <i>Gojek</i> Dengan skala penilaian (1-4)	
<i>Electronic Word of Mouth</i>	Media komunikasi untuk berbagi informasi mengenai suatu produk atau jasa yang telah dikonsumsi antar konsumen yang awalnya tidak saling mengenal dan bertemu sebelumnya yang disampaikan secara elektronik atau melalui sosial media Dengan skala penilaian (1-4)	Atribut
Kepuasan Pengguna	Perasaan senang atau kecewa seseorang yang berasal dari perbandingan antara kesannya terhadap kinerja (hasil) suatu produk layanan <i>GoFood</i> dan harapan-harapannya terhadap layanan tersebut Dengan skala penilaian (1-4)	Atribut
Loyalitas Pengguna	Wujud kesetiaan konsumen untuk menggunakan suatu produk atau jasa secara terus menerus karena merasa puas dengan layanan yang diberikan Dengan skala penilaian (1-4)	Atribut
Kepuasan Pengguna	Tingkat kepuasan pengguna layanan <i>GoFood</i> (Puas/Tidak Puas)	Label

Tabel 2. Nilai skala likert kuesioner

Nilai	Kategori
1	Sangat Tidak Setuju
2	Tidak Setuju
3	Setuju
4	Sangat Setuju

### 2.6. Pengujian Dengan RapidMiner

Untuk pengolahan data dalam penelitian ini dengan menggunakan RapidMiner. RapidMiner merupakan sebuah software atau perangkat lunak yang digunakan untuk proses pengolahan data. RapidMiner merupakan sebuah perangkat lunak yang bersifat terbuka (open source), yang digunakan untuk melakukan proses analisis data mining, text mining, dan analisis prediksi. Dalam proses ini, RapidMiner menggunakan berbagai teknik deskriptif dan prediksi untuk dapat membuat sebuah keputusan yang paling baik [5].

RapidMiner menyediakan GUI (Graphic User Interface) untuk merancang sebuah pipeline analitis, yang akan menghasilkan file XML (Extensible Markup Language) yang mendefinisikan proses analitis keinginan pengguna untuk diterapkan ke dalam bentuk data. File ini kemudian dibaca oleh RapidMiner untuk menjalankan analisis secara otomatis [5].

Dalam melakukan pengolahan data, RapidMiner memudahkan pengguna karena dapat melakukan perhitungan data dengan jumlah yang banyak



menggunakan operator-operator untuk memodifikasi data. Selain itu, RapidMiner juga merupakan salah satu software pilihan yang dapat melakukan ekstraksi data dengan metode data mining, dimana hasil yang diperoleh dapat ditampilkan secara visual dengan grafik.

**2.7. Evaluasi dan Validasi hasil**

Untuk proses validasi ini menggunakan Confusion Matrix, yang merupakan sebuah alat yang digunakan untuk melakukan analisis tingkat accuracy, precision, dan recall terhadap suatu model dari algoritma. Dengan menggunakan confusion matrix akan membantu peneliti dalam memberikan rincian klasifikasi terhadap kelas yang diprediksi. Sedangkan accuracy merupakan nilai dari ketepatan presentasi antara nilai yang diprediksi dengan nilai yang sebenarnya. Precision merupakan nilai akurasi dari kelas yang sudah diprediksi, dan recall merupakan nilai tingkat keberhasilan dari algoritma yang digunakan.

Dibawah ini merupakan model dari confusion matrix yaitu confusion table.

Tabel 3. Confusion Matrix

Classificati-on	Predicted Class		
		Positive	Negative
Kelas yang diamati	Positive	TP	FP
	Negative	FN	TN

Model untuk tabel diatas dapat dihitung menggunakan rumus sebagai berikut:

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

Keterangan:

TP (True Positive) = jumlah kasus positif yang diklasifikasikan sebagai positif

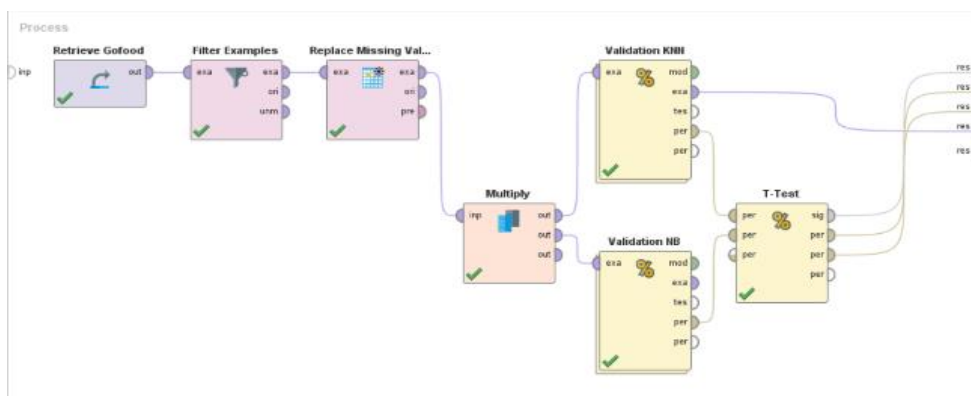
FP (False Positive) = jumlah kasus negatif yang diklasifikasikan sebagai positif

TN (True Negative) = jumlah kasus positif yang diklasifikasikan sebagai negatif

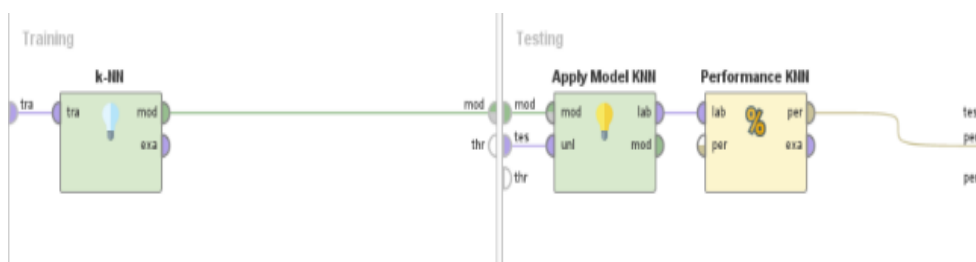
FN (False Negative) = jumlah kasus negatif yang diklasifikasikan sebagai negatif

Pada tahap ini dilakukan untuk melakukan perbandingan antara algoritma KNN dan Naive Bayes menggunakan Cross Validation dengan 10-fold cross validation, serta dengan melakukan pengujian T-Test dengan alpha ( $\alpha=0,05$ ). Proses validasi dengan menggunakan Cross Validation ini dipilih karena dinilai mampu untuk memberikan hasil yang lebih baik dibandingkan dengan metode validasi yang lain. Sedangkan penggunaan t-test dalam penelitian ini digunakan untuk melakukan pengujian hipotesis menggunakan satu objek penelitian yang menghasilkan dua perlakuan yang berbeda.

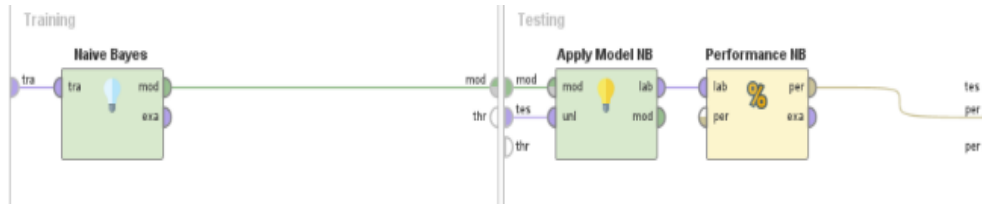
Untuk desain dari model pengukuran performance menggunakan Cross Validation dan T-Test dapat dilihat pada Gambar 3 di bawah ini. Sedangkan Gambar 4 dan Gambar 5 menunjukkan model validation KNN dan Naive Bayes.



Gambar 3. Desain Model Pengujian



Gambar 4. Model Validation KNN



Gambar 5. Model Validation Naive Bayes

### 3. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan dataset primer yang diperoleh melalui survei dengan menyebarkan kuesioner kepada para mahasiswa, yang selanjutnya akan dilakukan klasifikasi menggunakan algoritma KNN dan Naive Bayes. Model dalam penelitian ini akan diuji dengan menggunakan tools RapidMiner dengan versi 9.10 untuk memperoleh nilai Accuracy, Precision, Recall, dan T-Test.

#### 3.1. Hasil Pengujian Menggunakan Algoritma KNN

Hasil pengujian dengan menggunakan algoritma KNN dapat dilihat tabel dibawah ini. Dari Tabel 4 tersebut menunjukkan bahwa algoritma KNN mendapatkan hasil *Accuracy* sebesar 98,80%, *Precision* 98,80, dan *Recall* 100,00%.

Tabel 4. Hasil Pengujian Menggunakan Algoritma KNN  
Accuracy: 98,80%

	true Setuju	true Tidak Setuju	Class Precision
pred. Setuju	988	12	98,80%
pred. Tidak Setuju	0	0	0,00%
Class Recall	100,00%	0,00%	

#### 3.2. Hasil Pengujian Menggunakan Algoritma Naive Bayes

Untuk hasil pengujian dengan menggunakan algoritma Naive Bayes berdasarkan Tabel 5 menunjukkan bahwa hasil *Accuracy* 94,10%, *Precision* 99,57%, dan *Recall* 94,43%.

Tabel 5. Hasil Pengujian Menggunakan Algoritma Naive Bayes  
Accuracy: 94,10%

	true Setuju	true Tidak Setuju	Class Precision
pred. Setuju	933	4	99,57%
pred. Tidak Setuju	55	8	12,70%
Class Recall	94,43%	66,67%	

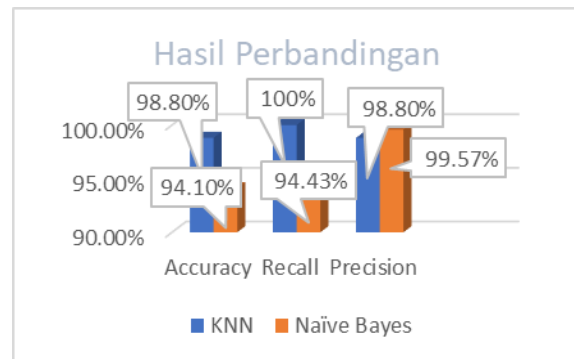
#### 3.3. Perbandingan Algoritma KNN dan Naive Bayes

Berdasarkan dari hasil pengujian yang telah dilakukan, diperoleh hasil perbandingan dari algoritma

KNN dan Naive Bayes yang dapat dilihat pada Gambar 6 dibawah ini. Dapat dilihat dengan jelas bahwa algoritma KNN memperoleh nilai tertinggi

dibandingkan dengan algoritma Naive Bayes, yang dapat dilihat dari tingkat Accuracy dan Recall. Dengan nilai algoritma KNN sebesar 98,80% untuk Accuracy dan 100,00% untuk Recall.

Untuk Naive Bayes memperoleh nilai Accuracy sebesar 94,10% dan Recall sebesar 94,43%. Sedangkan untuk tingkat Precision algoritma Naive bayes memperoleh nilai lebih tinggi dibandingkan dengan algoritma KNN, dengan nilai 99,57% untuk Naive Bayes dan 98,80% untuk KNN, dan dengan selisih sebesar 0,77%.



Gambar 6. Hasil Perbandingan Accuracy, Recall, Precision

#### 3.4. Hasil Pengujian T-Test

Pengujian T-Test ini dilakukan untuk menentukan dan mengetahui probabilitas nilai dari algoritma KNN dan Naive Bayes menggunakan Cross Validation apakah memperoleh hasil yang signifikan atau tidak. Dalam pengujian T-Test dapat dikatakan signifikan jika nilai alpha ( $\alpha$ ) yang diperoleh lebih kecil dari  $\alpha=0,050$ . Berdasarkan pengujian T-Test yang telah dilakukan diperoleh hasil seperti Tabel 1 dibawah ini, dengan nilai  $\alpha=0,000$  sehingga dapat dikatakan bahwa hasil yang diperoleh merupakan signifikan.

Tabel 1. Hasil Pengujian T-Test

A	B	C
0,988+/-0,004	0,988+/-0,004	0,000
0,941+/-0,033		

### 4. DISKUSI

Dalam melakukan penelitian ini, sangat dibutuhkan beberapa sumber sebagai referensi dan data pendukung yang dapat berupa teori. Data-data ini yang akan dijadikan sebagai bahan dalam mengembangkan dan menjalankan penelitian ini. Selain itu, data dan sumber ini bertujuan juga sebagai pembandingan antara penelitian terdahulu dengan



penelitian yang terbaru yang akan dilakukan oleh penulis.

Pada penelitian terdahulu, yaitu oleh (Hermanto and Romadhoni n.d.) dalam jurnal yang berjudul “Penerapan Metode Naive Bayes Untuk Prediksi Kepuasan Pelanggan Studi Kasus Bengkel Win Motor” menjelaskan bahwa penggunaan teknik mining dapat mengatasi masalah Bengkel Win Motor untuk mampu bersaing pada era globalisasi. Dengan melakukan prediksi menggunakan metode Naive Bayes untuk mengklasifikasikan kepuasan terhadap pelayanan yang diberikan oleh Bengkel Win Motor, sehingga nantinya akan di dapatkan besarnya tingkat akurasi kepuasan pelanggan dan membantu Bengkel Win Motor dalam mengevaluasi serta meningkatkan kualitas pelayanan terhadap para konsumen [5].

Menurut (Apandi and Sugianto 2019) dalam penelitiannya yang berjudul “Algoritma Naive Bayes Untuk Prediksi Kepuasan Perekaman e-KTP”. Dalam penelitian ini bertujuan untuk mengetahui sejak dini mengenai tingkat kepuasan masyarakat terhadap pelayanan perekaman e-KTP dengan menggunakan data mining dan *skala likert*. Untuk melakukan pengujian dengan menggunakan hasil model data latih, sehingga akan didapatkan akurasi dari masing-masing algoritma yang digunakan, yaitu Naive Bayes dan Decision Tree (sebagai algoritma pembandingan) [6].

Menurut (Sadewo et al. 2019) dalam penelitiannya yang berjudul “Algoritma Naive Bayes Dalam Memprediksi Kepuasan Nasabah”. Dengan penelitiannya untuk memprediksi kepuasan nasabah terhadap kualitas pelayanan Bank BTN Cabang Pematangsiantar, yang ditinjau dari aspek *Tangible* (Bukti Nyata), *Reliability* (Keandalan), *Assurance* (Jaminan), dan *Responsiveness* (Ketanggapan). Menggunakan *software RapidMiner* sehingga akan didapatkan sebuah model aturan yang akan diimplementasikan dengan algoritma Naive Bayes [7].

Dari beberapa sumber penelitian tersebut, adapun kesamaan dari penelitian terdahulu dengan penelitian ini adalah penggunaan *RapidMiner* sebagai alat untuk melakukan pengolahan data. Dalam penelitian ini menggunakan metode KNN dan Naive Bayes, dengan studi kasus serta proses perolehan data yang berbeda dengan penelitian terdahulu. Penggunaan dua algoritma ini bertujuan sebagai pembandingan antara nilai akurasi yang diperoleh setelah melakukan pengolahan data, sehingga dapat diketahui penggunaan algoritma yang terbaik [8].

## 5. KESIMPULAN DAN SARAN

Dari hasil pengujian yang dilakukan, dapat disimpulkan bahwa algoritma KNN merupakan algoritma yang terbaik dibandingkan dengan Naive Bayes, walaupun hasil *Accuracy* yang diperoleh dari kedua algoritma tersebut keduanya memiliki hasil dengan nilai yang baik. Namun KNN merupakan algoritma terbaik untuk digunakan dalam prediksi tingkat kepuasan pengguna layanan *GoFood* pada

mahasiswa, dengan tingkat *Accuracy* sebesar 98,80% dan *Recall* 100% sedangkan Naive Bayes memperoleh nilai *Accuracy* sebesar 94,10% dan *Recall* sebesar 94,43%.

Adapun saran untuk pengembangan penelitian selanjutnya adalah sebagai berikut:

1. Menggunakan dataset yang berbeda, yang dapat dilakukan dengan melakukan survei untuk memperoleh dataset terbaru atau dengan mengambil dataset dari *UCI Machine Learning Repository* dan *Kaggle*.
2. Dapat menggunakan algoritma yang berbeda dari penelitian ini, maupun menambahkan beberapa algoritma lain untuk memperoleh beberapa hasil dari nilai pengujian.
3. Dapat menggunakan atau menambahkan pengujian seperti *ROC Curve*.

## DAFTAR PUSTAKA

- [1] S. Agarwal, Data mining: Data mining concepts and techniques. 2014.
- [2] N. Nurhachita and E. S. Negara, “A Comparison Between Naïve Bayes and The K-Means Clustering Algorithm for The Application of Data Mining on The Admission of New Students,” *J. Intelett. Keislaman, Sos. dan Sains*, vol. 9, no. 1, pp. 51–62, 2020, doi: 10.19109/intelektualita.v9i1.5574.
- [3] V. Novita Sari and D. Suranti, “IMPLEMENTASI METODE BAYES DALAM EVALUASI KEPUASAN MAHASISWA TERHADAP LAYANAN UNIVERSITAS,” *Sintech J.*, vol. 4, 2021, doi: <https://doi.org/10.31598/sintechjournal.v4i1.569>.
- [4] Aprilla Dennis, *Belajar Data Mining dengan RapidMiner*, vol. 5, no. 4. 2013.
- [5] B. Hermanto and A. Romadhoni, “PENERAPAN METODE NAÏVE BAYES UNTUK PREDIKSI KEPUASAN PELANGGAN STUDI KASUS BENGKEL WIN MOTOR,” *SIGMA - J. Teknol. Pelita Bangsa*, vol. 9, 2019.
- [6] T. H. Apandi and C. A. Sugianto, “Algoritma Naive Bayes untuk Prediksi Kepuasan Pelayanan Perekaman e-KTP (Naive Bayes Algorithm for Satisfaction Prediction of e-ID Card Recording Service),” *JUITA J. Inform.*, vol. 7, no. 2, pp. 125–128, 2019, doi: 10.30595/juita.v7i2.3608.
- [7] M. G. Sadewo, A. Perdana Windarto, I. S. Damanik, S. Tunas, and B. Pematangsiantar, “Prosiding Seminar Nasional Riset Information Science (SENARIS) Algoritma Naïve Bayes Dalam Memprediksi Kepuasan Nasabah,” 2019, p. 318.
- [8] A. Prayoga Permana, K. Ainiyah, and K. Fahmi Hayati Holle, “Analisis Perbandingan

- Algoritma Decision Tree, kNN, dan Naive Bayes untuk Prediksi Kesuksesan Start-up,” *JISKA*, vol. 6, no. 3, pp. 178–188, 2021, [Online]. Available: <https://www.kaggle.com/manishkc06/startup-success-prediction>.
- [9] D. T. Larose and C. D. Larose, *DISCOVERING KNOWLEDGE IN DATA An Introduction to Data Mining Second Edition Wiley Series on Methods and Applications in Data Mining*. 2014.
- [10] N. Bayes Yunitasari et al., “Optimasi Backward Elimination untuk Klasifikasi Kepuasan Pelanggan Menggunakan Algoritme k-Nearest Neighbor (k-NN) dan,” *Technomedia J.*, vol. 6, 2021, doi: 10.33050/tmj.v6i1.
- [11] Yoga Religia and A. Amali, “Perbandingan Optimasi Feature Selection pada Naïve Bayes untuk Klasifikasi Kepuasan Airline Passenger,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 3, pp. 527–533, 2021, doi: 10.29207/resti.v5i3.3086.
- [12] Yoga Religia, Agung Nugroho, and Wahyu Hadikristanto, “Klasifikasi Analisis Perbandingan Algoritma Optimasi pada Random Forest untuk Klasifikasi Data Bank Marketing,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 1, pp. 187–192, 2021, doi: 10.29207/resti.v5i1.2813.
- [13] L. N. Martin, “Comparison of C4.5 and Naïve Bayes Algorithms for Assessment of Public Complaints Services,” *JITE (Journal Informatics Telecommun. Eng.* Available, vol. 3, no. 2, pp. 266–271, 2020.
- [14] T. Safitri, “Penerapan Algoritma Naïve Bayes Untuk Penentuan Calon Penerimaan Beasiswa Pada Sd Negeri 6 Ketapang,” *J. Inform. dan Sist. Inf.*, vol. 06, no. 01, pp. 43–52, 2020.
- [15] A. Heriyanto, “Penerapan Metode K-Nearest Neighbor (KNN) Untuk Klasifikasi Stanting Pada Balita,” *Publ. Ilm. Univ. Muhammadiyah Jember*, 2021.
- [16] F. Yulia, Lamsah, and Periyadi, *BUKU MANAJEMEN PEMASARAN\_compressed.pdf*, no. April. 2019.
- [17] Budi Indrawati, M. Wijayanti, and T. Yuniati, “Analisis Kualitas Layanan dan Kepercayaan Terhadap Kepuasan Konsumen yang Berimplikasi pada Loyalitas Pelanggan Gofooddi Kota Bekasi,” *Optim. J. Ekon. dan Kewirausahaan*, vol. 15, no. 2, p. 61, 2021.
- [18] I. M. Pasaribu, “ANALISIS PELAYANAN JASA APLIKASI GO JEK TERHADAP KEPUASAN KONSUMEN,” *J. Pros.*, no. 1, pp. 21–29, 2021.
- [19] A. Muzaki and A. Witanti, “Sentiment Analysis of the Community in the Twitter To the 2020 Election in Pandemic Covid-19 By Method Naive Bayes Classifier,” *J. Tek. Inform.*, vol. 2, no. 2, pp. 101–107, 2021, doi: 10.20884/1.jutif.2021.2.2.51.
- [20] Y. I. Kurniawan, A. Fatikasari, M. L. Hidayat, and M. Waluyo, “Prediction for Cooperative Credit Eligibility Using Data Mining Classification With C4.5 Algorithm,” *J. Tek. Inform.*, vol. 2, no. 2, pp. 67–74, 2021, doi: 10.20884/1.jutif.2021.2.2.49.
- [21] A. Pamuji, H. S. Setiawan, and B. K. Islam, “LINEAR REGRESSION FOR PREDICTION OF EXCESSIVE PERMISSIONS,” vol. 3, no. 2, pp. 467–474, 2022.
- [22] I. Agustina, J. Eska, and I. R. Harahap, “APPLICATION OF C4 . 5 ALGORITHM FOR DETERMINATION OF THE COMMUNITY OF RECIPIENTS OF PROSPEROUS FAMILY CARDS IN THE PENERIMA KARTU KELUARGA SEJAHTERA PADA DESA SUKARAMAI,” vol. 3, no. 2, pp. 1–7, 2022.