

**LAPORAN PROPOSAL**  
**PENERAPAN KELASIFIKASI KEPUASAN PELANGGAN**  
**GO-JEK MENGGUNAKAN METODE ALGORITMA**  
**NAIVE BAYES**



Disusun Oleh:

Liem Riki Wibowo

A11.2020.12486

**PROGRAM STUDI TEKNIK INFORMATIKA**  
**FAKULTAS ILMU KOMPUTER UNIVERSITAS**  
**DIAN NUSWANTORO**

2022

## ABSTRAK

Kepuasan pelanggan adalah suatu kondisi di mana harapan, persyaratan dan kebutuhan pelanggan terpenuhi. Hal ini dapat meningkatkan operasional bisnis karena dapat memberikan dampak positif berupa peningkatan keuntungan dan pujian positif terhadap layanan yang diberikan. Saat ini, pelanggan GO-JEK semakin sadar akan perlunya layanan berkualitas yang diberikan oleh pengemudi. Pelanggan akan membandingkan pelayanan yang diharapkan dengan pelayanan yang diterima. Tujuan dari penelitian ini adalah melakukan pemeringkatan kepuasan pelanggan GO-JEK dengan menerapkan teknik data mining menggunakan metode algoritma Naïve Bayes. Metode naïve Bayesian diterapkan dalam penelitian ini untuk menghitung probabilitas tertinggi untuk variabel independen yang teridentifikasi. Seperti pengaplikasian, ketepatan waktu, kenyamanan berkendara, kemudahan penggunaan dan harga. Klasifikasi ini menunjukkan apakah pelanggan GO-JEK puas atau tidak. Hasil penelitian ini berupa suatu sistem yang dapat membantu masyarakat dalam melakukan evaluasi terhadap layanan GO-JEK, sehingga GO-JEK dapat mempertimbangkan peningkatan layanan kepada pelanggan.

Kata kunci: Data Mining, Kepuasan Pelanggan, Naïve Bayes, Klasifikasi

# BAB I

## PENDAHULUAN

### 2.1 Latar Belakang

Perkembangan teknologi informasi pada saat ini sangat pesat, teknologi lalu lintas pun juga ikut berkembang. Seperti kebanyakan layanan transportasi online di Indonesia, termasuk Go-Jek. Go-Jek semakin populer dan menjadi sarana transportasi umum yang digemari karena praktis dan cepat. Saat ini kesadaran pelanggan Go-Jek akan perlunya mengapresiasi kualitas pelayanan yang diberikan oleh pengemudi semakin meningkat. Tingkat kepuasan pelanggan sendiri dapat dipengaruhi oleh kualitas pelayanan yang diberikan perusahaan kepada pelanggan (Santoso, 2014). Pengertian kepuasan pelanggan sendiri adalah kondisi terpenuhinya harapan, keinginan, dan kebutuhan konsumen. Setiap pelanggan membandingkan pelayanan yang diharapkan dengan pelayanan yang diterima. Kepuasan pelanggan sangat penting bagi suatu bisnis untuk meningkatkan keuntungan dan menerima pujian positif atas layanan yang diberikan. Oleh karena itu, dengan mengetahui loyalitas pelanggan, pelaku bisnis dapat mengidentifikasi pelanggan setia dan pelanggan yang akan beralih ke bisnis lain (Wijaya & Girsang, 2016).

Loyalitas pelanggan dapat membantu perusahaan untuk membuat perencanaan ditahun mendatang untuk menangani penurunan pelanggan (Moedjiono, Isak, & Kusdaryono, 2016). Sedangkan pendapat pelanggan terkadang tidak sesuai dengan skala penilaian dari perusahaan (Sipayung, Maharani & Zafanya, 2016). Media social menjadi salah satu sarana untuk membahas isu dan mengungkapkan pendapat pelanggan (susilawati, 2016). Penelitian terdahulu telah memberikan gambaran mengenai solusi yang dapat dilakukan untuk permasalahan yang sejenisnya. (Soepardi & Permata, 2015) Mengatakan dalam penelitiannya terkait dengan penilaian kepuasan masyarakat terhadap kinerja pemerintah, bahwa penilaian kepuasan masyarakat dalam melakukan analisa sentiment yang diutarakan melalui media social. Metode data mining naïve bayes yang diterapkan dalam penelitian ini untuk mengkalsifikasikan sentiment positif dan negatif.

Saat ini Go-Jek hanya bisa mengetahui tingkat kepuasan pelanggan melalui halaman komentar akun Go-Jek di media sosial. Oleh karena itu, para peneliti telah menerapkan teknik penambangan untuk mengatasi masalah tersebut, termasuk dalam bentuk sistem penilaian kepuasan pelanggan yang dapat membantu perusahaan Go-Jek menganalisis ulasan harga terhadap kepuasan pelanggan.

Metode yang digunakan adalah metode Naive Bayes yang merupakan metode perhitungan probabilitas yang paling sederhana dan memiliki akurasi yang sangat tinggi jika diterapkan pada database yang berisi data berukuran besar. (Wati, 2016). Menurut (Liu, 2016) membandingkan dengan metode klasifikasi yang lain, tingkat kesalahan yang minimum dimiliki Naive Bayes ini. Model Naive Bayes memprediksi probabilitas pada masa depan berdasarkan hasil perhitungan yang sudah dilakukan di masa lalu. Penelitian ini bertujuan untuk membangun sebuah sistem mengklasifikasi kepuasan dari pelayanan yang diberikan Driver Go-Jek kepada pelanggan dengan menggunakan metode Naive Bayes, serta mengetahui seberapa besar tingkat akurasi dalam membuat klasifikasi kepuasan pelanggan dalam menggunakan jasa Go-Jek. Sehingga dapat membantu dalam meningkatkan kualitas pelayanan Go-Jek dan mendapatkan hasil bisa atau tidaknya Go-Jek menjadi jasa antar jemput yang terpercaya. Menimbang dari latar belakang masalah di atas maka penulis mengambil penelitian skripsi ini dengan judul **“Penerapan Klasifikasi Kepuasan Pelanggan GO-JEK menggunakan metode algoritma Naive Bayes”**

## 2.2 Rumusan Masalah

Berdasarkan identifikasi masalah, maka dirumuskan permasalahan,

- Bagaimana memuaskan pelanggan pada perusahaan dengan kriteria dan proses yang sesuai?
- Bagaimana memanfaatkan penggalan data untuk menghasilkan kepuasan pada pelanggan dengan tingkat keakuratan yang tinggi.
- Bagaimana melihat hasil tingkat kinerja karyawan pada delivery perusahaan.

## 2.3 State Of Arts

Berikut beberapa penelitian mengenai data mining atau metode penelitian yang dijadikan acuan:

- Analisis Algoritma Naïve Bayes Untuk Sistem Klasifikasi Status Kepuasan Pelanggan Go-Jek (Soepriadi, A, Permata, M. 2016).

Tujuan dari penelitian ini adalah menerapkan metode algoritma Naive Bayes untuk menentukan peringkat kepuasan pelanggan pada layanan pick up dan drop off menggunakan aplikasi Go-Jek. Analisis proses klasifikasi status kepuasan pelanggan menggunakan variabel target/kategori yang sesuai dengan kategori pick-up/drop-off di aplikasi Go-Jek berdasarkan aspek non-moneter. Berdasarkan hasil penelitian yang dilakukan, model yang dilatih dengan algoritma Naive Bayes memberikan hasil dengan akurasi yang cukup baik yaitu 85,80%. Berdasarkan keyakinan klasifikasi nilai AUC yang diperoleh dari algoritma Naive Bayes adalah 0,930 sehingga tergolong klasifikasi sangat baik. Oleh karena itu, dapat disimpulkan bahwa algoritma Naive Bayes dapat diterapkan untuk menentukan peringkat kepuasan pelanggan.

- Pengembangan Sistem Identifikasi pelanggan go-jek Naive Bayes Dalam kepuasan pelanggan (I Wayan Supriana, dkk. 2018).

Dalam penelitian ini, kemajuan teknologi informasi digunakan untuk menentukan klasifikasinya. Sistem yang dibangun akan menentukan tingkat akurasi berdasarkan indikator kepuasan pelanggan. Metode identifikasi untuk menentukan kepuasan pelanggan menggunakan analisis Bayesian, khususnya Naive Bayes Classification. Hasil yang diperoleh berdasarkan analisis dan implementasi bahwa sistem yang dibangun mampu menentukan tingkat kepuasan pelanggan sebesar 75% berdasarkan data uji yang digunakan,

sehingga akan memudahkan dalam penyampaian sasaran program manajemen kepuasan berdasarkan penyampaian layanan.

- Penerapan Metode Naive Bayes Dalam Klasifikasi Kesalahan Driver (Liu, dkk. 2016).

Penerapan data mining untuk menentukan klasifikasi kesalahan mengemudi dapat digunakan untuk memprediksi apakah seseorang puas atau tidak dengan menggunakan algoritma Naive Bayes. Metode pengumpulan data yang digunakan dalam penelitian ini berupa wawancara kepada kepala departemen/driver Go-Jek itu sendiri kemudian dilakukan pengumpulan data berdasarkan data kantor atau karyawan Go-Jek.Go-Jek. Peneliti mengumpulkan 60 data latih dan data uji dengan menggunakan enam kriteria yaitu status mengemudi, jumlah tanggungan, jumlah karyawan, kondisi kendaraan, total pendapatan, dan status pemilik kendaraan. Semoga hasil penelitian ini dapat membantu pelanggan dalam menentukan apakah tingkat kepuasan pengemudi ada atau tidak.

Dari seluruh penelitian dan metode di atas, terbukti bahwa penggunaan metode Naive Bayes memiliki banyak keunggulan dalam hal prediksi. Dengan akurasi yang tinggi, metode Naive Bayes dipilih untuk digunakan dalam penelitian ini.

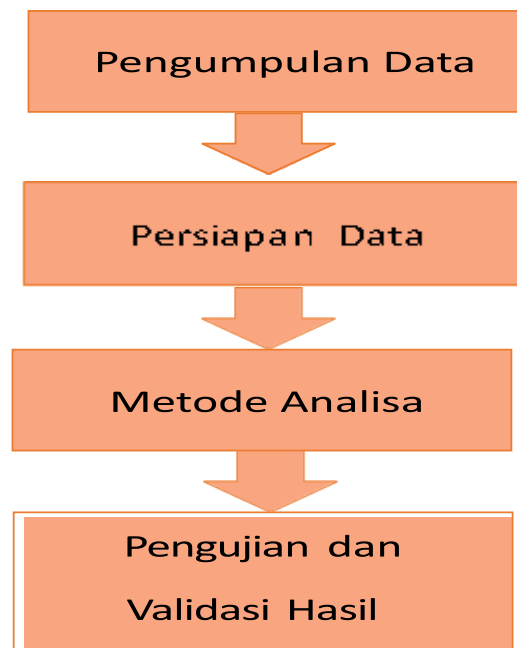
## BAB II

### METODE PENELITIAN

Metode penelitian merupakan salah satu rangkaian kegiatan ilmiah untuk mendapatkan data yang valid dengan langkah-langkah yang teratur dan sistematis (Sugiyono, 2016). Penelitian ini dilakukan untuk menemukan, mengembangkan, dan membuktikan suatu pengetahuan tertentu sehingga dapat ditarik kesimpulan berdasarkan faktor- faktor yang mempengaruhi.

#### 2.1 Tahapan Penelitian

Pada penelitian ini data yang digunakan dari data rumah tangga di desa Pasirsari. Data tersebut akan diolah menggunakan metode klasifikasi dengan algoritma Naive Bayes.



Gambar 3.1 Tahapan Penelitian

Tahapan pengumpulan data dan tahapan persiapan data berada di dalam metodologi data mining dan tahap pengujian akan di uraikan pada bab selanjutnya.

## 2.2 Metodologi Data Mining

Data mining metodologi yang sudah banyak digunakan dalam pengembangan data mining yakni CRISP-DM (Cross Industry Standard Process Model for Data Mining), terdiri dari enam fase yaitu pemahaman bisnis (Business Understanding), pemahaman data (data understanding), persiapan data (data preparation), pemodelan (modelling), evaluasi (Evaluation), dan penyebaran (deployment).

Pemahaman Bisnis (Business Understanding)

Objek Penelitian

Dalam penelitian ini penulis mengambil data Struktur perusahaan Go-Jek dari salah satu karyawan Go-Jek yang beralamatkan di kp.kandang roda Rt.002/004 Kec.Cikarang Selatan Kab.Bekasi.

Adapun struktur Perusahaan Go-Jek sebagai berikut :

Gambar 3.2 Struktur Organisasi Perusahaan Go-Jek



Tugas dan tanggungjawab dari masing-masing jabatan yaitu :

1. Direktur Utama

- Memelihara ketentraman dan ketertiban Karyawan Pada Pelanggan.
- Mengelola keuangan dan aset Perusahaan.
- Menyelenggarakan administrasi Perusahaan dengan baik.
- Menyelesaikan perselisihan Pelanggan Pada Driver.

2. Wakil Direktur

- Membahas dan menyepakati rencana peraturan Perusahaan Go-Jek.
- Menampung dan menyalurkan aspirasi Karyawan.
- Melakukan pengawasan kinerja Karyawan.

3. Karyawan IT

- Merancang Aplikasi Dengan Desain Yang Menarik.
- Menyaring Data Dari Pelanggan Yang Sudah Order Melalui Aplikasi.

- Mempersiapkan bahan untuk laporan Agar Bisa Diakses Melalui Aplikasi.

#### 4. Customer Service

- Melaksanakan administrasi kependudukan.
- Mempersiapkan bahan-bahan penyusunan perencanaan peraturan desa dan keputusan kepala desa.
- Melaksanakan kegiatan administrasi pertanahan.
- Melaksanakan kegiatan pencatatan monografi desa.

#### 5. Karyawan Akutansi

- Mengelola administrasi keuangan Perusahaan
- Membuat laporan pertanggungjawaban keuangan.
- Mempersiapkan bahan penyusunan Karyawan.

#### 6. Tukang Ojek

- Menerima Pesana Dari Pelanggan Lewat Aplikasi.
- Menyiapkan atau Menuju Kelokasi Yang Sudah Di Tentukan.

- Mengantar Pelanggan Pada Lokasi Yang Sudah Dipesan Pada Aplikasi Go-Jek.

### 3.3 Pemahaman Data (Data Understanding)

#### 3.3.1 Jenis Data dan Sumber Data

Pada penelitian ini menggunakan data yang berasal dari data perusahaan dengan jumlah data sebanyak 150 karyawan. Terdapat 4 atribut untuk mempengaruhi penentuan status kepuasan pelanggan go-jek yaitu :Status Aplikasi, ketepatan Waktu, Kenyamanan Dalam Berkendara, Keramahan.

##### 1. Jenis Data

Pada penelitian ini akan dilakukan menggunakan dua jenis data yaitu : data Kualitatif dan data kuantitatif.

##### 1. Data Kualitatif

Data yang dinyatakan dalam bentuk kata-kata atau bukan dalam bentuk angka. Data ini biasanya menjelaskan karakteristik atau sifat. Seperti pekerjaan (wiraswasta, supir, karyawan swasta),dll.

##### 2. Data kuantitatif

Data yang berisi keterangan yang dinyatakan dalam bentuk bilangan dan bersifat variabel.

##### 2. Sumber Data

Untuk mendapatkan data-data yang dapat menunjang penelitian ini, peneliti menggunakan beberapa metode pengumpulan data sebagai berikut:

##### 1. Sumber Data Primer

a) Metode Interview atau Wawancara

Metode wawancara dilakukan dengan cara melakukan wawancara terhadap sekretaris rukun warga untuk mendapatkan data rumah tangga.

b) Metode Studi Literatur

Pada metode studi literatur, penulis melakukan pencarian data berdasarkan sumber-sumber tertulis, baik berupa buku, arsip, jurnal, maupun dokumen-dokumen lain yang relevan dengan permasalahan yang berkaitan dengan algoritma Naive Bayes.

2. Sumber Data Sekunder

Dalam penulisan penelitian ini penulis tidak hanya menggunakan metode pengumpulan data secara wawancara dan studi literatur. Tetapi menggunakan pengumpulan data yang diperoleh langsung dari sumber objek penelitian. Dalam hal ini penulis mendapatkan data dari arsip yang disimpan di kantor desa.

Pengumpulan data dilakukan dengan mempertimbangkan penggunaan data berdasarkan jenis dan sumbernya. Data yang digunakan dalam penelitian ini adalah data sekunder, data yang penulis dapatkan adalah data Perusahaan aplikasi go-jek yang akan dijadikan bahan penelitian. Data yang digunakan dalam penelitian ini adalah data karyawan go-jek.

3.4 Persiapan Data (Data Preparation)

Dataset yang digunakan memiliki 1 variabel sebagai kelas yaitu status aplikasi go-jek —puas|| dan status aplikasi go-jek —tidak puas|| dan 4 variabel sebagai atribut. Variabel dan kategori yang digunakan pada penelitian ini adalah :

Tabel 3.1 Variabel dan kategori aplikasi go-jek

Variabel	Keterangan	Skala	Kategori
X1	Aplikasi Real	1 : Tidak Setuju 2 : Setuju	
3		: Sangat Setuju	

4 : Sangat Tidak Setuju

X2 Ketepatan Waktu Real 1 : Tidak Setuju 2 : Setuju

3 : Sangat Setuju

4 : Sangat Tidak Setuju

X3 Kenyamanan Berkendara Real 1 : Tidak Setuju 2 : Setuju

3 : Sangat Setuju

4 : Sangat Tidak Setuju

X4 Keramahan Real 1 : Tidak Setuju 2 : Setuju

3 : Sangat Setuju

4 : Sangat Tidak Setuju

#### 3.4.1 Pemodelan (Modelling)

Dalam penelitian ini akan dilakukan pengujian menggunakan klasifikasi dengan algoritma Naive bayes untuk mengetahui hasil perhitungan yang dianalisa dan untuk mengetahui apakah fungsi bekerja dengan baik atau tidak.. Setelah data dihitung secara manual, kemudian data diuji menggunakan tools RapidMiner untuk memastikan apakah hasil perhitungan manual dengan hasil yang diperoleh RapidMiner sama atau tidak.

#### 3.4.2 Evaluasi (Evaluation)

Melakukan pengecekan terhadap setiap nilai atribut dan model yang sudah dibangun. Kemudian melakukan evaluasi dengan cara mengamati dan menganalisa hasil dari algoritma yang digunakan untuk memastikan bahwa hasil pengujian benar dan sesuai hasil pembahasan, pengujian dilakukan untuk mengukur keakuratan hasil dari tiap model yang diusulkan.

Akurasi didefinisikan sebagai tingkat kedekatan antara nilai prediksi dengan nilai aktual. Pengukuran akurasi terhadap model dengan menggunakan confusion matrix yang menitik beratkan pada kelasnya. Confusion matrix merupakan table untuk mencatat hasil kerja klasifikasi. Berikut table Confusion Matrix untuk klasifikasi dua kelas :

Tabel 3.2 Perhitungan Akurasi, Presisi, Recall

	Classified as	
	+	-
+	True positives	False Negatives
-	False positives	True Negatives

Sumber : Han & Kamber dalam (Andriani, 2013)

Rumus confusion matrix :

$$\text{Akurasi} = \frac{TP+TN}{TP+FP+FN+TN}$$

$$\text{Presisi} = \frac{TP}{TP+FP}$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

Keterangan :

- Akurasi adalah proporsi jumlah prediksi yang benar.
- Precisi adalah proporsi kasus dengan hasil positif yang benar,

- TP (True positive) adalah jumlah record positif yang diklasifikasikan sebagai positif oleh classifier.
- TN (True negative) adalah jumlah record negatif yang diklasifikasikan sebagai negatif oleh classifier.
- FP (False positive) adalah jumlah record negatif yang diklasifikasikan sebagai positif oleh classifier.
- FN (False negative) adalah jumlah record positif yang diklasifikasikan sebagai negatif oleh classifier.

### 3.5 Metode Analisis Data

Naive Bayes merupakan metode probabilistik pengklasifikasian sederhana berdasarkan Teorema Bayes dimana pengklasifikasian dilakukan melalui training set sejumlah data secara efisien. Naive bayes mengasumsikan bahwa nilai dari sebuah input atribut pada kelas yang diberikan tidak tergantung dengan nilai atribut yang lain.

Metode analisis data menggunakan Naive Bayes Classifier (NBC) yang merupakan sebuah pengklasifikasi probabilitas sederhana yang mengaplikasikan Teorema Bayes dengan asumsi ketidaktergantungan (independen) yang tinggi.

Bentuk umum atau persamaan dari teorema Bayes adalah :

...(1)

Keterangan :

X : Data dengan class yang belum diketahui

H : Hipotesa data X merupakan suatu class spesifik

$P(H|X)$  : Probabilitas hipotesis H berdasar kondisi X (posteriori probability)  $P(H)$  : Probabilitas hipotesis H (prior probability)

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

Penjabaran lebih lanjut rumus Bayes tersebut dilakukan dengan menjabarkan  $(C|X_1, \dots, X_n)$  menggunakan aturan perkalian sebagai berikut.

$$P(C|x_1, \dots, x_n) = P(C) P(x_1, \dots, x_n | C)$$

$$= P(C)P(x_1 | C)P(x_2, \dots, x_n | C, x_1)$$

$$= P(C)P(x_1 | C)P(x_2 | C, x_1)P(x_3, \dots, x_n | C, x_1, x_2)$$

$$P(C)P(x_1 | C)P(x_2 | C, x_1)P(x_3 |$$

$$C, x_1, x_2)P(x_4, \dots, x_n | C, x_1, x_2, x_3)P(C)$$

$$= P(x_1 | C)P(x_2 | C, x_1)P(x_3 | C, x_1, x_2)$$

$$\dots P(x_n | C, x_1, x_2, x_3, \dots, x_{n-1}) \quad (2)$$

Dapat dilihat bahwa semakin banyak faktor-faktor yang semakin kompleks yang mempengaruhi nilai probabilitas maka semakin mustahil untuk menghitung nilai tersebut satu persatu. Akibatnya perhitungan semakin sulit untuk dilakukan, maka disinilah digunakan asumsi independensi yang sangat tinggi, bahwa masing-masing atribut dapat saling bebas.

Dengan asumsi tersebut, diperlukan persamaan :



$P(X_i|X_j) = \dots = P(X_i)$  untuk  $i \neq j$ , sehingga:

$$P(X_i | C, X_j) = P(X_i | C) \dots (3)$$

Dari persamaan (3) tersebut dapat di ambil kesimpulan bahwa asumsi independensi membuat syarat perhitungan menjadi lebih sederhana. Selanjutnya penjabaran  $P(C|X_1, \dots, X_n)$  dapat di sederhanakan menjadi persamaan (4) :

$$P(X_2|C)P(X_3|C) \dots$$

$$P(C|X_1, \dots, X_n) = P(X_1|C)$$

$$= \dots (4)$$

Keterangan :

= Perkalian ranting antar atribut

Dalam metode naive bayes diperlukan data latih dan data uji yang ingin diklasifikasikan, semakin banyak data latih yang yang dilibatkan, semakin baik hasil yang prediksi yang diberikan. Menghitung  $P(C_i)$  yang merupakan probabilitas prior untuk setiap sub kelas C yang akan dihasilkan menggunakan persamaan :

$$\dots (5)$$

Dimana :

$S_i$  : Jumlah data training dari kategori  $C_i$   $S$  : Jumlah total data training.

### 3.3 Instrumen Penelitian

Berdasarkan permasalahan yang telah diuraikan sebelumnya, maka bahan dan peralatan yang diperlukan untuk penelitian ini meliputi :

#### 3.6.1 Bahan

Dalam penelitian ini bahan yang dibutuhkan adalah data karyawan go-jek yang akan digunakan sebagai instrumentasi guna memperoleh data untuk menentukan status kesejahteraan rumah tangga.

#### 3.6.2 Peralatan

Peralatan dalam penelitian ini meliputi kebutuhan perangkat lunak dan kebutuhan perangkat keras. Dibawah ini merupakan peralatan atau tools yang dibutuhkan, diantaranya:

##### a) Perangkat Lunak (Software)

Perangkat Lunak, versi dan fungsi dapat dilihat pada tabel 3.4 dibawah ini : Tabel 3.3  
Perangkat Lunak (Software)

Microsoft Office Excel

2010 Digunakan untuk mengolah dataset

RapidMiner Studio

9.0 Digunakan untuk mengolah dataset dan untuk melihat hasil akurasi dari algoritma yang digunakan

b) Perangkat Keras (Hardware)

Selain perangkat lunak (software) dibutuhkan pula perangkat keras (hardware) sebagai pendukung penelitian data mining, yaitu laptop. Adapun spesifikasi laptop dijelaskan pada Tabel 3.4 dibawah ini

Tabel 3. 4 Perangkat Keras (Hardware)

Spesifikasi Hardware	Keterangan
Processor	Intel Core i3
RAM	2,00 GB
System Type	64-bit Operating System

4.1 Penentuan Kriteria

Dalam menganalisa kepuasan pelanggan pada perusahaan ojek online (GO- JEK) ada beberapa kriteria yang digunakan adalah sebagai berikut :

Tabel 4.1 Kriteria yang digunakan

Diketahui : Y X

Variabel Data Aplikasi Ojek Online (Go-Jek)

Y = Status Kepuasan Pelanggan

1 = Puas

2 = Tidak Puas X1 = Aplikasi

X2 = Ketetapan Waktu

X3 = Kenyamanan Berkendara X4 = Keramahan

#### 4.2 Perhitungan Naive Bayes

Dataset yang digunakan sebagai data training adalah sebanyak 36 data (lampiran 1). Sedangkan untuk data testing yang akan ditentukan hasil status rumah tangga berjumlah 13 data (lampiran 2).

##### 4.2.1 Perhitungan Probabilitas Prior ( $P(C_i)$ )

Dataset akan diproses dengan metode Naive Bayes dengan beberapa tahapan sehingga perlu menentukan besarnya nilai dari data, yaitu dari 36 data latih yang digunakan, diketahui

kelas C0 (Puas) sebanyak 23 data, dan kelas C1 (Tidak Puas) sebanyak 13 data. Perhitungan probabilitas prior dapat dilakukan dengan persamaan (5) yaitu :

Tabel 4. 1 Probabilitas Prior

Perhitungan Prior	Hasil
P (C0) 23/36	0,638
P (C1) 13/36	0.361

#### 4.2.2 Perhitungan Probabilitas Posterior X bersyarat C ( $P(X|C_i)$ )

Perhitungan probabilitas posterior dilakukan pada data latih sebanyak 36 data dengan menggunakan X sebagai vector penentuan status rumah tangga yaitu Xstatus kendaraan, Xnomer kendaraan, Xjumlah point, Xpendidikan, Xjenis kelamin. Sehingga  $P(X|C_i)$  dapat dijabarkan menjadi :

Tabel 4.3 Probabilitas Aplikasi

X1	Jumlah Kejadian			
"Dipilih"	Probabilitas			
Aplkasi	Puas	Tidak	Puas (C0)	Tidak (C1)
Sangat Setuju	8	2	0,347826087	0,153846154
Tidak Setuju	9	8	0,391304348	0,615384615
Setuju	6	3	0,260869565	0,230769231
Jumlah	23	13	1	1

Tabel 4.4 Probabilitas Ketepatan Waktu

X2	Jumlah Kejadian
—Dipilih	Probabilitas

Ketepatan Waktu	Puas	Tidak	Puas (C0)	Tidak Puas (C1)
Sangat Tidak Setuju	1	8	0,043478261	0,615384615
Setuju 19	1	0,826086957	0,076923077	
Tidak Setuju	3	4	0,130434783	0,307692308
Jumlah 23	13	1	1	

Tabel 4.5 Probabilitas Kenyamanan Dalam Berkendara

X3 Jumlah Kejadian

“Dipilih” Probabilitas

Kenyaman Dalam

Berkendara	Puas	Tidak	Puas (C0)	Tidak (C1)
Sangat Setuju	8	0	0,347826087	0
Setuju 14	5	0,608695652	0,384615385	
Tidak Setuju	1	8	0,043478261	0,615384615
Jumlah 23	13	1	1	

Tabel 4.6 Probabilitas Keramahan

X4 Jumlah Kejadian

“Dipilih” Probabilitas

Keramahan	Puas	Tidak	Puas (C0)	Tidak (C1)
Sangat Setuju	6	2	0,260869565	0,153846154
Setuju 13	4	0,565217391	0,307692308	
Tidak Setuju	4	7	0,173913043	0,538461538
Jumlah 23	13	1	1	

#### 4.3 Perhitungan Manual

Berikut ini perhitungan manual dengan menggunakan data uji yang dapat dilihat pada data training (lampiran 4.1), dengan menggunakan metode Naive Bayes.

Tabel 4.7 Data Uji

Aplikasi

Ketepatan Waktu      Kenyamanan dalam berkendara

keramahan

Prediksi

Sangat Setuju   Sangat Tidak Setuju   Sangat Setuju   Sangat Setuju   ?

##### 4.3.1 Pendefinisian Variabel

Berdasarkan tabel 4.7 dapat didefinisikan data uji X adalah sebagai berikut :  $X = \{X_{Aplikasi}=Sangat Setuju, X_{Ketepatan Waktu}=Sangat Tida Setuju, X_{Keramahan}=Sangat Setuju\}$

##### 4.3.2 Pendefinisian Probabilitas Prior $P(C_i)$

Hasil pendefinisian Probabilitas prior berdasarkan persamaan (5) menghasilkan nilai untuk kelas puas ( $C_0$ ) sebesar 0.638 dan untuk Tidak Puas ( $C_1$ ) sebesar 0.361.

##### 4.3.3 Perhitungan Probabilitas Data Uji

Berdasarkan data uji pada tabel 4.11 dapat dilakukan klasifikasi kedalam kelas puas ( $C_0$ ) dengan ketentuan nilai masing-masing kriteria yaitu : 0.348, 0.043, 0.348, 0.261.

Kemudian nilai dari masing-masing kriteria tersebut dikalikan  $P(X|C0) = 0.348 * 0.043 * 0.348 * 0.261 = 0.00136$

Untuk menghitung klasifikasi kedalam kelas tidak puas (C1) dengan ketentuan nilai masing-masing kriteria yaitu : 0.154, 0.615, 0, 0.154.

Kemudian nilai dari masing-masing kriteria tersebut dikalikan  $P(X|C1) = 0.154 * 0.615 * 0 * 0.154 = 0$

#### 4.3.4 Pemaksimalan $P(X|Ci) P(Ci)$

Perhitungan pemaksimal untuk klasifikasi kelas puas (C0) adalah dengan cara mengalikan  $P(X|C0)$  dengan  $P(C0)$  :

$$P(C0|X) = P(X|C0) * P(C0)$$

$$= 3,926 * 0,638 = 2,5048$$

Kemudian untuk kelas tidak puas (C1) adalah dengan cara mengalikan  $P(X|C1)$  dengan  $P(C1)$  :

$$P(C1|X) = P(X|C1) * P(C1)$$

$$= 0 * 0.361 = 0$$

Dari perhitungan di atas dapat dihasilkan nilai  $P(C0|X) = 2,5048$  sedangkan nilai  $P(C1|X) = 0$ . Berdasarkan nilai tersebut dapat di ambil kesimpulan bahwa  $P(C0|X)$

$> P(C1|X)$ . maka data uji tersebut diklasifikasikan kedalam kelas Miskin dalam kesejahteraan rumah tangga.

#### 4.4 Implementasi Klasifikasi Naive Bayes pada RapidMiner

Uji coba dilakukan untuk mengetahui apakah perhitungan yang telah dilakukan diatas sesuai untuk menentukan status kesejahteraan rumah tangga dengan metode Naive Bayes. Uji



coba dilakukan dengan menentukan 19 data testing yang telah dipilih (lampiran 1). Data testing tersebut akan dicari nilai prediksinya menggunakan RapidMiner 9.0.

#### 4.4.1 Proses Select Attributes

Melakukan select attributes yaitu untuk mengetahui hasil prediski dari RapidMiner, apakah hasil perhitungan manual sama atau tidak dengan hasil RapidMiner. Dapat dilihat pada gambar 4.1 Proses Select Attribute dan Hasil Prediksi RapidMiner dapat dilihat pada gambar 4.2.

Gambar 4.1 Proses Select Attribute

Gambar 4.2 Hasil Prediksi RapidMiner

#### 4.4.2 Akurasi Prediksi

Proses klasifikasi dengan RapidMiner menggunakan metode algoritma Naive Bayes pada data rumah tangga ini untuk membandingkan data testing dengan data training yang sudah diketahui rule-rulanya sebelumnya. Berikut langkahnya:

Pada tampilan process masukan operator Read Excel masukan masing- masing data training dan data testing, selanjutnya masukan operator Naive Bayes, Apply Model, dan Performance lalu sambungkan kabel seperti gambar dibawah ini :

#### Gambar 4.3 Proses Accuracy Prediksi

#### Gambar 4.4 Hasil Accuracy Data Testing

Hasil pengukuran data accuracy yang diperoleh dari data training mencapai 84.21%. Jumlah prediksi puas yang diklasifikasikan sebagai puas oleh classifier yaitu 13 data, dan jumlah prediksi tidak puas yang diklasifikasikan sebagai puas oleh classifier yaitu 0. Dengan pencapaian class precision 100.00%.

Sedangkan jumlah prediksi tidak puas yang diklasifikasikan sebagai puas oleh classifier yaitu 3 data, dan jumlah prediksi tidak puas yang diklasifikasikan sebagai tidak puas oleh classifier yaitu 3 data. Dengan pencapaian class precision 50.00%.

Untuk class recall dengan true tidak puas mencapai 81,25% sedangkan untuk

class recall dengan true puas mencapai 100.00%.

#### 4.4.3 Kurva ROC/AUC (Area Under Curve)

Menurut Vercellis dalam (Andriani, 2013) Kurva ROC menunjukkan akurasi dan membandingkan klasifikasi secara visual dan ROC mengekspresikan confusion matrix. ROC adalah grafik dua dimensi dengan false positives sebagai garis horizontal dan true positive sebagai garis vertikal. Tingkat keakurasian AUC dapat diklasifikasikan menjadi lima kelompok yaitu :

- a)  $0.90 - 1.00 = \text{Excellent Classification}$
- b)  $0.80 - 0.90 = \text{Good Classification}$
- c)  $0.70 - 0.80 = \text{Fair Classification}$
- d)  $0.60 - 0.70 = \text{Poor Classification}$
- e)  $0.50 - 0.60 = \text{Failure}$

Nilai AUC yang didapatkan dari pengujian berdasarkan kurva ROC menggunakan metode Naïve bayes sebesar 0.998, sehingga dari hasil tersebut berdasarkan kriteria diatas menunjukan klasifikasi yang dihasilkan termasuk kedalam kelompok excellent classification atau klasifikasi yang sangat baik. Hasil kurva ROC dapat dilihat pada gambar 4.5.

Gambar 4.5 Kurva ROC

Untuk mengetahui performance vector yang diperoleh, maka akan dijelaskan di bawah ini :

PerformanceVector:

accuracy: 84.21% ConfusionMatrix:

True: Puas      Tidak Puas Puas: 13      0

Tidak Puas:    3      3

precision: 50.00% (positive class: puas) ConfusionMatrix:

True: Puas      Tidak Puas Puas: 13      0

Tidak Puas:    3      3

recall: 100.00% (positive class: Tidak Puas) ConfusionMatrix:

True: Puas      Tidak Puas Puas: 13      0

Tidak Puas:    3      3

AUC (optimistic): 0.998 (positive class: Puas) AUC: 0.998 (positive class: Puas)

AUC (pessimistic): 0.998 (positive class: Puas )