

**KOMPARASI ALGORITMA MACHINE LEARNING DALAM
MEMPREDIKSI KEPUASAN PENGGUNAAN JASA
TRANSPORTASI BUS ROSALIA INDAH**



SKRIPSI

Diajukan untuk memenuhi salah satu syarat kelulusan Program Sarjana

DIMAS MIFTAKHUL FAKRI

NIM : 15200304

Program Studi Informatika Kampus Utama

Fakultas Teknik dan Informatika

Universitas Bina Sarana Informatika

2024

PERSEMBAHAN

Tidak masalah seberapa lambat kamu berjalan, selama kamu tidak pernah berhenti.

(Confusius)

Dengan mengucapkan puji syukur kepada Allah S.W.T, skripsi ini kupersembahkan untuk:

1. Kedua orang tua, yaitu Bapak Sarto dan Ibu Endang Murniasih tercinta yang telah membesarkanku dan selalu membimbing, mendukung, memotivasi, memberi arahan apa yang terbaik bagiku serta selalu mendoakan aku untuk meraih kesuksesanku.
2. Diriku sendiri sebagai bentuk apresiasi dan penghargaan atas segala usaha, kerja keras, dan ketekunan dalam menyelesaikan skripsi ini.
3. Teman-teman yang selalu mendukungku untuk berkembang dan jauh lebih baik kedepannya.

*Tanpa mereka,
aku dan karya ini tak akan pernah ada*

UNIVERSITAS

SURAT PERNYATAAN KEASLIAN SKRIPSI

Yang bertanda tangan di bawah ini:

Nama : Dimas Miftakhul Fakri
NIM : 15200304
Jenjang : Sarjana (S1)
Program Studi : Informatika
Fakultas : Teknik dan Informatika
Perguruan Tinggi : Universitas Bina Sarana Informatika

Dengan ini menyatakan bahwa Skripsi yang telah saya buat dengan judul: **“Komparasi Algoritma Machine Learning Dalam Memprediksi Kepuasan Penggunaan Jasa Transportasi Bus Rosalia Indah”** adalah asli (orisinal) atau tidak plagiat (menjiplak) dan belum pernah diterbitkan/dipublikasikan dimanapun dan dalam bentuk apapun.

Demikianlah surat pernyataan ini saya buat dengan sebenar-benarnya tanpa ada paksaan dari pihak manapun juga. Apabila di kemudian hari ternyata saya memberikan keterangan palsu dan atau ada pihak lain yang mengklaim bahwa (Skripsi) yang telah saya buat adalah hasil karya milik seseorang atau badan tertentu, saya bersedia diproses baik secara pidana maupun perdata dan kelulusan saya dari **Universitas Bina Sarana Informatika** dicabut/dibatalkan.

Dibuat di : Jakarta
Pada tanggal : 28 Juni 2024
Yang menyatakan,



A handwritten signature in black ink that reads 'Dimas'.

Dimas Miftakhul Fakri

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Yang bertandatangan di bawah ini, saya:

Nama : Dimas Miftakhul Fakri
NIM : 15200304
Jenjang : Sarjana (S1)
Program Studi : Informatika
Fakultas : Teknik dan Informatika
Perguruan Tinggi : Universitas Bina Sarana Informatika

Dengan ini menyatakan bahwa seluruh data, informasi, interpretasi serta pernyataan yang terdapat dalam karya ilmiah Penulis dengan judul “**Komparasi Algoritma Machine Learning Dalam Memprediksi Kepuasan Penggunaan Jasa Transportasi Bus Rosalia Indah**” ini, kecuali yang disebutkan sumbernya adalah hasil pengamatan, penelitian, pengelolaan, serta pemikiran saya.

Penulis menyetujui untuk memberikan ijin kepada pihak **Universitas Bina Sarana Informatika** untuk mendokumentasikan karya ilmiah saya tersebut secara internal dan terbatas, serta tidak untuk mengunggah karya ilmiah Penulis pada repository Universitas Bina Sarana Informatika.

Penulis bersedia untuk bertanggung jawab secara pribadi, tanpa melibatkan pihak **Universitas Bina Sarana Informatika**, atas materi/isi karya ilmiah tersebut, termasuk bertanggung jawab atas dampak atau kerugian yang timbul dalam bentuk akibat tindakan yang berkaitan dengan data, informasi, interpretasi serta pernyataan yang terdapat pada karya ilmiah saya ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta
Pada tanggal : 28 Juni 2024
Yang menyatakan,



Dimas Miftakhul Fakri

PERSETUJUAN DAN PENGESAHAN SKRIPSI

Skripsi ini diajukan oleh:

Nama : Dimas Miftakhul Fakri
NIM : 15200304
Jenjang : Sarjana (S1)
Program Studi : Informatika
Fakultas : Teknik dan Informatika
Perguruan Tinggi : Universitas Bina Sarana Informatika
Judul Skripsi : Komparasi Algoritma Machine Learning Dalam Memprediksi Kepuasan Penggunaan Jasa Transportasi Bus Rosalia Indah

Telah dipertahankan pada periode 2024-1 dihadapan penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh Sarjana Komputer (S.Kom) pada Program Sarjana (S1) Program Studi Informatika di Universitas Bina Sarana Informatika.

Jakarta, 22 Juli 2024

PEMBIMBING SKRIPSI

Pembimbing I : Waeisul Bismi, M.Kom.

DEWAN PENGUJI

Penguji I : Taufik Rahman, M.Kom.

Penguji II : Rian Septian Anwar, M.Kom.

PEDOMAN PENGGUNAAN HAK CIPTA

Skripsi sarjana yang berjudul “**Komparasi Algoritma Machine Learning Dalam Memprediksi Kepuasan Penggunaan Jasa Transportasi Bus Rosalia Indah**” adalah hasil karya tulis asli Dimas Miftakhul Fakri dan bukan hasil terbitan sehingga peredaran karya tulis hanya berlaku di lingkungan akademik saja, serta memiliki hak cipta. Oleh karena itu, dilarang keras untuk menggandakan baik sebagian maupun seluruhnya karya tulis ini, tanpa seizin penulis.

Referensi kepustakaan diperkenankan untuk dicatat tetapi pengutipan atau peringkasan isi tulisan hanya dapat dilakukan dengan seizin penulis dan disertai ketentuan pengutipan secara ilmiah dengan menyebutkan sumbernya.

Untuk keperluan perizinan pada pemilik dapat menghubungi informasi yang tertera di bawah ini:

Nama : Dimas Miftakhul Fakri
Alamat : Taman Tridaya Indah 1 Blok A.02 No.7, RT 002/009,
Tambun Selatan, Bekasi
No. Telp : 082322602008
E-mail : dimasmiftakhul25@gmail.com

UNIVERSITAS

KATA PENGANTAR

Alhamdulillah, dengan mengucapkan puji syukur kehadiran Allah SWT, yang telah melimpahkan rahmat dan karunia-Nya, sehingga pada akhirnya penulis dapat menyelesaikan tugas ini dengan baik. Skripsi ini penulis sajikan dalam bentuk buku yang sederhana. Adapun judul Skripsi yang penulis ambil sebagai berikut, **"Komparasi Algoritma Machine Learning Dalam Memprediksi Kepuasan Penggunaan Jasa Transportasi Bus Rosalia Indah"**.

Tujuan penulisan Skripsi ini dibuat sebagai salah satu syarat kelulusan Sarjana Universitas Bina Sarana Informatika. Sebagai bahan penulisan diambil berdasarkan hasil penelitian (eksperimen), observasi dan beberapa sumber literatur yang mendukung penulisan ini. Penulis menyadari bahwa tanpa bimbingan dan dorongan dari semua pihak, maka penulisan Skripsi ini tidak akan berjalan lancar. Oleh karena itu pada kesempatan ini, izinkanlah penulis menyampaikan ucapan terima kasih kepada:

1. Rektor Universitas Bina Sarana Informatika.
2. Dekan Fakultas Teknik dan Informatika Universitas Bina Sarana Informatika.
3. Ketua Program Studi Ilmu Komputer di Universitas Bina Sarana Informatika
4. Bapak Waeisul Bismi, M.Kom selaku Dosen Pembimbing Skripsi.
5. Staff / karyawan / dosen di lingkungan Universitas Bina Sarana Informatika.
6. Orang tua tercinta yang telah memberikan dukungan moral maupun spiritual.
7. Rekan-rekan mahasiswa kelas 15.8A.05
8. Teman-teman Bismania Community yang sudah membantu dalam penelitian ini.

Serta semua pihak yang terlalu banyak untuk disebut satu persatu sehingga terwujudnya penulisan ini. Penulis menyadari bahwa penulisan Skripsi ini masih jauh sekali dari sempurna, untuk itu penulis mohon kritik dan saran yang bersifat membangun demi kesempurnaan penulisan di masa yang akan datang.

Akhir kata semoga Skripsi ini dapat berguna bagi penulis khususnya dan bagi para pembaca yang berminat pada umumnya.

Jakarta, 28 Juni 2024

Penulis



Dimas Miftakhul Fakri



ABSTRAK

Dimas Miftakhul Fakri (15200304), Komparasi Algoritma *Machine Learning* Dalam Memprediksi Kepuasan Penggunaan Jasa Transportasi Bus Rosalia Indah

Beragamnya kegiatan seringkali mengharuskan manusia untuk mudah berpindah lokasi dari sebuah lokasi menuju ke lokasi yang lain. Untuk mencukupi keperluan tersebut manusia membutuhkan moda transportasi. Bus, menjadi sarana transportasi massal yang memberikan dampak baik serta berdampak positif pada lingkungan dengan mengurangi polusi udara serta kemacetan lalu lintas dan mudah dijangkau. Salah satu perusahaan bus yang terkenal dengan pelayanannya yaitu Rosalia Indah. Dalam menghadapi tantangan dalam memastikan kepuasan pengguna di tengah persaingan bisnis yang ketat. Memahami faktor-faktor yang memengaruhi kepuasan pengguna menjadi kunci untuk mempertahankan dan meningkatkan kualitas layanan. Penggunaan algoritma *machine learning* menawarkan solusi dalam menganalisis dan memprediksi kepuasan pengguna secara lebih akurat dan efisien. Dengan membandingkan dua algoritma *machine learning* yang populer, diharapkan dapat ditemukan algoritma yang paling efektif untuk meningkatkan kualitas layanan transportasi bus Rosalia Indah. Dari hasil pengujian yang dilakukan untuk mengetahui kepuasan pengguna Rosalia Indah dengan memanfaatkan metode *naïve bayes* dan *k-nearest neighbor*. Hasil dari *Naïve bayes* memiliki nilai rata-rata akurasi sejumlah 97.39% serta nilai AUC berjumlah 1.000 Sedangkan *k-nearest neighbors* memiliki tingkat rata-rata akurasi cukup tinggi yaitu 98.33% dengan nilai AUC sebesar 1.000. dalam hal ini *k-nearest neighbors* mempunyai tingkat akurasi lebih tinggi dibanding *naïve bayes* dengan selisih perbandingan yang tidak signifikan yaitu 0.94 %.

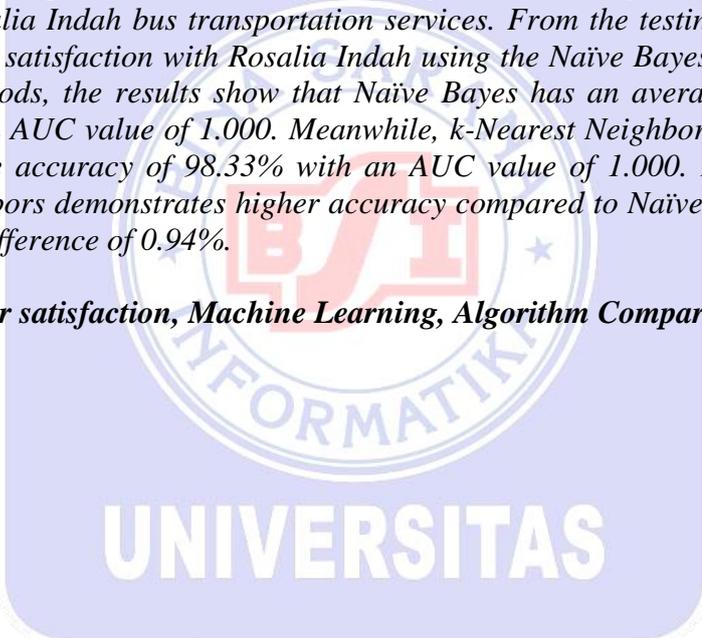
Kata kunci: Kepuasan Pengguna, *Machine Learning*, Komparasi Algoritma

ABSTRACT

Dimas Miftakhul Fakri (15200304), Comparison of Machine Learning Algorithms in Predicting User Satisfaction with Rosalia Indah Bus Transportation Services.

The variety of activities often requires people to easily move from one location to another. To fulfill these needs, humans need a mode of transportation. The bus, as a means of mass transportation, has a positive impact on the environment by reducing air pollution and traffic congestion and is easily accessible. One bus company known for its service is Rosalia Indah. In facing challenges in ensuring user satisfaction amidst intense business competition. Understanding the factors that influence user satisfaction is key to maintaining and improving service quality. The use of machine learning algorithms offers solutions in analyzing and predicting user satisfaction more accurately and efficiently. By comparing several popular machine learning algorithms, it is hoped that the most effective algorithm can be found to improve the quality of Rosalia Indah bus transportation services. From the testing conducted to determine user satisfaction with Rosalia Indah using the Naïve Bayes and k-Nearest Neighbor methods, the results show that Naïve Bayes has an average accuracy of 97.39% and an AUC value of 1.000. Meanwhile, k-Nearest Neighbors has a slightly higher average accuracy of 98.33% with an AUC value of 1.000. In this case, k-Nearest Neighbors demonstrates higher accuracy compared to Naïve Bayes, with an insignificant difference of 0.94%.

Keywords: User satisfaction, Machine Learning, Algorithm Comparison



UNIVERSITAS

DAFTAR ISI

LEMBAR JUDUL	i
PERSEMBAHAN.....	ii
SURAT PERNYATAAN KEASLIAN SKRIPSI	iii
SURAT PERNYATAAN PERSETUJUAN PUBLIKASI	iv
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS	iv
PERSETUJUAN DAN PENGESAHAN SKRIPSI	v
PEDOMAN PENGGUNAAN HAK CIPTA	vi
KATA PENGANTAR.....	vii
ABSTRAK	ix
ABSTRACT	x
DAFTAR ISI.....	xi
DAFTAR SIMBOL	xiii
DAFTAR GAMBAR.....	xv
DAFTAR TABEL	xvi
DAFTAR LAMPIRAN	xvii
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah	4
1.3 Tujuan dan Manfaat	4
1.4 Hipotesis.....	6
1.5 Batasan Masalah.....	6
BAB II LANDASAN TEORI	8
2.1 Tinjauan Pustaka	8
2.1.1 Transportasi	8
2.1.2 Kepuasan Pengguna	9
2.1.2.1 Kualitas Pelayanan	9
2.1.2.2 Harga	10
2.1.3 <i>Machine Learning</i>	10
2.1.4 <i>Supervised Learning</i>	11
2.1.5 <i>Data Mining</i>	11
2.1.6 <i>Pre-processing data</i>	11
2.1.7 <i>Splitting data</i>	12
2.1.8 <i>RapidMiner</i>	13
2.1.9 <i>Confusion Matrix</i>	13
2.1.10 <i>Kurva Roc</i>	13
2.2 Penelitian Terkait	14
BAB III METODOLOGI PENELITIAN	21
3.1 Proses dan Langkah Penelitian.....	21
3.2 Metode Pengolahan dan Analisis Data	22

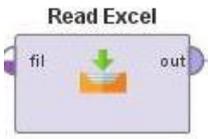
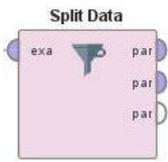
3.2.1	Analisis Masalah.....	22
3.2.2	Pengumpulan Data.....	22
3.3	Uji Validitas dan Reliabilitas	25
3.3.1	Uji Validitas.....	25
3.3.2	Uji Reliabilitas	25
3.4	Pengolahan Data.....	26
3.5	Pengujian Algoritma	26
3.5.1	Algoritma Naïve Bayes.....	27
3.5.2	<i>K-Nearest Neighbors</i>	28
3.6	Evaluasi Kinerja	28
BAB IV	HASIL DAN PEMBAHASAN	31
4.1	Hasil Penelitian	31
4.1.1	Hasil Pengumpulan data	31
4.2	Hasil <i>Pre-processing Data</i>	34
4.2.1	Uji Validitas dan Reliabilitas.....	34
4.2.2	<i>Data Cleaning</i>	36
4.2.3	Pelabelan data	38
4.2.4	<i>Splitting Data</i>	40
4.3	Hasil Pengujian Algoritma.....	41
4.4	Hasil Evaluasi.....	63
BAB V	KESIMPULAN	65
5.1	Kesimpulan	65
5.2	Saran.....	66
DAFTAR PUSTAKA		67
DAFTAR RIWAYAT HIDUP		71
LEMBAR KONSULTASI SKRIPSI.....		72
SURAT PERNYATAAN KEBENARAN/KEABSAHAN DATA HASIL RISET		
UNTUK KARYA ILMIAH		73
LAMPIRAN-LAMPIRAN		74

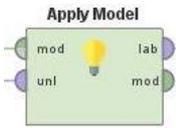
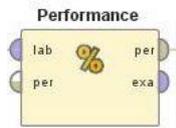
DAFTAR SIMBOL

a. Simbol Flowchart

SIMBOL	KETERANGAN
	<p><i>RECTANGLE</i></p> <p>Digunakan untuk pemodelan atau langkah suatu proses dalam alur kerja.</p>
	<p><i>LINE ARROW</i></p> <p>Digunakan untuk menunjukkan alur atau proses.</p>

b. Simbol RapidMiner

SIMBOL	KETERANGAN
	<p><i>READ EXCEL</i></p> <p>Digunakan untuk memasukan dataset kedalam <i>rapidminer</i> dalam format file Excel (.xlsx atau .xls) yang kemudian akan diolah.</p>
	<p><i>SPLIT DATA</i></p> <p>Digunakan untuk mengubah dataset menjadi <i>data training</i> dan <i>data testing</i> dengan rasio tertentu.</p>
	<p><i>NAÏVE BAYES</i></p> <p>Digunakan untuk melakukan pengkategorian data berdasarkan kategori tertentu dengan metode <i>naïve bayes</i>.</p>

 <p>The icon for k-NN shows a green box with a lightbulb in the center. To the left of the lightbulb is the text 'tra' and to the right is 'mod'. Below the lightbulb is the text 'exa'. A green checkmark is in the bottom-left corner.</p>	<p><i>K-NEAREST NEIGHBORS</i></p> <p>Digunakan untuk menentukan kelas atau kategori dari data input berdasarkan mayoritas kelas dari k tetangga terdekatnya menggunakan metode <i>k-nn</i>.</p>
 <p>The icon for Apply Model shows a green box with a lightbulb in the center. To the left of the lightbulb is the text 'mod' and to the right is 'lab'. Below the lightbulb is the text 'unl' and 'mod'.</p>	<p><i>APPLY MODEL</i></p> <p>Digunakan untuk membuat prediksi atau mengambil keputusan berdasarkan data yang digunakan.</p>
 <p>The icon for Performance shows a yellow box with a percentage sign in the center. To the left of the percentage sign is the text 'lab' and to the right is 'per'. Below the percentage sign is the text 'per' and 'exa'.</p>	<p><i>PERFORMANCE</i></p> <p>Digunakan mengukur akurasi hasil pemodelan yang dibuat.</p>



DAFTAR GAMBAR

Gambar III. 1 Kerangka Penelitian	21
Gambar III. 2 Rumus Slovin	24
Gambar III. 3 Interface RapidMiner	27
Gambar III. 4 Rumus Naïve Bayes	27
Gambar III. 5 Rumus K-Neares Neighbors.....	28
Gambar III. 6 Kurva ROC.....	30
Gambar IV. 1 Form Kuesioner.....	31
Gambar IV. 2 Penghitungan Rentang Skor Rata-rata	38
Gambar IV. 3 Rasio Split Data	40
Gambar IV. 4 Model Pengujian Naïve Bayes	42
Gambar IV. 5 Kurva ROC Pengujian Pertama Naïve Bayes	44
Gambar IV. 6 Kurva ROC Pengujian Kedua Naive Bayes.....	46
Gambar IV. 7 Kurva ROC Pengujian Ketiga Naïve Bayes	48
Gambar IV. 8 Kurva ROC Pengujian Keempat Naïve Bayes.....	50
Gambar IV. 9 Model Pengujian K-Nearest Neighbors	52
Gambar IV. 10 Kurva ROC Pengujian Pertama Metode K-nn.....	55
Gambar IV. 11 Kurva ROC Pengujian Kedua Metode K-nn.....	57
Gambar IV. 12 Kurva ROC Pengujian Ketiga Metode K-nn	59
Gambar IV. 13 Kurva ROC Pengujian Keempat Metode K-nn.....	61

DAFTAR TABEL

Tabel II. 1 Penelitian Terkait	17
Tabel III. 1 Skala Likert	23
Tabel III. 2 Confusion Matrix	29
Tabel IV. 1 Dataset Kepuasan Pengguna Jasa Bus Rosalia Indah	32
Tabel IV. 2 Uji Validitas	35
Tabel IV. 3 Uji Reliabilitas	36
Tabel IV. 4 Cleaning Data	37
Tabel IV. 5 Kategori Umur	37
Tabel IV. 6 Frekuensi menggunakan bus selama sebulan	38
Tabel IV. 7 Kategori Jenis Kelamin	38
Tabel IV. 8 Interval Nilai	39
Tabel IV. 9 Hasil Cleaning Data	39
Tabel IV. 10 Kepuasan Pengguna	40
Tabel IV. 11 Pengujian Confusion Matrix Naïve Bayes Rasio 90% dan 10%	43
Tabel IV. 12 Pengujian Confusion Matrix Naïve Bayes Rasio 80% dan 20%	45
Tabel IV. 13 Pengujian Confusion Matrix Naïve Bayes Rasio 70% dan 30%	47
Tabel IV. 14 Pengujian Confusion Matrix Naïve Bayes Rasio 60% dan 40%	49
Tabel IV. 15 Total Keseluruhan Pengujian Naïve Bayes	51
Tabel IV. 16 Pengujian Confusion Matrix K-nn Rasio 90% dan 10%	54
Tabel IV. 17 Pengujian Confusion Matrix K-nn Rasio 80% dan 20%	56
Tabel IV. 18 Pengujian Confusion Matrix K-nn Rasio 70% dan 30%	58
Tabel IV. 19 Pengujian Confusion Matrix K-nn Rasio 60% dan 40%	60
Tabel IV. 20 Total Keseluruhan Pengujian K-Nearest Neighbors	62
Tabel IV. 21 Keseluruhan Pengujian Bayes dan K-nn	63
Tabel IV. 22 Hasil Komparasi Naive Bayes dan K-nn	64

DAFTAR LAMPIRAN

Lampiran 1. Cek Plagiat.....	74
Lampiran 2. Kuesioner.....	82
Lampiran 3. Uji Validitas dan Reliabilitas.....	87



BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Beragamnya kegiatan mengharuskan manusia untuk mudah berpindah lokasi dari sebuah lokasi menuju ke lokasi yang lain. Untuk mencukupi akan keperluan perpindahan tersebut manusia membutuhkan moda transportasi. Transportasi merupakan salah satu aspek kegiatan masyarakat yang selalu dibutuhkan oleh masyarakat yang rutin beraktivitas untuk memenuhi kebutuhannya sehari-hari (Amir & Rahman, 2020). Salah satu jenis transportasi yang kerap digunakan adalah bus.

Transportasi bus adalah sarana transportasi massal yang digunakan untuk mengangkut banyak penumpang dari satu kota ke kota lainnya (Saphari, 2023). Selain murah dan nyaman, bus juga sangat diminati oleh masyarakat umum dan mudah diakses oleh seluruh lapisan masyarakat (Wiyanto, 2020).

Bus memiliki banyak jenis, mulai dari mini bus, medium bus, hingga bus besar. Transportasi bus sering dianggap sebagai bagian penting dari sistem transportasi publik di berbagai kota dan negara, ini memberikan aksesibilitas yang baik bagi banyak orang untuk berpindah dari satu daerah ke sebuah tempat yang lain, dan menjadi pilihan transportasi yang terjangkau bagi banyak orang. Dengan kapasitas yang dapat menampung lebih dari satu penumpang, bus menjadi pilihan favorit untuk masyarakat yang hendak bepergian dengan efisien dan terjangkau. Selain itu, Bus juga memiliki juga memberikan dampak positif bagi lingkungan dalam mengurangi polusi udara dengan mengurangi jumlah penggunaan mobil pribadi dan kemacetan lalu lintas.

Beberapa perusahaan penyedia jasa transportasi bus berlomba-lomba untuk memberikan pelayanan terbaiknya, salah satu perusahaan transportasi bus yang cukup terkenal di Indonesia adalah perusahaan bus asal Karanganyar, Jawa Tengah, yaitu Rosalia Indah. Perusahaan Otobus (PO) Rosalia Indah beroperasi dalam bidang transportasi khususnya layanan angkutan darat. Perusahaan Otobus (PO) Rosalia Indah sudah menjalankan kegiatan usahanya sejak tahun 1983. Perusahaan Otobus (PO) Rosalia Indah masuk kategori perusahaan swasta yang berkembang cukup pesat dan saat ini telah melayani perjalanan di pulau Jawa hingga pulau Sumatera.

Masalah utama sebagai salah satu pelayanan jasa transportasi yaitu apakah pelayanan yang diberikan dapat memuaskan para penggunanya. Dalam ketatnya kondisi persaingan bisnis transportasi saat ini, memahami faktor-faktor yang memengaruhi kepuasan pengguna merupakan hal faktor kunci bagi perusahaan transportasi seperti Rosalia Indah. Sebagai perusahaan yang berfokus pada bidang transportasi bus, Perusahaan Otobus (PO) Rosalia Indah menyadari banyak aspek yang dapat mempengaruhi tingkat kepuasan pengguna. Dengan memperhatikan dan meningkatkan standar produk yang ada, harga yang bersaing, serta memberikan pelayanan yang prima dan berkualitas, Perusahaan Otobus (PO) Rosalia Indah dapat semakin memenuhi harapan pelanggan dan memperkuat hubungan dalam jangka waktu yang lama dengan pelanggan yang ada (Sariatina & Ekawati, 2023).

Perusahaan Otobus (PO) Rosalia Indah, sebagai salah satu penyedia jasa transportasi bus terkemuka di Indonesia terus berupaya meningkatkan kualitas layanannya. Namun, dalam mengevaluasi kepuasan pengguna seringkali sulit untuk memperoleh gambaran yang akurat dan real-time karena metode tradisional seperti survei atau penilaian manual dapat menghabiskan waktu dan biaya yang cukup besar,

sehingga diperlukan metode yang lebih efisien dan dapat memberikan prediksi yang akurat terkait dengan kepuasan pengguna.

Dengan kemajuan teknologi dan ketersediaan data yang besar, penggunaan algoritma *machine learning* telah menjadi pilihan menarik dalam menganalisa dan memprediksi kepuasan pengguna terhadap layanan transportasi. Dengan melakukan komparasi antara dua algoritma *machine learning* yang populer dalam memprediksi kepuasan pengguna jasa transportasi bus Rosalia Indah, diharapkan dapat membantu dalam menentukan algoritma yang paling cocok dan efektif, sehingga memungkinkan penyedia jasa transportasi bus untuk meningkatkan kualitas layanan mereka.

Banyak penelitian yang telah dilakukan terkait kepuasan pengguna, salah satunya adalah penelitian oleh (Diansyah, 2022) tentang “Klasifikasi Tingkat Kepuasan Pengguna dengan Menggunakan Metode *K-Nearest Neighbour (KNN)*” Penelitian ini melibatkan pengujian dengan metode *k-neares neighbour* dengan hasil tingkat akurasi 94.12%.

Penelitian berikutnya terkait kepuasan pengguna juga dilakukan oleh (Iwandini et al., 2023) tentang “Analisa Sentimen Pengguna Transportasi Jakarta Terhadap Transjakarta Menggunakan Metode *Naives Bayes* dan *K-Nearest Neighbor*” pada penelitian tersebut digunakan dua metode untuk pengujian, yaitu *naïve bayes* dan *k-neares neighbors* untuk pengolahan datanya. Hasil dari pengujian tersebut menunjukkan bahwa *naïve bayes* memprediksi hasil akurasi 61,1%, dan *K-Nearest Neighbor* memprediksi hasil akurasi 75,7%.

Berdasarkan uraian tersebut, maka penulis memiliki ketertarikan untuk melakukan kegiatan penelitian yang dirangkum dan disusun dalam skripsi ini dan

diberi judul “**Komparasi Algoritma Machine Learning Dalam Memprediksi Kepuasan Penggunaan Jasa Transportasi Bus Rosalia Indah**”.

1.2 Rumusan Masalah

Melihat dari latar belakang masalah yang sudah dipaparkan, maka dapat dirumuskan permasalahan, diantaranya:

1. Bagaimana performa algoritma *machine learning* dalam memprediksi kepuasan pengguna jasa transportasi bus Rosalia Indah?
2. Algoritma *machine learning* mana yang memiliki akurasi tertinggi dalam memprediksi tingkat kepuasan pengguna?
3. Apa saja faktor utama yang mempengaruhi tingkat kepuasan pengguna layanan transportasi bus Rosalia Indah?

1.3 Tujuan dan Manfaat

Tujuan :

Tujuan dilakukannya penelitian ini adalah:

1. Menilai performa berbagai algoritma *machine learning* dalam memprediksi kepuasan pengguna jasa transportasi bus Rosalia Indah.
2. Membandingkan berbagai algoritma *machine learning* untuk menentukan algoritma yang memiliki tingkat akurasi tertinggi dalam memprediksi tingkat kepuasan pengguna.
3. Untuk mengidentifikasi apa saja faktor utama yang paling mempengaruhi terhadap tingkat kepuasan pengguna layanan transportasi bus Rosalia Indah berdasarkan hasil prediksi dari berbagai algoritma.

Manfaat :

Penelitian ini dilakukan dengan tujuan bisa menawarkan manfaat kepada berbagai kelompok yang membutuhkan, diantaranya:

1. Bagi Penulis

Menjadi tambahan ilmu bagi penulis dalam memperluas pengetahuan tentang penggunaan teknologi dalam industri transportasi tentang bagaimana algoritma *machine learning* dapat digunakan untuk memprediksi kepuasan pengguna jasa transportasi.

2. Bagi Pembaca

Penelitian ini dapat menjadi referensi penting dan sumber tambahan bagi peneliti yang ingin melakukan studi serupa dalam konteks lain atau dengan variabel yang berbeda yang membahas tentang dampak dari kualitas pelayanan perusahaan penyedia jasa transportasi terhadap kepuasan pelanggan.

3. Bagi Perusahaan

Capaian pada penelitian ini diharapkan bisa menghadirkan wawasan yang berharga kepada perusahaan dalam memanfaatkan teknologi *machine learning* untuk meningkatkan pelayanan dan kepuasan pelanggan. Hal ini dapat mendorong inovasi dan pengembangan berkelanjutan dalam penyediaan layanan transportasi yang lebih efisien dan memuaskan, terutama yang berkaitan dengan upaya peningkatan pelayanan terhadap kepuasan pengguna jasa transportasi bus.

1.4 Hipotesis

Hipotesis pada penelitian ini dapat dirumuskan seperti berikut ini:

H_0 : Tidak terdapat perbedaan yang signifikan pada kinerja algoritma *machine learning* dalam memprediksi kepuasan penggunaan jasa terhadap layanan transportasi bus rosalia indah.

H_1 : Terdapat perbedaan yang signifikan pada kinerja algoritma *machine learning* dalam memprediksi kepuasan penggunaan jasa terhadap layanan transportasi bus rosalia indah.

1.5 Batasan Masalah

Pada penelitian ini, terdapat batasan masalah yang perlu diperhatikan, yaitu :

Penelitian ini hanya terbatas pada penelitian tentang tingkat kepuasan pengguna transportasi bus Rosalia Indah yang penggunanya adalah pelanggan yang telah menggunakan jasa transportasi bus Rosalia Indah. Penelitian terbatas dilakukan dengan pengumpulan data melalui kuesioner kepada responden di sosial media. *Attribute* yang digunakan untuk menentukan tingkat kepuasan penggunaan jasa transportasi bus Rosalia Indah terdiri dari bus dengan *attribute* kualitas layanan secara keseluruhan dengan targetnya adalah tingkat kepuasan pengguna. Dalam pengolahan data, penelitian ini akan menggunakan *rapidminer* sebagai alat analisis data. *Algoritma* yang akan dibandingkan pada penelitian ini dibatasi dengan dua algoritma *machine learning* yang populer untuk menentukan tingkat akurasi kepuasan pengguna jasa transportasi bus rosalia indah.

BAB II

LANDASAN TEORI

2.1 Tinjauan Pustaka

Landasan teori pada penelitian ini mencakup sejumlah konsep penting tentang kepuasan pengguna jasa transportasi bus Rosalia Indah. Pemahaman terhadap konsep-konsep ini membantu dalam merancang penelitian dan menganalisis faktor utama apa saja yang berpengaruh kepada kepuasan pengguna jasa layanan transportasi bus Rosalia Indah.

2.1.1 Transportasi

Transportasi merupakan sebuah kendaraan yang dioperasikan oleh manusia untuk bermigrasi dari sebuah tempat ke tempat lain (Ilhamsyah et al., 2020). Menurut Miro “Transportasi merupakan proses pemindahan, pengerakkan, dan pengangkutan sebuah benda dari suatu tempat ke sebuah tempat dimana di tempat tersebut benda itu bermanfaat dalam mencapai suatu target” (Ferdila & Anwar, 2021). Nasution mengemukakan “Transportasi merupakan salah satu sektor penggerak perekonomian melalui pemindahan barang dan manusia (*the promoting sector*) dan penyedia layanan jasa (*the servicing sector*) bagi pembangunan ekonomi (Amir & Rahman, 2020).

Di Indonesia terdapat beberapa jenis moda transportasi massal, seperti darat, laut, dan udara. Dari beragam jenis transportasi, transportasi darat merupakan moda transportasi yang menjadi pilihan utama dan populer di masyarakat.

2.1.2 Kepuasan Pengguna

Menurut Kotler dan Keller mendefinisikan bahwa “Kepuasan pelanggan merupakan perasaan bahagia maupun kekecewaan pengguna yang disebabkan karena membandingkan performa yang dirasakan dari sebuah produk atau pelayanan dengan ekspektasi mereka” (Rooroh et al., 2020). Jika performa ataupun hasil belum memenuhi ekspektasi, maka pelanggan tersebut akan merasa kecewa. Jika kinerja sesuai dengan ekspektasi, kemudian pelanggan tersebut akan senang.

Menurut Pratama menyatakan bahwa “Beberapa indikator dalam menilai kepuasan pengguna meliputi: kesesuaian layanan yang memenuhi harapan, pelayanan yang sebanding dengan harga yang ditawarkan, dan kesesuaian layanan yang sebanding dengan yang diterima oleh pengguna” (Hamid & Susanti, 2023).

Dalam penelitian ini faktor utama yang digunakan untuk menentukan kepuasan pelanggan adalah bagaimana pengguna mempersepsikan kualitas layanan dan harga.

2.1.2.1 Kualitas Pelayanan

Tjiptono mengatakan bahwa “Kualitas pelayanan adalah usaha untuk memenuhi kebutuhan dan keinginan pelanggan, serta ketepatan penyampaian dengan cara yang sesuai dengan harapan pelanggan” (Hamid & Susanti, 2023). Pelanggan akan merasakan senang apabila mereka mendapatkan kualitas pelayanan memuaskan dan seperti dengan harapan yang mereka inginkan (Tirta Komara et al., 2021).

2.1.2.2 Harga

Tjiptono mengemukakan bahwa “Harga merupakan nilai moneter atau ukuran lain (produk dan layanan yang lain) yang diganti untuk mendapatkan hak menggunakan atau memiliki sebuah produk atau layanan” (Rooroh et al., 2020). Harga bisa menjadi indikator kepuasan pengguna, harga yang murah, harga yang wajar dan sesuai dengan kualitas layanan serta fasilitas yang dirasakan oleh pelanggan tentu akan menimbulkan rasa puas pada pelanggan (Setiawan & Frianto, 2021).

2.1.3 Machine Learning

Machine learning adalah algoritma matematika dalam program komputer mempelajari pola yang berasal dari data dan akan menghasilkan sebuah prediksi di masa depan (Saphari, 2023)

Machine learning merupakan bagian integral dari *Artificial Intelligence* (AI) yang memanfaatkan cara kerja ilmu komputer dan statistik guna menciptakan bentuk yang mencerminkan pola yang terdapat dalam data (Giarsyani, 2020). *Machine learning* berisi algoritma yang lazim atau umum yang mampu menciptakan sesuatu yang memikat serta berguna dari data tanpa memerlukan penulisan kode yang spesifik (Rahmat et al., 2023).

Pada penelitian ini penggunaan algoritma *machine learning* dibatasi pada beberapa algoritma yang populer untuk menentukan tingkat akurasi kepuasan pengguna jasa transportasi bus Rosalia Indah.

2.1.4 *Supervised Learning*

Supervised learning merupakan cara pembelajaran yang mengidentifikasi pola korelasi antara variabel input dan output (label) berdasarkan *data training* yang dibagikan (Agustina & Rahmah, 2022).

2.1.5 *Data Mining*

Data Mining adalah metode yang memakai ketrampilan menyerupai matematika dan statistik untuk mendapatkan informasi bernilai dari kelompok data, dan perusahaan bisa mendapatkan manfaat dari hal ini karena bisa membantu mereka mengevaluasi apa yang terjadi di perusahaannya, dan membuat prediksi apa yang akan terjadi di masa depan (Iwandini et al., 2023).

Data mining merupakan bagian pada tahap KDD (*Knowledge Discovery in Databases*) yang meliputi dari jumlah tahap seperti memilih data, *pre-processing*, transformasi, *data mining*, dan evaluasi kinerja (Zai, 2022).

2.1.6 *Pre-processing data*

Data Pre-processing atau pra-pemrosesan data adalah tahap dalam persiapan data yang bertujuan untuk mengolah data untuk menghasilkan data yang berkualitas dalam format yang sesuai dengan kebutuhan (Firdaus & Setiadi, 2023). *Data pre-processing* merupakan proses pengolahan data mentah untuk menjadi data yang relevan dengan tahap *data mining* yang dilakukan (Febriansyah et al., 2019). Tahapan *Pre-processing* ini tujuan utama adalah memastikan bahwa data yang akan digunakan untuk analisis telah dibersihkan serta diverifikasi keakuratannya. Berikut ini adalah beberapa metode yang digunakan dalam data *pre-processing*:

1. *Cleaning data*

Proses ini bertujuan untuk menghapus nilai yang tidak valid, memperbaiki kesalahan yang terdapat pada data dan mendeteksi inkonsistensi data yang mungkin ada pada dataset (Febriansyah et al., 2019).

2. *Labeling data*

Dataset yang telah dibersihkan akan diberikan label yang digunakan untuk mengklasifikasikan data uji ke dalam golongan atau kelas tertentu (Firdaus & Setiadi, 2023)

2.1.7 *Splitting data*

Split data adalah tahapan memisahkan dataset menjadi dua klasifikasi, yakni data uji dan data latih yang akan dimanfaatkan untuk keperluan algoritma (Prasetyo & Hermawan, 2023). Dalam tahapan ini dataset akan diubah menjadi dua kategori, yaitu data latih (*training*) dan data uji (*testing*).

A. *Data training*

Data training adalah himpunan data yang mempunyai atribut label atau kelas yang dimanfaatkan oleh mesin supaya bisa memahami karakteristik dataset sehingga dapat menghasilkan sebuah model atau pola data (Musu et al., 2021).

B. *Data testing*

Data testing merupakan sekumpulan data yang dilabeli atau dikelompokan yang dipakai dalam menilai ketepatan pola maupun model yang digunakan untuk mengklasifikasikan *data testing* (Musu et al., 2021).

2.1.8 *RapidMiner*

RapidMiner adalah sebuah perangkat lunak yang mengadopsi prinsip dari algoritma *data mining* untuk pengolahan data (Novianti, 2019). *RapidMiner* beroperasi menggunakan ide *data mining* dengan mengambil pola dari kumpulan data berukuran besar, dan menggabungkannya menggunakan metode statistika, *artificial intelligence*, serta menyusunnya dalam format *database* (Febriansyah et al., 2019).

Pada penelitian ini metode yang diterapkan adalah algoritma *machine learning* yang dibatasi pada beberapa algoritma yang populer untuk menentukan tingkat akurasi kepuasan pengguna jasa transportasi bus Rosalia Indah.

2.1.9 *Confusion Matrix*

Confusion matrix merupakan alat yang dapat digunakan untuk mengevaluasi kinerja sebuah model dalam mengklasifikasikan dataset pada perangkat lunak *machine learning* (Mardiyyah et al., 2024). Salah satu tujuan utama dari *confusion matrix* adalah untuk menghitung nilai *accuracy*, *precision* dan *recall* sebagai representasi dari prediksi dan keadaan aktual dari data yang dihasilkan oleh algoritma *machine learning* (Mardiyyah et al., 2024).

Metode ini bekerja dengan cara membandingkan matriks prediksi dengan kelas asli yang memuat informasi aktual serta nilai klasifikasi yang diprediksi (Styawati et al., 2021).

2.1.10 *Kurva Roc*

Kurva ROC merupakan visualisasi grafik dua dimensi yang menggambarkan *confusion matrix* dan memvisualisasikan akurasi serta perbandingan klasifikasi (Maulana et al., 2023). Kurva ROC digambarkan dengan dua buah garis berwarna

biru dan merah, yaitu garis *horizontal* menggambarkan *false positif* dan garis *vertikal* menggambarkan *true positif*.

2.2 Penelitian Terkait

Penulis pada pembuatan skripsi ini memaparkan beberapa penelitian terkait yang relevan dan berhubungan dengan latar belakang masalah yang dibahas dalam skripsi ini. Pada Table II.1 rangkuman beberapa penelitian terkait yang berhubungan dengan skripsi ini dan berikut penjelasannya:

Dalam penelitian yang dilakukan oleh (Diansyah, 2022) yang berjudul “Klasifikasi Tingkat Kepuasan Pengguna dengan Menggunakan Metode *K-Nearest Neighbour (KNN)*” Tujuan dari penelitian ini untuk mengklasifikasikan tingkat kepuasan pengguna layanan bus Trans Metro Pekanbaru, dengan tujuan mendapatkan pengetahuan terkait tingkat kepuasan penumpang. Data yang digunakan sebanyak 170 set data yang dikumpulkan melalui kuesioner yang disebarakan kepada penumpang bus Trans Metro Pekanbaru. Untuk mencari nilai k optimal, penelitian ini menggunakan metode *k-Fold Cross Validation*, sementara dalam mengklasifikasi tingkat kepuasan pengguna, digunakan metode *K-Nearest Neighbor*. Untuk akurasi yang dicapai dari pengujian ini menghasilkan akurasi sebesar 94,12% dengan nilai k optimal adalah 5.

Penelitian berikutnya oleh (Iwandini et al., 2023) yang berjudul “Analisa Sentimen Pengguna Transportasi Jakarta Terhadap Transjakarta Menggunakan Metode *Naives Bayes* dan *K-Nearest Neighbor*”. Penelitian ini bertujuan untuk mengembangkan sistem yang mampu mendeteksi sentimen positif dan negatif dalam informasi publik, khususnya dalam konteks data Twitter terkait Transjakarta. Sistem ini menggunakan metode *Naive Bayes Classifier* dan *K-Nearest Neighbor Classifier*

untuk klasifikasi sentimen, dengan pembobotan TF-IDF sebagai teknik ekstraksi fitur dengan masukan sistem berupa tweet tentang Transjakarta. Hasil pengujian menunjukkan bahwa metode *Naive Bayes* menghasilkan akurasi sebesar 61,1%, sedangkan metode *K-Nearest Neighbor* mencapai akurasi 75,7%. Dapat dilihat bahwa *K-Nearest Neighbor* lebih unggul dalam menentukan sentimen dari data Twitter terkait penggunaan transportasi Transjakarta.

Penelitian selanjutnya dilakukan oleh (Salsabilah et al., 2023) yang berjudul “Komparasi Algoritma *Naive Bayes* dan *K Nearest Neighbor* dalam Kepuasan Pengguna Fitur *TikTok Shop*” Menggali faktor-faktor yang memengaruhi kepuasan pengguna terhadap fitur TikTok Shop pada aplikasi tersebut. Studi ini merespons pertumbuhan transaksi jual beli online yang semakin meningkat di Indonesia, dengan fokus pada fitur *TikTok Shop* yang diperkenalkan pada tahun 2021 sebagai bagian dari *social commerce*. Penelitian ini membandingkan algoritma *Naive Bayes* dan *K Nearest Neighbor* dalam menganalisis data, dengan fokus pada pengaruh *service quality*, *system quality*, dan *information quality*. Hasil penelitian menunjukkan bahwa *service quality* memiliki pengaruh tertinggi terhadap kepuasan pengguna *TikTok Shop*, diikuti oleh *system quality*, dan *information quality*. Selain itu, perbandingan algoritma menunjukkan bahwa *Naive Bayes* mencapai bobot akurasi tertinggi sebesar 98.36% dengan pengelompokan data latih sebesar 70% dan data uji sebesar 30%

Penelitian yang selanjutnya dilakukan oleh (Yuliarina & Hendry, 2022) yang berjudul “*Comparison of Prediction Analysis of Gofood Service Users Using the Knn & Naive Bayes Algorithm With Rapidminer Software*” Pada penelitian tersebut menyoroti pentingnya peran *GoFood* sebagai penyedia layanan jasa yang semakin esensial dalam kehidupan modern. Oleh karena itu, diperlukan analisis survei untuk

mengumpulkan data kepuasan pelanggan yang dapat digunakan dalam memprediksi kepuasan pengguna layanan *GoFood*. Algoritma *K-Nearest Neighbor (KNN)* dan *Naive Bayes* digunakan untuk menguji 1.000 data pengguna *GoFood* yang telah dikumpulkan. Hasil pengujian dengan *Cross Validation* dan *T-Test* menunjukkan bahwa algoritma *KNN* memberikan performa terbaik dengan akurasi 98,80% dan *recall* 100%, sedangkan algoritma *Naive Bayes* menghasilkan akurasi 94,10% dan *recall* 94,43%. Hasil ini menunjukkan bahwa algoritma *KNN* lebih unggul dalam memprediksi kepuasan pengguna layanan *GoFood*.

Penelitian yang selanjutnya dilakukan oleh (Hakim & Sugiyono, 2024) yang berjudul “Analisa Sentimen Terhadap Kereta Cepat Jakarta - Bandung Menggunakan Algoritma *Naive Bayes* Dan *K-Nearest Neighbor*” pada penelitian tersebut digunakan Penelitian ini menggunakan metode *Naive Bayes* dan *K-Nearest Neighbor (KNN)* untuk mengklasifikasikan 2.000 tweet, yang terdiri dari 1.000 data latih dan 1.000 data uji. Hasil analisis menunjukkan bahwa 44,18% masyarakat memiliki sentimen positif, sementara 55,82% bersentimen negatif terhadap proyek kereta cepat ini. Temuan ini menunjukkan bahwa sebagian besar masyarakat Indonesia cenderung memiliki respon negatif terhadap proyek tersebut. Berdasarkan hasil klasifikasi menggunakan algoritma *KNN* dengan *split data* 0.8 : 0.2 dan nilai $k=3$, diperoleh akurasi sebesar 99,76%, yang mengindikasikan bahwa algoritma *KNN* mampu mengklasifikasikan data secara efektif dan akurat.

Tabel II. 1 Penelitian Terkait

Tahun	Judul	Penulis	Metode	Hasil
2022 – Jurnal Sistim Informasi dan Teknologi (Jsisfotek)	Klasifikasi Tingkat Kepuasan Pengguna dengan Menggunakan Metode <i>K-Nearest Neighbour</i> (<i>KNN</i>)	Surya Diansyah	Dalam penelitian ini, dilakukan pencarian nilai <i>k</i> optimal menggunakan teknik <i>k-Fold Cross Validation</i> , sementara untuk mengklasifikasi tingkat kepuasan layanan pengguna bus dengan menggunakan metode <i>K-Nearest Neighbor</i> .	Tingkat akurasi yang dihasilkan dari pengujian ini adalah sebesar 94.12% dengan nilai <i>k</i> optimal adalah <i>k</i> = 5.
2023 – <i>Journal of Information System Research</i> (<i>JOSH</i>)	Analisa Sentimen Pengguna Transportasi Jakarta Terhadap Transjakarta Menggunakan Metode <i>Naives Bayes</i> dan <i>K-Nearest Neighbor</i>	Ismia Iwandini, Agung Triayudi, & Gatot Soepriyono	Metode yang digunakan dalam klasifikasi sentimen ini yaitu perbandingan metode <i>Naive Bayes Classifier</i> dan juga <i>K-Nearest Neighbor Classifier</i> dengan pembobotan TF-IDF.	Dari hasil pengujian didapatkan bahwa akurasi algoritma <i>Naive Bayes</i> untuk analisis sentimen data <i>Twitter</i> terkait penggunaan transportasi Transjakarta mencapai 61,1%, sedangkan metode <i>K-Nearest Neighbor</i> memperoleh hasil akurasi sebesar 75,7%. Hal ini menunjukkan metode <i>K-nearest-neighbor</i>

				memberikan tingkat akurasi yang lebih tinggi.
2023 – <i>Journal of Emerging Information System and Business Intelligence (JEISBI)</i>	Komparasi Algoritma <i>Naive Bayes</i> dan <i>K-Nearest Neighbor</i> dalam Kepuasan Pengguna Fitur <i>Tiktok Shop</i> .	Balqist Naurah Salsabilah & I Kadek Dwi Nuryana	Penelitian ini menggunakan algoritma <i>Naive Bayes</i> dan <i>K Nearest Neighbor</i> karena hasil tingkat akurasi dalam melakukan klasifikasi data memberikan hasil yang baik. Untuk mencari tahu faktor-faktor yang memiliki pengaruh yakni menggunakan <i>Gain Ratio</i> dan mengurutkannya berdasarkan yang paling tinggi hingga terendah.	Hasil yang didapatkan dari pengujian yakni <i>service quality</i> menjadi pengaruh paling tinggi pada kepuasan pengguna fitur <i>Tiktok Shop</i> . Hasil dari perbandingan algoritma, <i>Naive Bayes</i> memiliki bobot akurasi terbaik sebesar 98.36% dengan 70% <i>data training</i> dan 30% <i>data testing</i> .
2022 – Jurnal Teknik Informatika (JUTIF)	<i>Comparison of Prediction Analysis of Gofood Service Users Using the Knn & Naive Bayes Algorithm With Rapidminer Software.</i>	Agista Nindy Yuliarina & Hendry	Dengan memanfaatkan algoritma <i>KNN</i> dan <i>Naive Bayes</i> , yang keduanya merupakan algoritma yang optimal untuk melakukan pengujian terhadap 1.000 data pengguna <i>GoFood</i> yang telah dikumpulkan.	Hasil pengujian menggunakan <i>Cross Validation</i> dan <i>T-Test</i> menyatakan algoritma <i>KNN</i> adalah algoritma paling unggul dengan <i>Accuracy</i> 98,80% dan <i>Recall</i> 100% dibandingkan <i>Naive Bayes</i> yang mendapatkan hasil <i>Accuracy</i> sebesar 94,10% dan <i>Recall</i> sebesar 94,43%

2024 – Jurnal Sains dan Teknologi	Analisa Sentimen Terhadap Kereta Cepat Jakarta – Bandung Menggunakan Algoritma <i>Naïve Bayes</i> Dan <i>K- Nearest Neighbor</i> .	Zaky Rahman Hakim & Sugiyono	Penelitian ini menggunakan metode <i>Naïve Bayes</i> dan metode <i>K-Nearest Neighbor</i> .	Berdasarkan klasifikasi model algoritma <i>K- Nearest Neighbor</i> , dengan split data memakai rasio 0.8 : 0.2 dengan nilai $k = 3$ pada dataset kereta cepat Jakarta – Bandung, diperoleh nilai <i>accuracy</i> sebesar 99.76%. Hal ini menunjukkan bahwa algoritma <i>K- NN</i> dapat melakukan klasifikasi data dengan tingkat akurasi yang sangat tinggi dan efektif.
--	---	---------------------------------------	--	--

Sumber: Bahan Penelitian

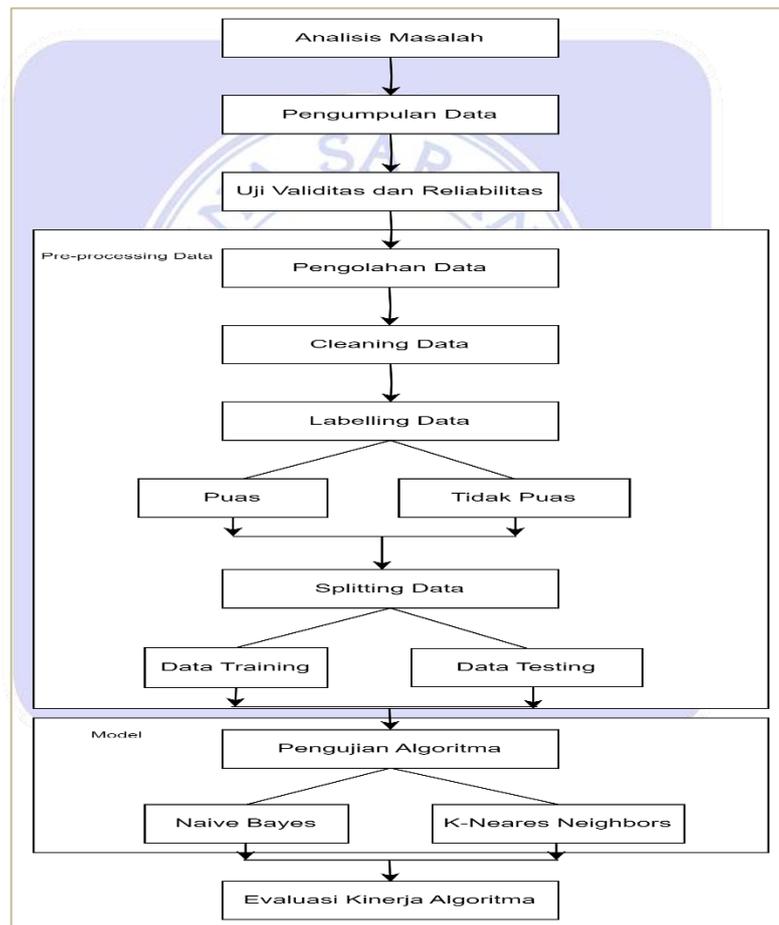


BAB III

METODOLOGI PENELITIAN

3.1 Proses dan Langkah Penelitian

Proses dan langkah penelitian dalam penyusunan skripsi tentang “Komparasi Algoritma *Machine Learning* Dalam Memprediksi Kepuasan Penggunaan Jasa Transportasi Bus Rosalia Indah” digambarkan dalam gambar III.1 berikut:



Sumber: Hasil Penelitian

Gambar III. 1 Kerangka Penelitian

3.2 Metode Pengolahan dan Analisis Data

Metode mengolah dan menganalisis data pada penelitian ini dilakukan untuk menjawab pertanyaan penelitian mengenai komparasi algoritma *machine learning* dalam memprediksi kepuasan pengguna bus rosalia indah.

3.2.1 Analisis Masalah

Pada tahapan ini, diidentifikasi masalah yang sebenarnya dan faktor apa saja yang dibutuhkan untuk menganalisis kepuasan pengguna jasa transportasi bus Rosalia Indah dengan menerapkan metode *data mining* menggunakan dua algoritma *machine learning* yang populer. Metode analisis data yang dipakai untuk penelitian ini merupakan analisis kuantitatif yaitu proses mengumpulkan, menganalisis, dan menafsirkan data numerik untuk mengidentifikasi pola, hubungan, atau tren yang terdapat dalam data tersebut. Data dikumpulkan memakai alat atau perangkat penelitian, dan analisisnya dilakukan secara kuantitatif untuk pengujian hipotesis yang telah dirumuskan.

3.2.2 Pengumpulan Data

Penelitian ini memakai data primer yang berasal dari pengumpulan data secara langsung oleh peneliti atau pihak pertama yang didalamnya berisi penilaian pelanggan mengenai kualitas keseluruhan pelayanan jasa transportasi bus rosalia indah dan informasi yang didapatkan langsung dari sumber utama, yakni pelanggan.

Adapun langkah-langkah yang dilakukan dalam proses pengambilan data adalah sebagai berikut:

1. Pembuatan Kuesioner

Kuesioner dibuat menggunakan platform *Google Form* untuk memudahkan pengumpulan data secara *online*.

2. Penyusunan Kuesioner

Pertanyaan kuesioner dirancang untuk menggali informasi tingkat kepuasan pengguna Bus Rosalia di sosial media. Pertanyaan mencakup faktor-faktor seperti kualitas pelayanan, harga, tingkat kepuasan pelanggan pengguna transportasi bus rosalia indah, dan data demografis pengguna. Dalam merancang pertanyaan peneliti menyediakan 5 pilihan jawaban pertanyaan menggunakan skala likert. dimana masing-masing opsi jawaban diberi nilai (1, 2, 3, 4, dan 5) seperti yang digambarkan pada Tabel III.1 dibawah ini.

Tabel III. 1 Skala Likert

No	Keterangan	Nilai
1	Sangat Tidak Setuju	1
2	Tidak Setuju	2
3	Netral	3
4	Setuju	4
5	Sangat Setuju	5

Sumber: Hasil Penelitian

3. Distribusi Kuesioner

Data dikumpulkan dengan memberikan daftar pertanyaan kepada responden melalui kuesioner yang kemudian disebarluaskan secara acak melalui *platform online* menggunakan media sosial kepada para pengguna bus Rosalia Indah yang pernah menggunakan jasa transportasi bus Rosalia Indah.

4. Populasi dan Sampel

Dalam penelitian kuantitatif, populasi didefinisikan sebagai area generalisasi yang mencakup objek atau subjek dengan kualitas dan karakteristik tertentu yang ditentukan oleh peneliti untuk diteliti dan kemudian digunakan untuk dibuat kesimpulan (Suriani et al., 2023). Populasi pada penelitian ini adalah pengguna media sosial yang pernah menggunakan jasa bus Rosalia Indah.

Sampel adalah bagian dari populasi yang karakteristiknya ingin diteliti dan dianggap mampu mewakili populasi secara keseluruhan (Suriani et al., 2023). Metode pengampilan sampel yang diaplikasikan pada penelitian ini adalah rumus slovin seperti gambar berikut:

$$n = \frac{N}{1 + N \times e^2}$$

Sumber: Bahan Penelitian

Gambar III. 2 Rumus Slovin

Keterangan:

n = Ukuran sampel yang dibutuhkan

N = Jumlah total populasi

e = *Margin of error* yang diinginkan

Dalam penelitian ini jumlah populasi (N) yang diperlukan adalah 300 orang. Dan besarnya *Margin of error* yang diinginkan (e) sebesar 1% sehingga diperoleh perhitungan sebagai berikut:

$$n = \frac{300}{1 + 300 \times 0.01^2} \quad n = 291$$

Berdasarkan perhitungan diatas, sampel yang diperlukan adalah sekitar 291 responden. Sampel pada penelitian ini adalah responden pengguna jasa transportasi bus Rosalia Indah. Teknik sampling yang diterapkan pada penelitian ini adalah *purposive sampling* yaitu teknik sampling berdasarkan kriteria. Adapun kriteria yang digunakan pada penelitian adalah:

1. Responden pernah menggunakan jasa transportasi bus Rosalia Indah.
2. Berusia lebih dari 15 tahun.

3.3 Uji Validitas dan Reliabilitas

Pada tahapan ini dilakukan uji validitas dan realibitas terhadap jawaban yang telah dikumpulkan melalui kuesioner dengan langkah-langkah berikut:

3.3.1 Uji Validitas

Uji validitas dilakukan untuk mengevaluasi keterkaitan atau hubungan antar item pertanyaan dalam satu variabel. Alat ukur yang digunakan dalam pengujian validitas adalah kuesioner berisi daftar pertanyaan yang telah diisi oleh responden untuk melihat apakah sebuah alat ukur tersebut valid atau tidak valid. Sebuah kuesioner dinyatakan valid apabila pertanyaan pada kuesioner tersebut dapat memaparkan sesuatu yang diukur oleh kuesioner (Janna & Herianto, 2021).

Data atau jawaban kuesioner dikatakan valid apabila nilai r hitung $>$ r tabel, namun apabila nilai r hitung $<$ r tabel maka data atau jawaban kuesioner dinyatakan tidak valid, hal tersebut dibantu dengan program SPSS (*Statistical Product and Service Solution*).

3.3.2 Uji Reliabilitas

Reliabilitas adalah serangkaian pengukuran atau serangkaian alat ukur yang memiliki konsistensi ketika pengukuran tersebut dilakukan berulang kali dengan alat ukur yang sama (Rooroh et al., 2020). Pada penelitian ini metode uji reliabilitas menggunakan metode *Cronbach's Alpha*.

Suatu variabel dikatakan reliabel apabila koefisien *Cronbach Alpha* menunjukkan nilai $>$ 0,60 dan sebuah kuesioner dikatakan kurang reliabel jika koefisien *Cronbach Alpha* menunjukkan nilai $<$ 0,60 (Kumrotin & Susanti, 2021)

3.4 Pengolahan Data

Pengolahan data pada penelitian ini dilakukan dengan langkah-langkah berikut:

1. Preprocessing data

Tahapan preprocessing data terdiri dari:

a. Cleaning data

bertujuan untuk mengumpulkan data yang kemudian dibersihkan guna menghilangkan *missing values*, duplikasi, serta data yang tidak relevan.

b. Pelabelan data

Selanjutnya data yang sudah dibersihkan akan diberi label menjadi dua klasifikasi, yaitu Puas dan Tidak Puas.

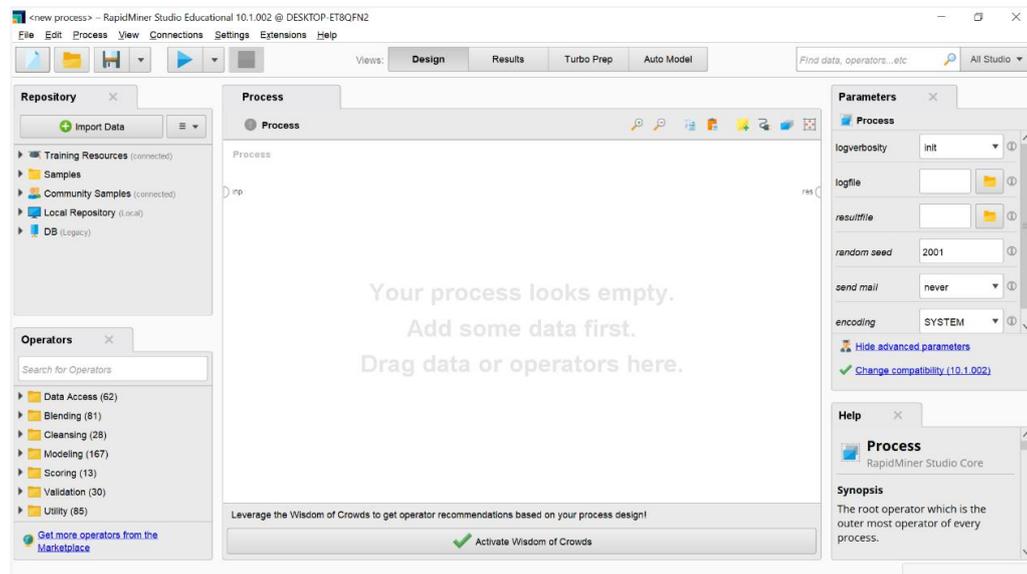
2. Splitting Data

Dari dataset yang telah dikumpulkan akan diuji menggunakan operator *split data*, dan dataset tersebut akan dibentuk menjadi dua tipe, yaitu data latih (*data training*) dan data uji (*data testing*) yang kemudian diujikan menggunakan beberapa algoritma *machine learning* dengan bantuan *rapidminer*.

3.5 Pengujian Algoritma

Pada tahap ini, peneliti menentukan serta mengimplemetasikan metode yang sesuai agar bisa memperoleh hasil akurasi yang sempurna. Dalam penelitian ini algoritma yang diterapkan untuk pengujian yaitu algoritma *naive bayes*, dan *k-neares neighbors*. Pemilihan algoritma didasarkan pada popularitas dan efektivitas dalam tugas klasifikasi. Pengujian dijalankan dengan berbagai pengujian memanfaatkan

operator *split data* dengan menggunakan *tools software RapidMiner* dengan tampilan pada gambar III.3..



Sumber: Hasil Penelitian

Gambar III. 3 Interface RapidMiner

3.5.1 Algoritma Naïve Bayes

Algoritma *naïve bayes* adalah metode *data mining* yang bisa dipakai untuk pengelompokkan, dan dapat menyediakan hasil yang bisa memperkirakan peluang yang terjadi di masa depan berdasarkan hasil klasifikasi (Gunawan et al., 2021).

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

Sumber (Gunawan et al., 2021)

Gambar III. 4 Rumus Naïve Bayes

Keterangan:

1. X : Data yang *classnya* tidak diketahui.
2. H : Hipotesis atau *class* yang ingin diteliti.
3. P(H|X) : Probabilitas posterior dari kelas H setelah diberikan data X.
4. P(H) : Probabilitas prior dari kelas H.
5. P(X|H) : Probabilitas *likelihood* X jika kelas H adalah benar.

6. $P(X)$: Probabilitas prior dari kelas X.

3.5.2 K-Nearest Neighbors

Metode *k-nearest neighbor* (k-nn) merupakan sebuah teknik memanfaatkan algoritma *supervised* dimana hasil dari *Query instance* bisa dikategorikan menurut sebagian besar label *class* pada *k-nn* (Kartarina et al., 2021). *K-nn* dijalankan dengan mengidentifikasi sekumpulan k objek pada *data training* yang mendekati (serupa) seperti objek dalam data yang baru (*data testing*), sampai memungkinkan untuk menentukan kesamaan antara *data testing* dan *data training* sehingga kecocokan banyaknya data dapat mempengaruhi hasil dari penelitian tersebut (Wu dalam penelitian Muharrom, 2019).

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^n (a_r(x_i) - a_r(x_j))^2}$$

Sumber (Gunawan et al., 2021)

Gambar III. 5 Rumus K-Neares Neighbors

Keterangan :

1. $d(x_i, x_j)$: Jarak *Euclidean* antara data (x_i) dan (x_j)
2. (x_i) : data ke- i
3. (x_j) : data ke- j
4. (a_r) : himpunan data yang akan dibandingkan
5. i, j : posisi dari atribut yang ingin dibandingkan dari kedua data.

3.6 Evaluasi Kinerja

Pemilihan fitur-fitur yang paling relevan untuk analisis dilakukan untuk meningkatkan akurasi model evaluasi kinerja model dilakukan dengan menggunakan data uji. Metode pengujian yang dipilih oleh peneliti dalam penelitian ini adalah

dengan menggunakan *tools software rapidminer* untuk mendapatkan *accuracy* yang dipakai untuk mengukur performa masing-masing algoritma *machine learning*. Selain itu, *confusion matrix* juga digunakan untuk memberikan gambaran lebih detail mengenai prediksi yang benar (positif) dan salah (negatif) dan Kurva ROC untuk menilai hasil prediksi.

1. *Confusion Matrix*

Confusion matrix menggunakan evaluasi model *confussion matrix* seperti pada tabel III.2.

Tabel III. 2 Confusion Matrix

KLASIFIKASI	KELAS PREDIKSI	
	Class= Puas	Class= Tidak Puas
class precision	Class= Puas	Class= Tidak Puas
	Class= Puas Positive- (TP)	Class= Tidak Puas Negative- (FN)
	Class- Tidak Puas Positive- (FP)	Class- Puas Negative (TN)

Sumber : Hasil Penelitian

Tingkat akurasi dapat dihitung dengan menerapkan rumus sebagai berikut:

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

TP : Jumlah data positif yang diprediksi sebagai positif dengan benar oleh model.

FP : Jumlah data yang seharusnya negatif tetapi salah diprediksi sebagai positif oleh model.

TN : Jumlah data negatif yang diprediksi sebagai negatif dengan benar oleh model.

FN : Jumlah data yang sebenar positif yang tetapi salah dikategorikan sebagai Negatif oleh model.

2. Kurva ROC

Kurva ROC diilustrasikan dengan garis *horizontal* yang menunjukkan *false positif* dan garis *vertikal* menunjukkan *true positif* seperti pada gambar III.5. Teknik yang sering digunakan untuk mengelompokkan tingkat keakuratan adalah menggunakan AUC.



Sumber : Hasil Penelitian

Gambar III. 6 Kurva ROC

Untuk pengklasifikasian *data mining*, nilai AUC terbagi dalam sejumlah kategori:

- a. 0.90-1.00 = *Excellent Classification*
- b. 0.80-0.90 = *Good Classification*
- c. 0.70-0.80 = *Fair Classification*
- d. 0.60-0.70 = *Poor Classification*
- e. 0.50-0.60 = *Failure*

BAB IV

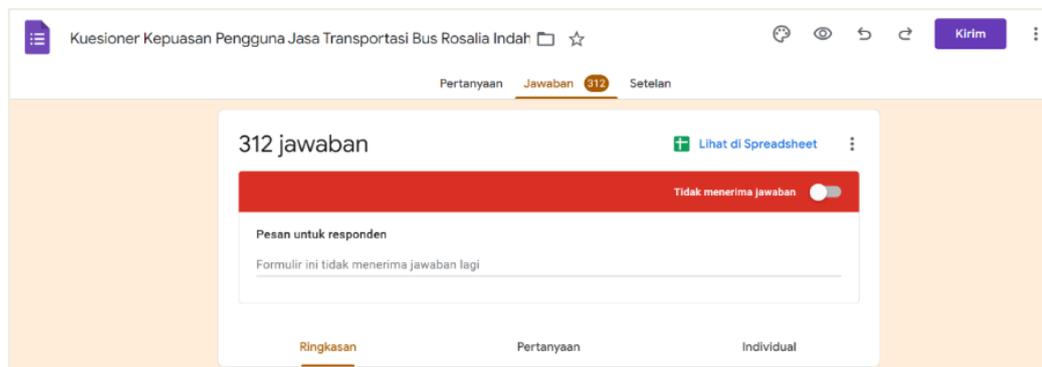
HASIL DAN PEMBAHASAN

4.1 Hasil Penelitian

Penelitian ini dilakukan untuk memahami sejauh mana pengguna merasakan kepuasan terhadap layanan bus rosalia indah. Data yang dianalisa adalah data kepuasan pengguna bus rosalia indah yang dikumpulkan cara memberikan daftar pertanyaan kepada responden dengan cara menyebarkan secara acak kuesioner penelitian kepada pengguna bus rosalia indah di media sosial (*facebook, instagram, dan twitter*). Data tersebut akan dianalisa dengan memanfaatkan beberapa algoritma *machine learning* yang populer seperti *naïve bayes*, dan *k-nearest neighbors* menggunakan bantuan tools *RapidMiner*.

4.1.1 Hasil Pengumpulan data

Dalam upaya meningkatkan kualitas layanan dan memahami tingkat kepuasan pelanggan, penulis telah melaksanakan survei selama satu bulan dengan menggunakan kuesioner yang dibagikan kepada para pengguna bus Rosalia Indah melalui media sosial. Dari survei tersebut, terkumpul data sebanyak 312 responden.



Sumber: Hasil Penelitian

Gambar IV. 1 Form Kuesioner

Adapun beberapa faktor yang digunakan dalam kuesioner ini, diantaranya:

a. Data Demografis Pengguna

Berisi data responden, seperti nama, usia, jenis kelamin, domisili.

b. Kualitas Pelayanan

Berisi penilaian responden terhadap pelayanan yang diberikan pihak rosalia indah, seperti keselamatan, kenyamanan, keamanan, fasilitas, dan lain-lain.

c. Harga

Berisi penilaian responden terhadap harga yang diberikan pihak rosalia indah untuk dibandingkan.

d. Tingkat Kepuasan Pengguna

Berisi penilain responden terhadap aspek tingkat kepuasan penggunaan bus secara keseluruhan.

Hasil dari penelitian terhadap kepuasan pengguna bus rosalia indah bisa dilihat pada tabel IV.1

Tabel IV. 1 Dataset Kepuasan Pengguna Jasa Bus Rosalia Indah

Nama Lengkap	Umur	Jenis Kelamin	Pekerjaan	Frekuensi menggunakan bus	Domisili	Layanan bus Rosalia Indah	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12	B13	C1	C2	C3	C4	D1	D2	D3	D4	D5		
Valent Aif Naryento	21:30 tahun	Laki-laki	Mahasiswa	Pelajar	12 kali	Jabodetabek	Keluarga	4	5	5	5	5	4	4	5	4	5	5	4	5	4	4	4	4	4	5	5	4	5	5
AHMAD SYAFRI WIJAYA	21:30 tahun	Laki-laki	Mahasiswa	Pelajar	Jarang menggunakan	Jabodetabek	Keluarga	1	2	2	2	1	3	2	1	2	1	2	1	2	2	1	2	2	1	2	5	2	1	2
Ahlan Rochmatul Abdan	21:30 tahun	Laki-laki	Karyawan	Jarang menggunakan	Jabodetabek	Langganan sendi	5	4	5	4	3	5	3	5	4	5	5	3	3	4	3	4	5	5	4	5	5	4	5	4
Trianov Vidyandaru	21:30 tahun	Laki-laki	Freelance	Jarang menggunakan	Jabodetabek	Keluarga	4	4	3	4	2	4	3	3	3	3	4	4	5	3	3	1	3	4	3	3	3	3	4	
Diren Bahkar Khan	21:30 tahun	Laki-laki	Belum bekerja	Jarang menggunakan	Jabodetabek	Media Sosial	4	5	4	4	4	4	4	5	4	5	4	4	3	4	4	4	3	4	3	3	3	3	4	
Thang Saqid Imanullah	15:20 tahun	Laki-laki	Mahasiswa	Pelajar	Jarang menggunakan	Lampung	Busmania	5	5	5	5	5	4	5	5	5	5	5	5	3	1	5	2	5	2	5	5	5	5	
Ringga Raka	21:30 tahun	Laki-laki	Karyawan	12 kali	Jabodetabek	Keluarga	4	4	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	
Anis Jaka Purnama	21:30 tahun	Laki-laki	Karyawan	Jarang menggunakan	Surabaya	Media Sosial	3	5	3	4	4	4	5	5	3	5	5	3	5	5	5	5	5	5	4	4	5	4	4	
Yedi	21:30 tahun	Laki-laki	Wiraswasta	Jarang menggunakan	Jabodetabek	Keluarga	5	3	3	4	5	5	2	5	5	4	5	5	3	3	4	5	4	5	4	5	4	5	4	
Ahino Rifat Hasan	15:20 tahun	Laki-laki	Mahasiswa	Pelajar	12 kali	Surabaya	Media Sosial	5	5	4	5	5	4	5	5	5	5	5	5	5	5	5	5	5	5	4	5	4	5	
Adelvudha Saputra	15:20 tahun	Laki-laki	Mahasiswa	Pelajar	3-5 kali	Jabodetabek	Media Sosial	5	5	4	5	4	4	5	5	4	5	4	4	4	5	4	5	5	5	5	5	5	5	
Fano	21:30 tahun	Laki-laki	Mahasiswa	Pelajar	Jarang menggunakan	Lampung	Keluarga	3	4	3	4	2	4	4	5	3	3	3	2	2	4	4	3	4	4	3	4	4	3	
Gabriel Adi	21:30 tahun	Laki-laki	Mahasiswa	Pelajar	3-5 kali	Jabodetabek	Keluarga	4	5	3	5	3	3	4	5	4	4	5	4	3	5	5	4	3	4	4	5	5	4	
Suebun Sofyan	21:30 tahun	Laki-laki	Karyawan	12 kali	Jabodetabek	Keluarga	1	5	5	4	5	4	4	4	2	5	3	1	5	5	5	3	4	3	4	3	4	3	4	
Muhammad Ghaisan Pad	15:20 tahun	Laki-laki	Mahasiswa	Pelajar	3-5 kali	Yogyakarta	Keluarga	3	5	3	4	4	5	4	4	5	4	5	4	3	3	4	5	4	4	3	5	4	4	
Gilang Albar	21:30 tahun	Laki-laki	Karyawan	Jarang menggunakan	Jabodetabek	Keluarga	4	5	3	4	4	2	4	5	5	4	5	3	3	4	4	3	4	3	4	5	5	5		
M Rizki Albar	15:20 tahun	Laki-laki	Mahasiswa	Pelajar	Jarang menggunakan	Jember	Keluarga	5	4	4	5	3	5	5	3	4	3	5	2	3	3	5	4	4	3	4	3	5		
Marvin Adhyas Anggoro	15:20 tahun	Laki-laki	Mahasiswa	Pelajar	Jarang menggunakan	Jabodetabek	Media Sosial	4	4	4	5	4	5	5	5	5	5	5	4	5	4	4	5	4	5	4	5	5		
Elang Raka Prayoga	21:30 tahun	Laki-laki	Mahasiswa	Pelajar	12 kali	Jabodetabek	Media Sosial	5	5	5	5	5	3	4	5	5	5	5	5	3	3	5	4	5	5	5	5	5		
Arif Achyrian	21:30 tahun	Laki-laki	Karyawan	12 kali	Jabodetabek	agen dekat rumah	5	4	4	5	4	5	4	5	4	5	4	5	4	5	4	5	4	5	4	4	5	4	5	
ian	21:30 tahun	Laki-laki	Mahasiswa	Pelajar	Jarang menggunakan	Jabodetabek	Media Sosial	4	5	5	5	4	4	5	5	5	5	5	3	4	4	5	5	5	5	5	5	5		
Muhammad Riza	15:20 tahun	Laki-laki	Mahasiswa	Pelajar	Jarang menggunakan	Jabodetabek	Media Sosial	4	4	3	3	5	5	4	3	4	3	1	3	3	4	3	4	3	4	3	4	4		
Muhammad Araya Swal	21:30 tahun	Laki-laki	Wiraswasta	12 kali	Semarang	Media Sosial	5	4	4	5	5	4	4	5	5	5	5	3	4	5	5	3	5	5	5	5	5	4		
Alii Ton	15:20 tahun	Laki-laki	Mahasiswa	Pelajar	Jarang menggunakan	Jabodetabek	Media Sosial	4	5	3	4	4	4	4	4	4	4	4	4	4	4	4	5	4	2	4	4	4		
Ferd Firdaus	21:30 tahun	Laki-laki	Mahasiswa	Pelajar	Jarang menggunakan	Surabaya	Media Sosial	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5		
Fadri Khyal Ham Sunu	15:20 tahun	Laki-laki	Mahasiswa	Pelajar	Jarang menggunakan	Jawa Tengah	Media Sosial	5	5	5	4	5	4	5	5	5	5	5	4	5	5	4	4	4	3	5	5			
Vitoria Pratama	21:30 tahun	Laki-laki	Karyawan	12 kali	Jabodetabek	Media Sosial	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5			
Yovie Athab Yudistira	15:20 tahun	Laki-laki	Mahasiswa	Pelajar	12 kali	Yogyakarta	Illian	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5			

Sumber: Hasil Penelitian

Keterangan :

- Nama : Nama responden
- Umur : Umur responden

- Jenis Kelamin : Jenis kelamin responden
- Pekerjaan : Profesi responden
- Frekuensi menggunakan bus : Seberapa sering responden menggunakan bus Rosalia Indah dalam sebulan
- Domisili : Tempat tinggal responden
- layanan bus Rosalia Indah : Darimana responden mengetahui informasi mengenai layanan bus Rosalia Indah.
- B1: Penilaian terhadap ketepatan waktu keberangkatan dan kedatangan bus.
- B2: Penilaian terhadap kemudahan dalam pemesanan/reservasi tiket bus Rosalia Indah melalui agen (Offline) maupun melalui website/aplikasi Rosalia Indah (Online).
- B3 : Penilaian terhadap responsivitas layanan pelanggan (Rosalia Indah Contact Center) dalam menangani keluhan atau pertanyaan.
- B4 : Penilaian terhadap keramahan dan profesionalisme Staf bus (Agen) bus Rosalia Indah dalam melayani pembelian tiket bus.
- B5 : Penilaian terhadap fasilitas ruang tunggu yang disediakan agen bus Rosalia Indah
- B6 : Penilaian terhadap fasilitas yang disediakan di dalam bus Rosalia Indah (misalnya, Selimut, Bantal, AC, Wi-Fi, Snack, toilet)
- B7 : Penilaian terhadap kondisi armada bus Rosalia Indah yang digunakan
- B8 : Penilaian terhadap kerapihan penampilan crew bus (Pengemudi dan pramugara/pramugari) bus Rosalia Indah
- B9 : Penilaian terhadap keramahan dan profesionalisme crew bus (pramugara/pramugari) dalam melayani penumpang dan menangani keluhan atau masalah selama perjalanan
- B10 : Penilaian terhadap profesionalisme Pengemudi dalam mengemudikan bus

- B11 : Penilaian terhadap kebersihan dan kenyamanan bus Rosalia Indah
- B12 : Penilaian terhadap keamanan dan keselamatan selama perjalanan dengan bus rosalia indah
- B13 : Penilaian terhadap kompensasi yang ditawarkan saat bus mengalami masalah/problem (bus mogok, bus terlambat)
- C1 : Penilaian terhadap harga tiket bus Rosalia Indah
- C2 : Penilaian tentang harga tiket yang ditawarkan apakah terjangkau dan sesuai dengan kemampuan responden
- C3 : Penilaian tentang harga tiket yang ditawarkan apakah sebanding dengan kualitas pelayanan dan fasilitas yang diberikan oleh pihak Rosalia Indah
- C4 : Penilaian tentang harga tiket bus Rosalia Indah apakah cukup bersaing jika dibandingkan dengan harga yang diberikan dari perusahaan bus lain
- D1 : Penilaian terhadap pelayanan yang diberikan oleh pihak bus Rosalia Indah
- D2 : Peluang responden akan menjadikan bus Rosalia Indah sebagai Transportasi utama untuk berpergian ke luar kota
- D3 : Peluang responden akan merekomendasikan bus Rosalia Indah kepada orang terdekatnya
- D4 : Peluang responden akan memberikan informasi yang positif tentang bus Rosalia Indah
- D5 : Tingkat kepuasan responden dengan keseluruhan layanan yang diberikan oleh bus Rosalia Indah

4.2 Hasil *Pre-processing Data*

Langkah yang selanjutnya dilakukan yaitu melakukan pengecekan terhadap dataset pengguna bus rosalia indah yaitu melakukan tahapan *pre-processing data*.

4.2.1 Uji Validitas dan Reliabilitas

Kuesioner yang telah dijawab oleh responden kemudian diuji validitas dan reliabilitasnya untuk memastikan bahwa instrumen kuesioner yang digunakan benar-

benar mengukur tentang aspek kepuasan pengguna dan konsistensi hasil yang diperoleh dari instrumen penelitian.

A. Uji Validitas

Uji validitas dilakukan menggunakan r tabel. Nilai r tabel untuk sampel 312 dengan tingkat signifikan 1% menunjukkan r tabel sebesar 0.148. Uji validitas dilakukan untuk mengetahui apakah hasil jawaban responden kuesioner valid atau tidak. Kriteria dikatakan valid apabila nilai r hitung > nilai r tabel. Adapun hasil dari pengujian validitas dari penelitian ini dapat dilihat pada tabel IV.2 berikut:

Tabel IV. 2 Uji Validitas

Variabel	α	r tabel	r hitung	Keterangan
A01	0.01	0.148	0.824	Valid
A02	0.01	0.148	0.868	Valid
A03	0.01	0.148	0.812	Valid
B01	0.01	0.148	0.803	Valid
B02	0.01	0.148	0.813	Valid
B03	0.01	0.148	0.747	Valid
B04	0.01	0.148	0.826	Valid
B05	0.01	0.148	0.744	Valid
B06	0.01	0.148	0.726	Valid
B07	0.01	0.148	0.816	Valid
B08	0.01	0.148	0.799	Valid
B09	0.01	0.148	0.819	Valid
B10	0.01	0.148	0.827	Valid
B11	0.01	0.148	0.836	Valid
B12	0.01	0.148	0.845	Valid
B13	0.01	0.148	0.749	Valid
C01	0.01	0.148	0.897	Valid
C02	0.01	0.148	0.884	Valid
C03	0.01	0.148	0.783	Valid
C04	0.01	0.148	0.840	Valid
D01	0.01	0.148	0.882	Valid
D02	0.01	0.148	0.846	Valid
D03	0.01	0.148	0.903	Valid
D04	0.01	0.148	0.898	Valid
D05	0.01	0.148	0.850	Valid

Sumber: Bahan Penelitian

Pada Tabel IV.2 dapat dilihat bahwa semua pernyataan pada semua indikator memperoleh nilai r hitung yang lebih besar ($>$) dari r tabel sebesar 0.148 sehingga uji validitas semua item pertanyaan dalam kuesioner dinyatakan valid.

B. Uji Reliabilitas

Setelah melakukan uji validitas langkah selanjutnya adalah melakukan uji reliabilitas. Uji reliabilitas dilakukan untuk mengetahui konsistensi hasil yang diperoleh dari instrumen penelitian. Pengujian dilakukan untuk mengetahui nilai dari *Cronbach's Alpha*, Suatu variabel data dikatakan reliabel apabila nilai dari *Cronbach's Alpha* lebih dari ($>$) 0,6. Pengujian dilakukan terhadap variabel data demografi pengguna, kualitas pelayanan, harga, dan tingkat kepuasan secara keseluruhan. Hasil uji reliabilitas pada penelitian ini dapat dilihat pada Tabel IV.3 berikut:

Tabel IV. 3 Uji Reliabilitas

Variabel	Cronbach's Alpha	Standar Alpha	Keterangan
A	0.6	0.779	Reliabel
B	0.6	0.951	Reliabel
C	0.6	0.873	Reliabel
D	0.6	0.923	Reliabel

Sumber: Bahan Penelitian

Hasil uji reliabilitas menunjukkan bahwa untuk masing-masing variabel r alpha memiliki nilai lebih besar dari nilai *Cronbach's Alpha* 0.6, Hal ini menunjukkan hasil uji reliabilitas dari keseluruhan variabel dalam penelitian ini adalah reliabel.

4.2.2 Data Cleaning

Data yang sudah dikumpulkan dalam tahap ini akan dilakukan *cleaning* data yang meliputi *missing values* dan duplikat data. Hasil dari pengecekan *data cleaning* dapat dilihat pada tabel IV.4.

Tabel IV. 4 Cleaning Data

No.	Cleaning	Total
1	<i>Missing value</i>	0
2	Duplikat data	0

Sumber: Hasil Penelitian

Tahapan Selanjutnya akan dilakukan perubahan sebanyak 3 kategori dari kategorial menjadi angka pada beberapa data yang sudah *dicleaning* tadi. Kategori yang dirubah antara lain :

- a. Umur
- b. Jenis Kelamin
- c. Frekuensi menggunakan bus

Perubahan yang pertama yang pada kategori umur, umur responden pada penelitian ini diubah dari *character* menjadi *numerik* seperti pada tabel IV.5, dan tabel kategori umur diubah menjadi kategori A1 seperti pada tabel IV.9.

Tabel IV. 5 Kategori Umur

No.	Keterangan	Value
1	15-20 tahun	1
2	21-30 tahun	2
3	31-40 tahun	3
4	41-50 tahun	4
5	Lebih dari 50 tahun	5

Sumber: Hasil Penelitian

Kategori selanjutnya yang akan diubah yaitu Frekuensi responden menggunakan bus dalam sebulan. Pada kategori ini akan dilakukan perubahan jawaban kategori dari *character* menjadi *numerik* seperti pada tabel IV.6, dan tabel kategori diubah menjadi A2 seperti pada tabel IV.9.

Tabel IV. 6 Frekuensi menggunakan bus selama sebulan

No.	Keterangan	Value
1	Jarang menggunakan	1
2	1-2 kali	2
3	3-5 kali	3
4	Lebih dari 5 kali	4

Sumber: Hasil Penelitian

Kategori terakhir yang akan diubah yaitu Jenis Kelamin responden menggunakan bus dalam sebulan. Pada kategori tersebut dilakukan perubahan jawaban kategori dari *character* menjadi *numerik*. Hasil dari perubahan kategori tersebut bisa dilihat pada tabel IV.7, dan tabel kategori diubah menjadi A3 seperti pada tabel IV.9.

Tabel IV. 7 Kategori Jenis Kelamin

No.	Keterangan	Value
1	Laki-laki	1
2	Perempuan	2

Sumber: Hasil Penelitian

4.2.3 Pelabelan data

Menurut hasil pada penelitian yang dilakukan kepada 312 responden dengan teknik mengumpulkan data melalui kuesioner. Tahapan selanjutnya yaitu pelabelan data. Untuk pelabelan datanya di gunakan skala likert dengan pilihan jawaban setuju dan tidak setuju dengan beragam skala nilai. Semua pertanyaan berjumlah 25 dengan nilai minimum 1 dan maksimum 5, serta jumlah kelas 2 (K), maka perhitungan *range* atau rata-rata skor dilakukan dengan rumus berikut:

$$1 \frac{(25 \times 5) - (25 \times 1)}{2} = 50$$

Sumber: Bahan Penelitian

Gambar IV. 2 Penghitungan Rentang Skor Rata-rata

Dengan memakai 25 kategori, maka skala sebesar 125 dikurangi 25 dibagi 2 akan menghasilkan 50. Berdasarkan kategori jawaban responden maka skala pelabelan data bisa digambarkan pada tabel IV.8.

Tabel IV. 8 Interval Nilai

No.	Interval	Klasifikasi
1	25 – 74	Tidak Puas
2	75 – 125	Puas

Sumber: Hasil Penelitian

Kemudian dilakukan penjumlahan sejumlah kategori pada dataset yang kemudian penjumlahan dari kategori tersebut akan dirubah menjadi label klasifikasi puas dan tidak puas berdasarkan tabel IV.8 seperti dataset seperti pada tabel IV.9.

Tabel IV. 9 Hasil Cleaning Data

Nama	Pekerjaan	Domisili	Dari mana Anda mengikuti tentang layanan bus Ros Indah?	A1	A2	A3	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12	B13	C1	C2	C3	C4	D1	D2	D3	D4	D5	Total	Klasifikasi	
				▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼
Valent Arit Nanyanto	Mahasiswa/Pelajar	Jabodetabek	Keluarga/Teman	2	1	2	4	5	5	5	5	4	4	5	4	5	5	4	5	4	4	4	4	5	5	4	5	5	105	Puas	
AHMAD SYAHRI WIJAYANTO	Mahasiswa/Pelajar	Jabodetabek	Keluarga/Teman	1	1	2	1	2	2	2	1	3	2	1	2	1	2	1	2	1	2	2	1	2	5	2	1	2	44	Tidak Puas	
Alifan Rochmatul Abdani	Karyawan	Jabodetabek	Langganan sendiri	1	1	2	5	4	5	4	3	5	3	5	4	5	5	3	3	4	3	4	5	5	4	5	5	4	97	Puas	
Trianov Wiloyandaru	Freelance	Jabodetabek	Keluarga/Teman	1	1	2	4	4	3	4	2	4	3	3	3	3	4	4	5	3	3	1	3	4	3	3	3	4	77	Puas	
Orion Bahhar Khan	Belum bekerja	Jabodetabek	Media Sosial	1	1	2	4	5	4	4	4	4	4	5	4	5	4	4	3	4	4	4	3	4	3	3	3	4	90	Puas	
Thariq Sajid Imanullah	Mahasiswa/Pelajar	Lampung	Buana	1	1	1	5	5	5	5	5	4	5	5	5	5	5	5	3	1	5	2	5	2	5	5	5	5	100	Puas	
Ringga Raka	Karyawan	Jabodetabek	Keluarga/Teman	2	1	2	4	4	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	113	Puas	
Aris Jaka Purnama	Karyawan	Surabaya	Media Sosial	1	1	2	3	3	3	4	4	4	5	5	3	5	5	3	5	5	5	5	5	4	4	5	4	4	99	Puas	
yedii	Wiraswasta	Jabodetabek	Keluarga/Teman	1	1	2	5	3	3	4	5	5	2	5	5	4	5	5	3	3	4	5	4	5	4	5	5	4	97	Puas	
Alfino Rifat Hasan	Mahasiswa/Pelajar	Surabaya	Media Sosial	2	1	1	5	5	4	5	5	5	4	5	5	5	5	5	5	5	5	5	5	4	5	4	5	4	5	110	Puas
Adeyudha Saputra	Mahasiswa/Pelajar	Jabodetabek	Media Sosial	3	1	1	5	5	4	5	4	4	5	5	5	4	5	4	4	4	4	5	4	5	5	5	5	5	5	106	Puas
Fero	Mahasiswa/Pelajar	Lampung	Keluarga/Teman	1	1	2	3	4	4	3	4	2	4	4	5	3	5	3	3	2	2	4	4	3	4	4	3	3	80	Puas	
Gabriel Adi	Mahasiswa/Pelajar	Jabodetabek	Keluarga/Teman	3	1	2	4	5	3	5	3	3	4	5	4	5	4	3	5	5	4	3	4	4	5	5	5	4	97	Puas	
Sjabon Sofyan	Karyawan	Jabodetabek	Keluarga/Teman	2	1	2	1	5	5	4	5	4	4	4	2	5	3	1	5	5	5	3	4	3	4	3	4	3	4	88	Puas
Muhammad Ghaisan Radhi	Mahasiswa/Pelajar	Jogjakarta	Keluarga/Teman	3	1	1	3	5	3	4	4	5	4	4	5	4	5	4	3	3	4	5	4	4	3	5	4	4	94	Puas	
Gilang Abadi	Karyawan	Jabodetabek	Keluarga/Teman	1	1	2	4	5	3	4	4	2	4	5	5	4	5	5	3	3	4	4	3	4	5	5	5	5	95	Puas	
M Rizki Akbar	Mahasiswa/Pelajar	Jember	Keluarga/Teman	1	1	1	5	4	4	5	3	5	5	3	4	3	5	2	3	3	5	4	4	4	3	4	3	5	89	Puas	
Marvin Adhyas Anggoro	Mahasiswa/Pelajar	Jabodetabek	Media Sosial	1	1	1	4	4	4	5	4	5	5	5	5	5	5	4	5	4	4	5	5	4	5	5	5	5	105	Puas	
Elang Rakai Prayoga	Mahasiswa/Pelajar	Jabodetabek	Media Sosial	2	1	2	5	5	5	5	5	3	4	5	5	5	5	5	3	3	5	4	5	5	5	5	5	5	5	107	Puas
Arif Adhityanto	Karyawan	Jabodetabek	agen dekat rumah	2	1	2	5	4	4	5	4	5	4	5	5	4	5	4	5	4	5	4	5	4	5	4	4	5	5	105	Puas
Ian	Mahasiswa/Pelajar	Jabodetabek	Media Sosial	1	1	2	4	5	5	5	4	4	5	5	5	5	5	5	3	4	4	4	5	5	5	5	5	5	5	107	Puas
Muhamad Riza	Mahasiswa/Pelajar	Jabodetabek	Media Sosial	1	1	1	4	4	4	3	3	5	5	4	4	3	4	3	1	3	3	4	3	4	3	4	4	4	4	82	Puas
Muhammad Araysa Swatara	Wiraswasta	Semarang	Media Sosial	2	1	2	5	4	4	5	5	4	4	5	5	5	5	3	4	5	5	3	5	5	5	5	5	4	105	Puas	
Alif Tion	Mahasiswa/Pelajar	Jabodetabek	Media Sosial	1	1	1	4	5	3	4	4	4	4	4	4	4	4	4	4	4	4	5	4	2	4	4	4	4	90	Puas	
Ferdinandus	Mahasiswa/Pelajar	Surabaya	Media Sosial	1	1	2	5	3	3	5	5	5	5	5	5	5	5	5	3	5	5	5	5	5	5	5	5	5	5	114	Puas
Fadri Kiyi Itham Sunu	Mahasiswa/Pelajar	Jawa Tengah	Media Sosial	1	1	1	5	5	5	4	5	5	4	5	5	5	5	5	4	5	5	4	4	4	3	5	5	5	5	105	Puas
Vittorio Ratama	Karyawan	Jabodetabek	Media Sosial	2	1	2	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	115	Puas
Yovie Arhab Yudistira	Mahasiswa/Pelajar	Jogjakarta	Nilan	2	1	1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	114	Puas
Sultan MDS	Mahasiswa/Pelajar	Jabodetabek	Keluarga/Teman	1	1	1	4	5	4	4	4	5	5	4	4	4	4	4	3	4	4	4	4	4	4	4	4	4	4	95	Puas
Filtri Ali	Mahasiswa/Pelajar	Kediri	Melihat Langsung	2	1	1	4	4	3	4	3	4	4	5	4	4	4	4	3	3	4	4	3	4	4	4	4	4	4	88	Puas
Benedictus	Wiraswasta	Palembang	Media Sosial	2	1	2	4	5	5	5	4	4	4	5	5	5	5	4	4	5	4	4	5	5	5	5	4	4	4	104	Puas
Putra	Karyawan	Jabodetabek	Media Sosial	1	1	2	3	4	4	3	3	3	3	3	3	3	3	3	2	3	2	2	5	3	3	3	3	3	3	69	Tidak Puas

Sumber: Hasil Penelitian

Dari hasil pelabelan data pada tabel IV.9 didapatkan sebanyak 263 responden dinyatakan Puas terhadap pelayanan yang diberikan perusahaan otobus (PO) rosalia indah dan 49 responden dinyatakan Tidak Puas terhadap pelayanan yang diberikan

perusahaan otobus (PO) rosalia indah dari total 312 responden seperti pada tabel IV.10.

Tabel IV. 10 Kepuasan Pengguna

No.	Keterangan	Total
1	Puas	263
2	Tidak Puas	49
	Total	312

Sumber: Hasil Penelitian

4.2.4 *Splitting Data*

Tahapan berikutnya adalah melakukan *splitting data*. Data yang telah diberikan label kemudian akan dilakukan proses pembagian data dimana total 312 data pada dataset akan dikelompokkan ke dalam dua klasifikasi, yaitu *data training* dan *data testing*. pada tahap ini akan dilakukan beberapa pengujian memanfaatkan operator *split data* dengan menggunakan *naive bayes* dan *k-neares neighbors* memakai bantuan *tools rapidminer* seperti pada gambar IV.3.

Gambar IV. 3 Rasio Split Data



Sumber: Hasil Penelitian

Dalam tahapan akan ada total empat pengujian *split data* yang dilakukan:

- Percobaan pertama akan melibatkan *data training* sebesar 90% dan 10% *data testing* sebesar 10%.
- percobaan kedua menggunakan *data training* sebesar 80% dan *data testing* sebesar 20%.

- c. percobaan ketiga melibatkan *data training* sebesar 70% dan *data testing* sebesar 30%.
- d. Pengujian terakhir akan menggunakan *data training* sebesar 60% dan *data testing* sebesar 40%.

4.3 Hasil Pengujian Algoritma

Setelah melakukan proses *data preprocessing*, selanjutnya adalah melakukan pengujian dua algoritma *machine learning* seperti *naive bayes* dan *k-neares neighbors* dalam menentukan nilai *accuracy* dan AUC.

Pada tahap ini dataset akan dilakukan 4 kali pengujian menggunakan *split validation*, yaitu membagi data menjadi dua klasifikasi, yakni menjadi data latih (*Training*) dan data uji (*Testing*). Tujuan dilakukannya 4 kali pengujian adalah supaya dapat mengetahui algoritma yang bisa memprediksi nilai *accuracy* dan AUC yang paling tinggi.

A. Algoritma *Naïve Bayes*

Pengujian yang pertama adalah pengujian menggunakan algoritma *naïve bayes* dengan *RapidMiner*.

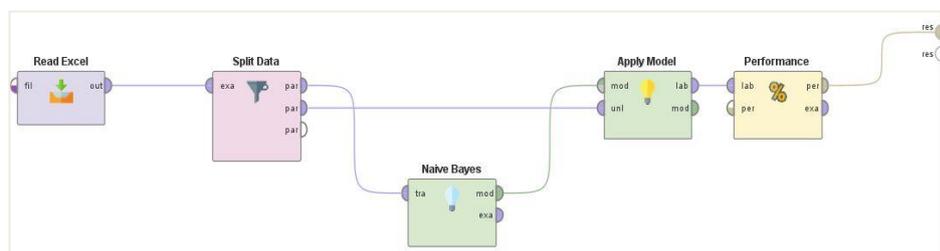
Untuk membuat pemodelan *naïve bayes* menggunakan *rapidminer* bisa dengan melakukan *drag & drop operator* yang diperlukan dan menyambungkan koneksi antar operator. Langkah Pertama yang dilakukan adalah menambahkan operator *Read Excel* untuk memilih dataset yang telah melalui *pr-eprocessing*. Selanjutnya ubat *dataset* tersebut pada bagian klasifikasi (Puas/Tidak Puas) menjadi *label*.

Operator yang selanjutnya yaitu *split data*, operator ini akan dipakai dalam membagi *dataset* kedalam dua klasifikasi, yakni data latih (*training*) dan data uji (*testing*) dengan rasio 0.9-0.1, 0.8-0.2, 0.7-0.3, 0.6-0.4.

Selanjutnya yaitu menambahkan operator *Naïve bayes* yang berfungsi melakukan pengkategorian data berdasarkan kategori tertentu.

Yang selanjutnya adalah menambahkan operator *Apply Model* untuk pengambilan keputusan berdasarkan data yang ada.

Yang terakhir yaitu operator *Performance* digunakan sebagai pengukuran dari hasil pengujian yang dibuat. Untuk model pengujian *naïve bayes* bisa dilihat pada Gambar IV. 4.



Sumber: Hasil Penelitian

Gambar IV. 4 Model Pengujian Naïve Bayes

Keterangan:

1. Read Excel : digunakan untuk membaca data yang disimpan dalam format file Excel (.xlsx atau .xls) ke dalam lingkungan pemrograman atau perangkat lunak analisis data.
2. Split Data : digunakan dalam mengkategorikan kedalam dua kelas, yakni data latih (*training*) dan data uji (*testing*).
3. Naïve Bayes : digunakan untuk mengklasifikasikan data ke dalam kategori tertentu.
4. Apply Model : untuk membuat prediksi atau mengambil keputusan berdasarkan data yang baru atau data yang tidak pernah terlihat sebelumnya.
5. Performance: digunakan mengukur akurasi hasil pemodelan yang dibuat.

Dari dataset yang sudah diberi label kemudian akan dilakukan percobaan menggunakan operator *split data*, dimana sejumlah 312 data yang didapatkan dari pengumpulan data akan diubah kedalam *data training* dan *data testing* memakai *RapidMiner* menggunakan operator *split data* dengan rasio tertentu mencari nilai *accuracy* dari *Confusion Matrix* dengan menerapkan model *naive bayes* dengan empat kali percobaan.

a) Pengujian pertama memakai 90% *data training* dan 10% *data testing*

Tabel IV. 11 Pengujian Confusion Matrix Naïve Bayes Rasio 90% dan 10%

	true Puas	true Tidak Puas	class precision
pred. Puas	28	0	100.00%
pred. Tidak Puas	1	2	66.67%
class recall	96.55%	100.00%	

Sumber: Hasil Penelitian

Berdasarkan tabel IV.11 menunjukkan bahwa, diketahui dari 312 data pengguna bus rosalia indah, 90% dirubah menjadi *data training* dan 10% untuk *data testing* dengan jumlah 31 data yang dimanfaatkan untuk pengujian. Hasil pengujian *confusion matrix* pada tabel IV.11 terdapat 28 pengguna diperkirakan puas maka hasilnya pun sama seperti perkiraan, yakni puas. 0 pengguna diperkirakan tidak puas tetapi hasilnya tidak seperti yang diperkirakan, yaitu puas, sedangkan 1 pengguna yang diprediksikan tidak puas tetapi hasilnya tidak sesuai prediksi, yaitu puas. Dan 2 pengguna diprediksikan tidak puas maka hasilnya sama seperti yang diprediksikan yakni tidak puas. Percobaan pertama untuk mencari nilai akurasi dengan penerapan algoritma *Naive Bayes* yaitu sebesar 96.77%, dan untuk mencari nilai akurasi dengan cara berikut:

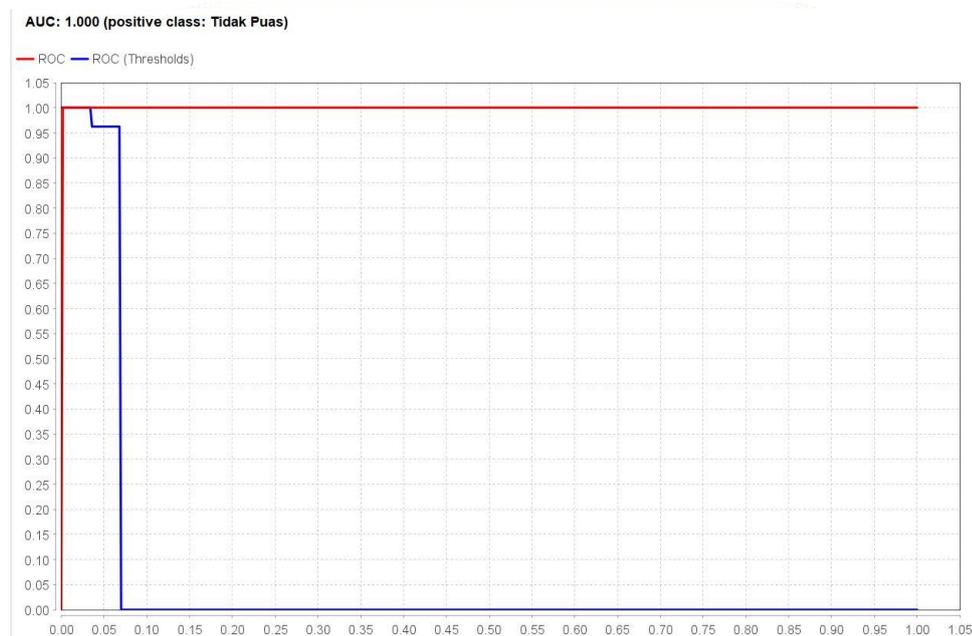
$$TP = 28$$

$$FP = 1$$

$$TN = 2$$

$$FN = 0$$

$$\begin{aligned} \text{Rumus} &= \frac{(TP+TN)}{(TP+TN+FP+FN)} \\ &= \frac{(28+2)}{(28+2+1+0)} \\ &= \frac{30}{31} \\ &= 0.9677 \text{ (96.77\%)} \end{aligned}$$



Sumber: Bahan Penelitian

Gambar IV. 5 Kurva ROC Pengujian Pertama Naïve Bayes

Pada gambar IV.5 terdapat Kurva ROC *naïve bayes* yang menggambarkan nilai AUC (*Area Under Curve*) dari *confusion matrix* pada Tabel IV.11 sebesar 1.000 dimana hasil tersebut termasuk dalam kategori *Excellent Classification*.

b) Pengujian kedua menggunakan 80% *data training* dan 20% *data testing*

Tabel IV. 12 Pengujian Confusion Matrix Naïve Bayes Rasio 80% dan 20%

accuracy: 98.39%

	true Puas	true Tidak Puas	class precision
pred. Puas	58	0	100.00%
pred. Tidak Puas	1	3	75.00%
class recall	98.31%	100.00%	

Sumber: Hasil Penelitian

Berdasarkan tabel IV.12 menunjukkan bahwa, diketahui dari 312 data pengguna bus rosalia indah, 80% dirubah menjadi data uji (*training*) dan 20% sebagai data uji (*testing*) dengan jumlah 62 data yang dipakai untuk pengujian. Hasil pengujian *confusion matrix* pada tabel 1V.12 terdapat 58 pengguna diprediksikan puas maka hasilnya pun seperti yang sudah diprediksikan, yakni puas. 1 pengguna diperkirakan tidak puas tetapi hasilnya tidak seperti yang diperkirakan, yaitu puas, sedangkan 0 pengguna yang diperkirakan tidak puas tetapi hasilnya tidak sesuai perkiraan, yaitu puas. Ada 3 pengguna yang diperkirakan tidak puas maka hasilnya sama seperti yang perkiraan yakni tidak puas. Percobaan kedua untuk mencari nilai akurasi dengan penerapan algoritma *Naive Bayes* yaitu sebesar 98.39%, dan dapat dihitung untuk menentukan nilai akurasi dengan cara berikut:

$$TP = 58$$

$$FP = 1$$

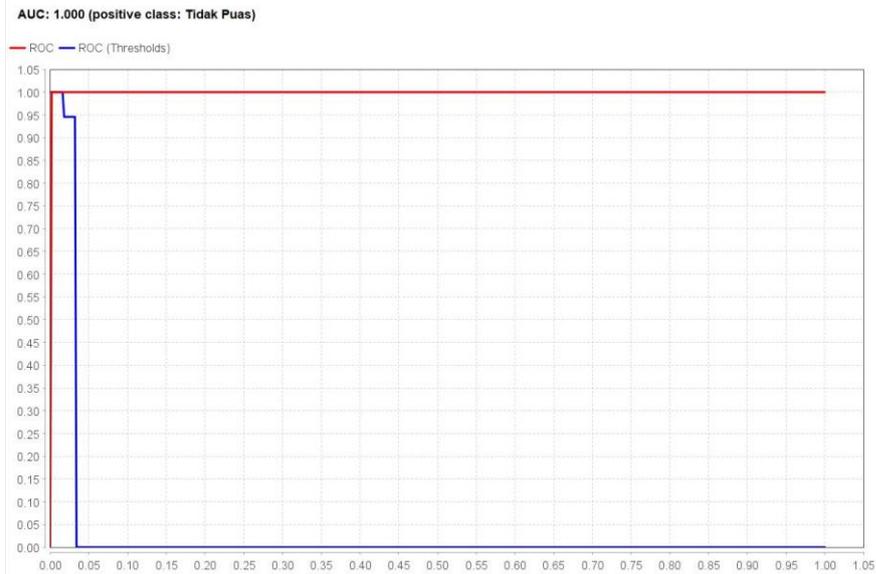
$$TN = 3$$

$$FN = 0$$

$$\begin{aligned} \text{Rumus} &= \frac{(TP+TN)}{(TP+TN+FP+FN)} \\ &= \frac{(58+3)}{(58+3+1+0)} \end{aligned}$$

$$= \frac{61}{62}$$

$$= 0.9839 \text{ (98.39\%)}$$



Sumber: Hasil Penelitian

Gambar IV. 6 Kurva ROC Pengujian Kedua Naive Bayes

Gambar IV.6 menggambarkan Kurva ROC yang mempunyai nilai *AUC* (*Area Under Curve*) dari *confusion matrix* pada Tabel IV.12 yakni sebesar 1.000 dimana hasil tersebut termasuk dalam kategori *Excellent Classification*.

- c) Pengujian menggunakan 70% *data training* dan 30% *data testing*

Tabel IV. 13 Pengujian Confusion Matrix Naïve Bayes Rasio 70% dan 30%

accuracy: 96.81%			
	true Puas	true Tidak Puas	class precision
pred. Puas	80	0	100.00%
pred. Tidak Puas	3	11	78.57%
class recall	96.39%	100.00%	

Sumber: Bahan Penelitian

Berdasarkan tabel IV.13 menunjukkan bahwa, dari 312 data pengguna bus rosalia indah, 70% diubah menjadi data uji (*training*) dan 30% untuk data uji (*testing*) dengan jumlah 94 data yang digunakan untuk pengujian. Hasil pengujian *confusion matrix* pada tabel IV.13 terdapat 80 pengguna diprediksikan puas maka hasilnya pun seperti yang sudah diprediksikan, yakni puas. 3 pengguna diperkirakan tidak puas tetapi hasilnya tidak seperti yang diperkirakan, yaitu puas, sedangkan 0 pengguna yang diperkirakan tidak puas tetapi hasilnya tidak sesuai perkiraan, yaitu puas. Ada 11 pengguna yang diprediksikan tidak puas maka hasilnya sama seperti yang diprediksikan adalah tidak puas. Percobaan ketiga untuk mencari nilai akurasi dengan penerapan algoritma *Naive Bayes* yaitu sebesar 96.81%, dan untuk menghitung akurasinya dengan cara berikut:

$$TP = 80$$

$$FP = 3$$

$$TN = 11$$

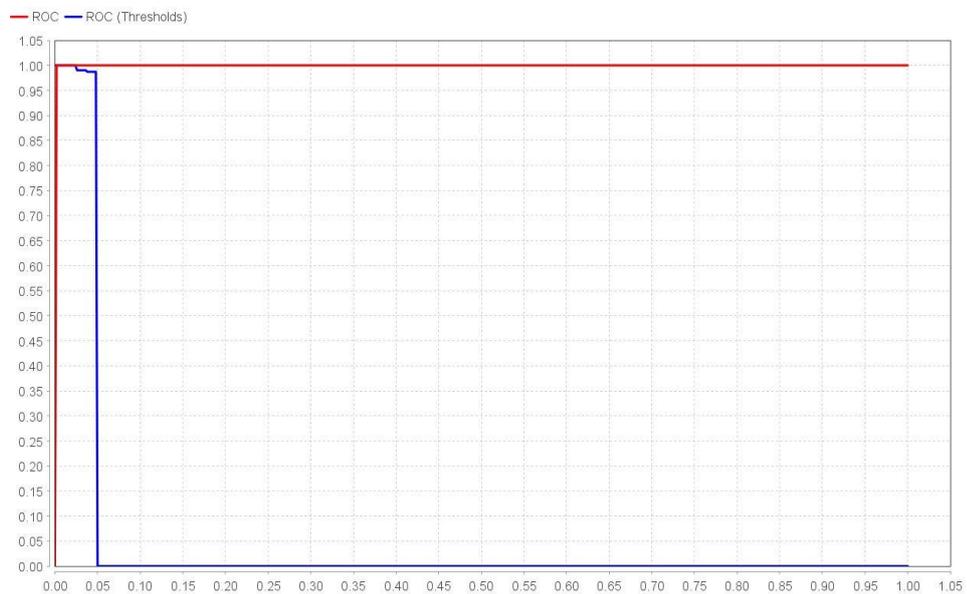
$$FN = 0$$

$$\begin{aligned} \text{Rumus} &= \frac{(TP+TN)}{(TP+TN+FP+FN)} \\ &= \frac{(80+11)}{(80+11+3+0)} \end{aligned}$$

$$= \frac{91}{94}$$

$$= 0.9681 \text{ (96.81\%)}$$

AUC: 1.000 (positive class: Tidak Puas)



Sumber: Bahan Penelitian

Gambar IV. 7 Kurva ROC Pengujian Ketiga Naïve Bayes

Gambar IV.7 mengilustrasikan Kurva ROC yang mempunyai nilai *AUC* (*Area Under Curve*) dari *confusion matrix* pada Tabel IV.13 yakni sebesar 1.000 dimana hasil tersebut termasuk dalam kategori *Excellent Classification*.

d) Pengujian menggunakan 60% *data training* dan 40% *data testing*

Tabel IV. 14 Pengujian Confusion Matrix Naïve Bayes Rasio 60% dan 40%

accuracy: 97.60%			
	true Puas	true Tidak Puas	class precision
pred. Puas	104	0	100.00%
pred. Tidak Puas	3	18	85.71%
class recall	97.20%	100.00%	

Sumber: Bahan Penelitian

Berdasarkan tabel IV.14 memperlihatkan bahwa, sebanyak 312 data pengguna bus rosalia indah, 60% akan dijadikan menjadi data uji (*training*) dan 40% dijadikan sebagai data uji (*testing*) dengan jumlah 125 data yang dipakai dalam pengujian. Hasil pengujian *confusion matrix* pada tabel IV.14 terdapat 103 pengguna diprediksikan puas maka hasilnya pun seperti yang sudah diprediksikan, yakni puas. 3 pengguna diperkirakan tidak puas tetapi hasilnya tidak seperti yang diperkirakan, yaitu puas, sedangkan 0 pengguna yang diperkirakan tidak puas tetapi hasilnya tidak sesuai perkiraan, yaitu puas. Ada 18 pengguna yang diprediksikan tidak puas maka hasilnya sama seperti yang diprediksikan adalah tidak puas. Percobaan terakhir dalam mencari nilai akurasi dengan penerapan algoritma *naive bayes* yaitu sebesar 97.60%, dan bisa dihitung untuk menemukan nilai akurasi dengan cara berikut:

$$TP = 104$$

$$FP = 3$$

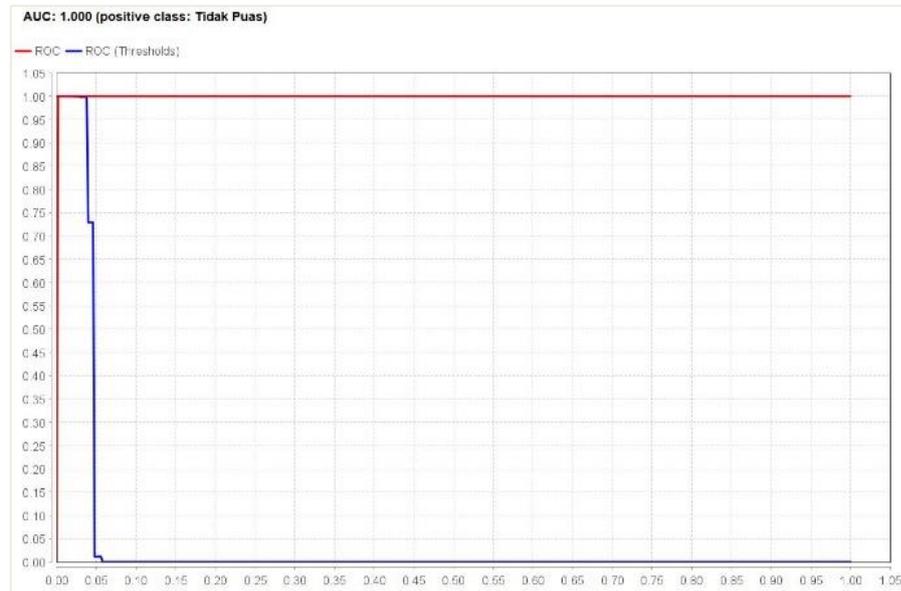
$$TN = 18$$

$$FN = 0$$

$$\begin{aligned} \text{Rumus} &= \frac{(TP+TN)}{(TP+TN+FP+FN)} \\ &= \frac{(104+18)}{(104+18+3+0)} \end{aligned}$$

$$= \frac{122}{125}$$

$$= 0.976 \text{ (97.60\%)}$$



Sumber: Hasil Penelitian

Gambar IV. 8 Kurva ROC Pengujian Keempat Naïve Bayes

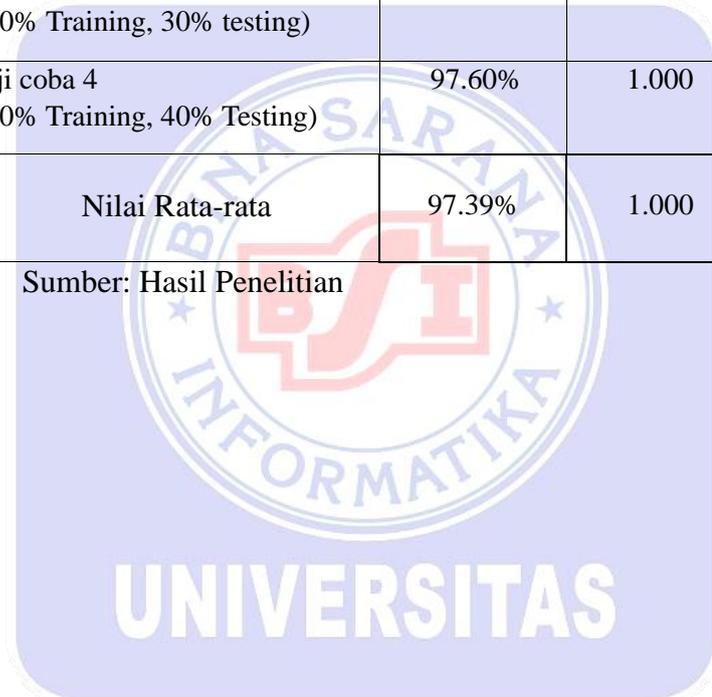
Pada gambar IV.8 terdapat grafik ROC dari percobaan *confusion matrix* pada Tabel IV.14 dengan nilai *AUC (Area Under Curve)* sejumlah 1.000 dan termasuk dalam kategori *Excellent Classification*.

Hasil dari pengujian 4 *split data* menggunakan algoritma *naïve bayes* dapat dilihat di tabel IV.15 berikut.

Tabel IV. 15 Total Keseluruhan Pengujian Naive Bayes

Percobaan <i>Split Data</i>	<i>Accuracy</i>	AUC
Uji coba 1 (90% Training, 10% testing)	96.77%	1.000
Uji coba 2 (80% Training, 20% testing)	98.39%	1.000
Uji coba 3 (70% Training, 30% testing)	96.81%	1.000
Uji coba 4 (60% Training, 40% Testing)	97.60%	1.000
Nilai Rata-rata	97.39%	1.000

Sumber: Hasil Penelitian



B. Pengujian *K-Neares Neighbors*

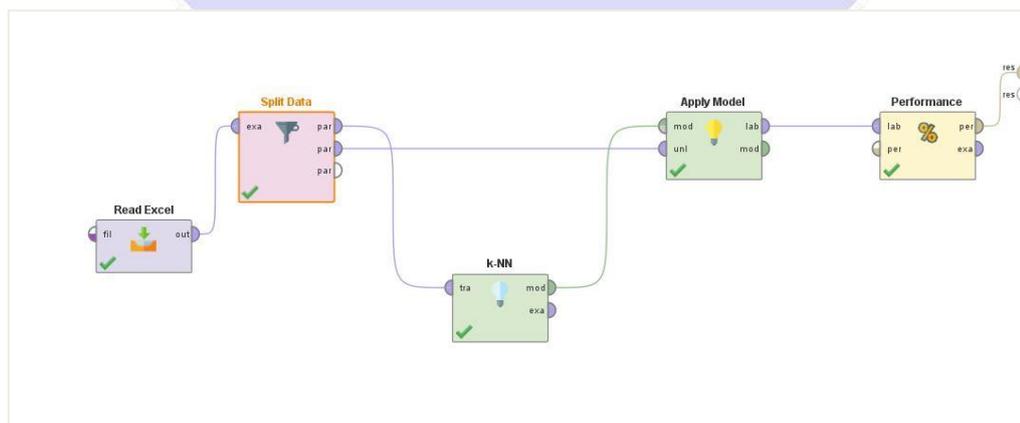
Percobaan berikutnya adalah pengujian menggunakan algoritma *k-nearest neighbors*. Untuk membuat pemodelan *k-nn* memakai *rapidminer* bisa dengan melakukan *drag & drop operator* yang diperlukan untuk menyambunngkan koneksi antar operator. Langkah pertama yang dilakukan adalah menambahkan operator *Read Excel* untuk memilih dataset yang telah melalui *pre-eprocessing*. Selanjutnya mengubah *dataset* tersebut pada bagian klasifikasi (Puas/Tidak Puas) menjadi *label*.

Operator yang selanjutnya yaitu *split data*, operator ini akan digunakan dalam membagi dataset kedalam dua kelas, yakni data latih (*training*) dan data uji (*testing*) dengan rasio 0.9-0.1, 0.8-0.2, 0.7-0.3, 0.6-0.4.

Selanjutnya yaitu menambahkan operator *K-nn* yang berfungsi untuk menentukan kategori dari data yang ada.

Selanjutnya adalah menambahkan operator *Apply Model* untuk pengambilan keputusan berdasarkan data yang ada.

Operator yang terakhir yaitu *Performance* digunakan sebagai pengukuran dari hasil pengujian yang dibuat. Untuk model pengujian *k-nn* bisa dilihat pada gambar IV.9.



Sumber: Hasil Penelitian

Gambar IV. 9 Model Pengujian K-Nearest Neighbors

Keterangan:

1. Read Excel : digunakan untuk membaca data yang disimpan dalam format fileExcel (.xlsx atau .xls) ke dalam lingkungan pemrograman atau perangkat lunak analisis data.
2. Split Data : digunakan dalam mengkategorikan *dataset* kedalam dua kelas, yakni data latih (*training*) dan data uji (*testing*).
3. K-Nearest Neighbors : digunakan menentukan kelas atau kategori dari data input berdasarkan mayoritas kelas dari k tetangga terdekatnya.
4. Apply Model : untuk membuat prediksi atau mengambil keputusan berdasarkan data yang baru atau data yang belum pernah dilihat sebelumnya.
5. Performance: digunakan untuk menghitung keakuratan hasil pemodelan yang dibuat.

Tahap selanjutnya adalah pengujian menggunakan algoritma *k-neares neighbors*. Data yang sudah diberikan label kemudian akan dilakukan percobaan menggunakan operator *split data*, dimana sejumlah 312 data yang didapatkan dari pengumpulan data akan dijadikan kedalam *data training* dan *data testing* memanfaatkan *RapidMiner* menggunakan operator *split data* pada rasio tertentu dalam mencari nilai *accuracy* dari *Confusion Matrix* dengan menerapkan model *K-nn* dengan empat kali percobaan.

a) Pengujian menggunakan 90% *data training* dan 10% *data testing*

Tabel IV. 16 Pengujian Confusion Matrix K-nn Rasio 90% dan 10%

accuracy: 96.77%			
	true Puas	true Tidak Puas	class precision
pred. Puas	28	0	100.00%
pred. Tidak Puas	1	2	66.67%
class recall	96.55%	100.00%	

Sumber: Hasil Penelitian

Berdasarkan tabel IV.16 memperlihatkan dari total 312 data pengguna bus rosalia indah, 90% diubah menjadi data uji (*training*) dan 10% untuk data uji (*testing*) dengan banyaknya 31 data akan dipakai sebagai percobaan. Menentukan nilai $k=\sqrt{n}$ dari *data training* $\sqrt{281} : 17$. Hasil percobaan *confusion matrix* pada tabel IV.16 didapati sebesar 28 pengguna dikategorikan puas maka hasilnya pun seperti yang sudah diprediksi, yakni puas. Ada 1 pengguna yang diperkirakan tidak puas tetapi hasilnya tidak seperti yang diperkirakan, yaitu puas, sedangkan 0 pengguna yang diperkirakan tidak puas tetapi hasilnya tidak sesuai perkiraan, yaitu puas. Ada 2 pengguna yang diprediksikan tidak puas maka hasilnya sama seperti yang diprediksikan adalah tidak puas. Percobaan pertama untuk mencari nilai akurasi memakai algoritma *K-nearest neighbors* yakni sebesar 96.77%, sementara menghitung tingkat akurasi dengan cara berikut:

$$TP = 28$$

$$FP = 1$$

$$TN = 2$$

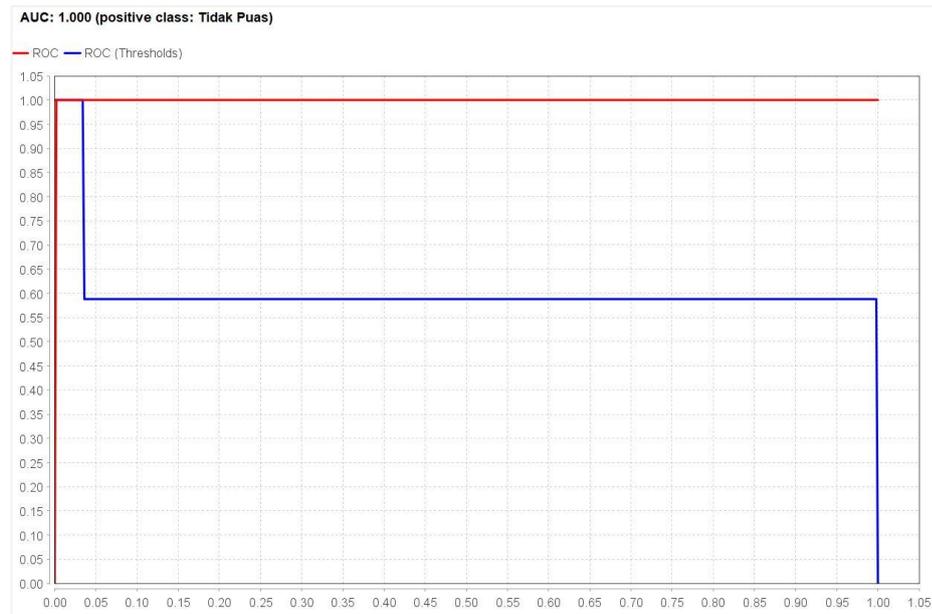
$$FN = 0$$

$$\text{Rumus} = \frac{(TP+TN)}{(TP+TN+FP+FN)}$$

$$= \frac{(28+2)}{(28+2+1+0)}$$

$$= \frac{30}{31}$$

$$= 0.9677 \text{ (96.77\%)}$$



Sumber: Hasil Penelitian

Gambar IV. 10 Kurva ROC Pengujian Pertama Metode *K-nn*

Gambar IV.10 menggambarkan Kurva ROC *K-nn* yang mempunyai nilai *AUC* (*Area Under Curve*) dari *confusion matrix* pada Tabel IV.16 sebesar 1.000 dan termasuk dalam kategori *Excellent Classification*.

b) Pengujian menggunakan 80% *data training* dan 20% *data testing*

Tabel IV. 17 Pengujian Confusion Matrix K-nn Rasio 80% dan 20%

accuracy: 98.39%

	true Puas	true Tidak Puas	class precision
pred. Puas	58	0	100.00%
pred. Tidak Puas	1	3	75.00%
class recall	98.31%	100.00%	

Sumber: Hasil Penelitian

Berdasarkan tabel IV.17 menunjukkan bahwa diketahui dari 312 data kepuasan pengguna bus rosalia indah, 80% akan dipakai untuk data latih (*training*) dan 20% untuk data uji (*testing*) atau sejumlah 62 data akan dipakai dalam pengujian. Mencari nilai $k = \sqrt{n}$ dari *data training* $\sqrt{250} : 16$. Hasil ujicoba pada tabel IV.17 didapati sejumlah 58 pengguna diramalkan puas dan hasilnya selaras pada yang diramalkan yaitu puas. terdapat 1 pengguna diramalkan tidak puas tetapi tidak sesuai dengan ramalannya ternyata puas, sedangkan 0 pengguna yang diprediksi tidak puas tetapi hasilnya adalah puas. Dan ada 3 pengguna yang diperkirakan tidak puas maka hasilnya sama seperti yang perkiraan yakni tidak puas. Tingkat akurasi pada percobaan kedua dengan algoritma *K-nearest neighbors* sebesar 98.39%, dan untuk menghitung tingkat akurasi dengan rumus berikut:

$$TP = 58$$

$$FP = 1$$

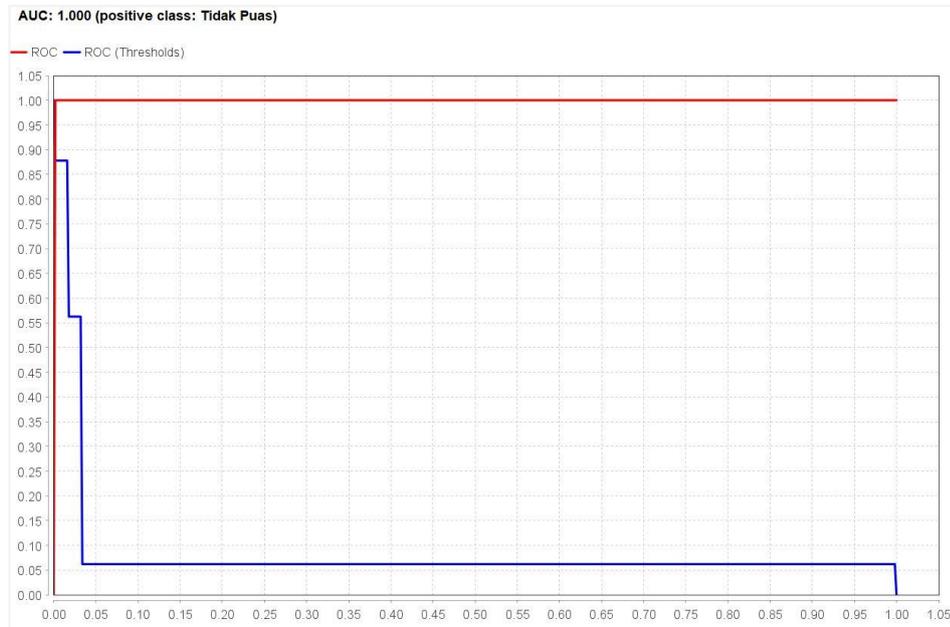
$$TN = 3$$

$$FN = 0$$

$$\begin{aligned} \text{Rumus} &= \frac{(TP+TN)}{(TP+TN+FP+FN)} \\ &= \frac{(58+3)}{(58+3+1+0)} \end{aligned}$$

$$= \frac{61}{62}$$

$$= 0.9839 \text{ (98.39\%)}$$



Sumber: Hasil Penelitian

Gambar IV. 11 Kurva ROC Pengujian Kedua Metode K-nn

Gambar IV.11 memperlihatkan Kurva ROC K-nn yang mempunyai nilai *AUC* (*Area Under Curve*) dari *confusion matrix* pada Tabel IV.17 sebesar 1.000 dan termasuk dalam kategori *Excellent Classification*.

c) Pengujian menggunakan 70% *data training* dan 30% *data testing*

Tabel IV. 18 Pengujian Confusion Matrix K-nn Rasio 70% dan 30%

accuracy: 98.94%

	true Puas	true Tidak Puas	class precision
pred. Puas	82	0	100.00%
pred. Tidak Puas	1	11	91.67%
class recall	98.80%	100.00%	

Sumber: Hasil Penelitian

Berdasarkan tabel IV.18 memperlihatkan bahwa diketahui dari 312 data kepuasan pengguna bus rosalia indah, 70% dikategorikan menjadi data latih (*training*) dan 30% menjadi data uji (*testing*) atau sejumlah 94 data dimanfaatkan pada pengujian. Untuk mencari nilai $k = \sqrt{n}$ dari *data training* $\sqrt{218} : 15$. Hasil ujicoba pada tabel IV.18. Sejumlah 82 pengguna diperkirakan puas dan hasilnya sama seperti yang diperkirakan yakni puas, terdapat 1 pengguna diperkirakan tidak puas tetapi tidak sesuai perkiraan dengan hasilnya yaitu puas, sedangkan 0 pengguna diperkirakan tidak puas tetapi hasilnya tidak sama dengan perkiraan hasilnya yakni puas. dan 11 pengguna yang diperkirakan tidak puas maka hasilnya sama seperti yang perkiraan yakni tidak puas. Tingkat akurasi pada percobaan ketiga menggunakan *k-nearest neighbors* adalah 98.94%, dan bisa dihitung untuk mencari nilai akurasi dengan cara berikut:

$$TP = 82$$

$$FP = 1$$

$$TN = 11$$

$$FN = 0$$

$$\text{Rumus} = \frac{(TP+TN)}{(TP+TN+FP+FN)}$$

$$= \frac{(82+11)}{(82+11+1+0)}$$

$$= \frac{93}{94}$$

$$= 0.9894 \text{ (98.94\%)}$$



Sumber: Hasil Penelitian

Gambar IV. 12 Kurva ROC Pengujian Ketiga Metode K-nn

Gambar IV.12 mengilustrasikan Kurva ROC K-nn yang mempunyai nilai *AUC* (*Area Under Curve*) dari *confusion matrix* pada Tabel IV.18 sebesar 1.000 dan termasuk dalam kategori *Excellent Classification*.

d) Pengujian menggunakan 60% *data training* dan 40% *data testing*

Tabel IV. 19 Pengujian Confusion Matrix K-nn Rasio 60% dan 40%

accuracy: 99.20%

	true Puas	true Tidak Puas	class precision
pred. Puas	106	0	100.00%
pred. Tidak Puas	1	18	94.74%
class recall	99.07%	100.00%	

Sumber: Hasil Penelitian

Berdasarkan tabel IV.19 memperlihatkan bahwa, sebanyak 312 data kepuasan pengguna bus rosalia indah, 60% untuk data latih (*training*) dan 40% untuk data uji (*testing*) dengan jumlah 125 data yang dipakai dalam pengujian. Mendapatkan nilai $k=\sqrt{n}$ dari *data training* $\sqrt{187} : 14$. Hasil pengujian *confusion matrix* pada tabel IV.19 terdapat 106 pengguna diprediksikan puas maka hasilnya pun seperti yang sudah diprediksikan, yakni puas. 1 pengguna diperkirakan tidak puas tetapi hasilnya tidak seperti yang diperkirakan, yaitu puas, sedangkan 0 pengguna yang diperkirakan tidak puas tetapi hasilnya tidak sesuai perkiraan, yaitu puas. Ada 18 pengguna yang diprediksikan tidak puas maka hasilnya sama seperti yang diprediksikan adalah tidak puas. Percobaan terakhir dalam mencari nilai akurasi dengan penerapan algoritma *naive bayes* yaitu sebesar 99.20%, dan bisa dihitung untuk menemukan nilai akurasi dengan cara berikut:

$$TP = 106$$

$$FP = 1$$

$$TN = 18$$

$$FN = 0$$

$$\begin{aligned} \text{Rumus} &= \frac{(TP+TN)}{(TP+TN+FP+FN)} \\ &= \frac{(106+18)}{(106+18+1+0)} \end{aligned}$$

$$= \frac{124}{125}$$

$$= 0.992 \text{ (99.20\%)}$$



Sumber: Hasil Penelitian

Gambar IV. 13 Kurva ROC Pengujian Keempat Metode K-nn

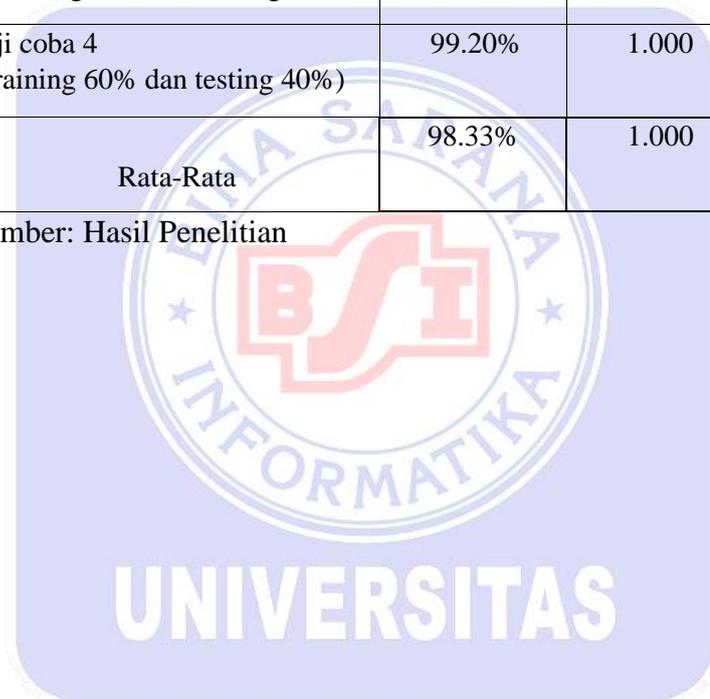
Gambar IV.13 mengilustrasikan Kurva ROC K-nn yang mempunyai nilai *AUC* (*Area Under Curve*) dari *confusion matrix* pada Tabel IV.19 sebesar 1.000 dan termasuk klasifikasi *Excellent Classification*.

Hasil dari pengujian 4 *split data* menggunakan algoritma *K-Nearest Neighbors* bisa dilihat dalam tabel IV.20 berikut:

Tabel IV. 20 Total Keseluruhan Pengujian K-Nearest Neighbors

Percobaan Split Data	Accuracy	AUC
Uji coba 1 (Training 90% dan testing 10%)	96.77%	1.000
Uji coba 2 (training 80% dan Testing 20%)	98.39%	1.000
Uji coba 3 (Training 70% dan testing 30%)	98.94%	1.000
Uji coba 4 (training 60% dan testing 40%)	99.20%	1.000
Rata-Rata	98.33%	1.000

Sumber: Hasil Penelitian



4.4 Hasil Evaluasi

Untuk menganalisa kepuasan pengguna diperlukan metode yang paling akurat. Algoritma *naïve bayes* dan *k-nearest neighbors* dipilih berdasarkan hasil dari penelitian terdahulu menunjukkan apabila *naïve bayes* dan *k-nearest neighbors* mempunyai keahlian yang signifikan dalam mengklasifikasi suatu data.

Hasil pengujian yang sudah dijalankan untuk mengetahui perkiraan hasil klasifikasi kepuasan pengguna rosalia indah, bisa disimpulkan dari hasil pengujian memanfaatkan metode *naïve bayes* memiliki nilai rata-rata akurasi sejumlah 97.39% serta nilai AUC berjumlah 1.000. sedangkan *k-nearest neighbors* memiliki tingkat rata-rata akurasi sebesar 98.33% dengan nilai AUC sebesar 1.000. Penelitian ini melakukan percobaan menggunakan beberapa metode pengelompokan yang jitu dalam memperkirakan kepuasan pengguna rosalia indah. Untuk hasil komparasi bisa dilihat pada tabel IV.21.

Tabel IV. 21 Keseluruhan Pengujian Bayes dan K-nn

No	Percobaan	Naive Bayes		K-NN	
		Accuracy	AUC	Accuracy	AUC
1	Percobaan 1 (Training 90% dan Testing 10 %)	96.77%	1.000	96.77%	1.000
2	Percobaan 2 (Training 80% dan Testing 20 %)	98.39%	1.000	98.39%	1.000
3	Percobaan 3 (Training 70% dan Testing 30 %)	96.81%	1.000	98.94%	1.000
4	Percobaan 4 (Training 60% dan Testing 40 %)	97.60%	1.000	99.20%	1.000
	Rata-Rata	97.39%	1.000	98.33%	1.000

Sumber: Hasil Penelitian

Tabel IV. 22 Hasil Komparasi Naive Bayes dan K-nn

Algoritma	Accuracy	AUC
Naive Bayes	97.39%	1.000
K-NN	98.33%	1.000

Sumber: Hasil Penelitian

Dari hasil pengujian menggunakan *naïve bayes* dan *k-nn* tersebut dapat memprediksi kepuasan pengguna bus rosalia indah dengan faktor kualitas pelayanan dan harga saja. Dan diperoleh hasil bila algoritma *k-nearest neighbors* menunjukkan tingkat akurasi keseluruhan akurasi nilai cukup besar dibandingkan algoritma *naïve bayes* dengan selisih mencapai 0.94 % dan nilai AUC sebesar 0. Adapun faktor yang dapat mempengaruhi hasil tersebut, diantaranya adalah jumlah data dan banyaknya atribut yang digunakan sangat mempengaruhi tingkat akurasi.



BAB V

KESIMPULAN

5.1 Kesimpulan

Dalam penelitian ini diperoleh kesimpulan, diantaranya:

1. Hasil dari penelitian ini menunjukkan kalau algoritma *k-nearest neighbors* mempunyai tingkat keseluruhan akurasi cukup besar daripada algoritma *naïve bayes* dengan perbedaan tidak terlalu jauh, hanya sebanyak 0.94%.
2. Berdasarkan hasil dari komparasi dua algoritma data mining, yakni algoritma *naïve bayes* dan algoritma *k-nearest neighbors*, disimpulkan apabila algoritma *k-nearest neighbors* menunjukkan tingkat keseluruhan akurasi cukup besar, yakni sebesar 98.33% dengan AUC 1.000, sedangkan algoritma *naïve bayes* mencatatkan tingkat keseluruhan akurasi berjumlah 97.39% dan AUC 1.000.
3. Beberapa faktor yang memengaruhi hasil dari tingkat akurasi kepuasan pengguna pada penelitian ini diantaranya adalah harga dan kualitas pelayanan.
4. Dalam penelitian ini, penulis menguji hipotesis berdasarkan penelitian tentang kepuasan pengguna yang menunjukkan bahwa pengujian menggunakan metode algoritma *K-nearest neighbors* memiliki perbedaan signifikan pada tingkat akurasi dengan *naïve bayes* dalam memprediksi tingkat kepuasan pengguna bus rosalia indah

5.2 Saran

Untuk pengembangan penelitian ini kedepannya, peneliti memberikan beberapa saran yang dapat dilakukan, yaitu:

1. Karena pengujian ini hanya membandingkan beberapa algoritma populer lainnya untuk akurasi mengetahui tingkat kepuasan pengguna, untuk kedepannya bisa dilakukan dengan menggunakan algoritma lain seperti *decision tree* dan *random forest* untuk penelitian yang berkaitan dengan kepuasan pengguna.
2. Penggunaan data yang lebih beragam dengan penambahan hal-hal atau faktor yang beragam untuk meningkatkan hasil akurasi dan membantu meningkatkan pelayanan perusahaan otobus rosalia indah sehingga meminimalisir penumpang berpindah layanan ke perusahaan bus lainnya.
3. Membandingkan kinerja Naïve Bayes dan KNN dalam memprediksi tingkat kepuasan pengguna dengan menggunakan *confusion matrix* yang lain seperti *precision*, *recall*, dan *F1-score*.



UNIVERSITAS

DAFTAR PUSTAKA

- Agustina, D., & Rahmah, F. (2022). Analisis Sentimen pada Sosial Media Twitter terhadap MRT Jakarta Menggunakan Machine Learning. *Insearch (Information System Research) Journal*, 2, 1–6.
- Amir, M., & Rahman, A. (2020). Analisis Dampak Transportasi Online Terhadap Transportasi Konvensional (Bentor) di Kota Makassar. *Mirai Management*, 5(1), 313–329. <https://journal.stieamkop.ac.id/index.php/miraipg.313>
- Diansyah, S. (2022). Klasifikasi Tingkat Kepuasan Pengguna dengan Menggunakan Metode K-Nearest Neighbour (KNN). *Jurnal Sistim Informasi Dan Teknologi*, 4, 7–12. <https://doi.org/10.37034/jsisfotek.v4i1.114>
- Febriansyah, R. F., Sukardi, F. R., & Aini, Q. (2019). Rekomendasi Moda Transportasi Mahasiswa Dengan Algoritma Naive Bayes (Studi Kasus: Fakultas Sains Dan Teknologi Uin Syarif *Simetris: Jurnal Teknik Mesin ...*, 10(2), 733–740. <https://jurnal.umk.ac.id/index.php/simet/article/view/3566>
- Ferdila, M., & Anwar, K. U. (2021). Analisis Dampak Transportasi Ojek Online Terhadap Pendapatan Ojek Konvensional di Kota Jambi. *IJIEB: Indonesian Journal of Islamic Economics and Business*, 6(2), 2021. <http://e-journal.lp2m.uinjambi.ac.id/ojp/index.php/ijieib>
- Firdaus, L., & Setiadi, T. (2023). Perbandingan Algoritma Naive Bayes, Decision Tree, dan KNN untuk Klasifikasi Produk Populer Adidas US dengan Confusion Matrix. *Jurnal Sistem Komputer Dan Informatika (JSON)*, 5(2), 185–195. <https://doi.org/10.30865/json.v5i2.6124>
- Giarsyani, N. (2020). Komparasi Algoritma Machine Learning dan Deep Learning untuk Named Entity Recognition : Studi Kasus Data Kebencanaan. *Indonesian Journal of Applied Informatics*, 4(2), 138. <https://doi.org/10.20961/ijai.v4i2.41317>
- Gunawan, M., Zarlis, M., & Roslina, R. (2021). Analisis Komparasi Algoritma Naïve Bayes dan K-Nearest Neighbor Untuk Memprediksi Kelulusan Mahasiswa Tepat Waktu. *Jurnal Media Informatika Budidarma*, 5(2), 513. <https://doi.org/10.30865/mib.v5i2.2925>
- Hakim, Z. R., & Sugiyono, S. (2024). Analisa Sentimen Terhadap Kereta Cepat Jakarta – Bandung. *Jurnal Sains Dan Teknologi*, 5(3), 939–945.

<https://doi.org/10.55338/saintek.v5i3.1423> Analisa

- Hamid, A., & Susanti, F. (2023). PENGARUH HARGA DAN KUALITAS LAYANAN TERHADAP KEPUASAN PELANGGAN (Studi pada Mahasiswa Pengguna Layanan Go-Jek di STIE-KBP Padang). *Jurnal Economina*, 2(3), 836–847. <https://doi.org/10.55681/economina.v2i3.405>
- Ilhamsyah, F., Ginting, R., & Setiawan, A. (2020). Analisis Kepuasan Pengguna Jasa Transportasi Umum Jak Lingko Di Jakarta Selatan. *Business Management, Economic, and Accounting National Seminar*, 1(1), 1059–1076.
- Iwandini, I., Triayudi, A., & Soepriyono, G. (2023). Analisa Sentimen Pengguna Transportasi Jakarta Terhadap Transjakarta Menggunakan Metode Naives Bayes dan K-Nearest Neighbor. *Journal of Information System Research (JOSH)*, 4(2), 543–550. <https://doi.org/10.47065/josh.v4i2.2937>
- Janna, N. M., & Herianto. (2021). Artikel Statistik yang Benar. *Jurnal Darul Dakwah Wal-Irsyad (DDI)*, 18210047, 1–12.
- Kartarina, K., Sriwinarti, N. K., & Juniarti, N. luh P. (2021). Analisis Metode K-Nearest Neighbors (K-NN) Dan Naive Bayes Dalam Memprediksi Kelulusan Mahasiswa. *JTIM : Jurnal Teknologi Informasi Dan Multimedia*, 3(2), 107–113. <https://doi.org/10.35746/jtim.v3i2.159>
- Kumrotin, E. L., & Susanti, A. (2021). Pengaruh Kualitas Produk, Harga, Dan Kualitas Pelayanan Terhadap Kepuasan Konsumen Pada Cafe Ko.We.Cok Di Solo. *J-MIND (Jurnal Manajemen Indonesia)*, 6(1), 1. <https://doi.org/10.29103/j-mind.v6i1.4870>
- Mardiyyah, N. W., Rahaningsih, N., Ali, I., & Neighbor, K. (2024). PENERAPAN DATA MINING MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR PADA PREDIKSI PEMBERIAN KREDIT DI SEKTOR FINANSIAL. 8(2), 1491–1499.
- Maulana, R., Raihan, M., & Santoso, I. (2023). Komparasi Algoritma Naive Bayes Dan K-Nearest Neighbor Pada Analisis Sentimen Terhadap Ulasan Pengguna Aplikasi Tokopedia. *Jurnal Teknologi Informasi: Jurnal Keilmuan Dan Aplikasi Bidang Teknik Informatika*, 7(2), 177–189. <https://doi.org/10.47111/jti.v7i2.10071>
- Muharrom, M. (2019). Komparasi Algoritma Klasifikasi Naive Bayes Dan K-Nearest Neighbors Untuk Prediksi Penyakit Jantung. *Syntax : Jurnal*

- Informatika*, 8(1), 44–56. <https://doi.org/10.35706/syji.v8i1.1856>
- Musu, W., Ibrahim, A., & Heriadi. (2021). Pengaruh Komposisi Data Training dan Testing terhadap Akurasi Algoritma C4.5. *Prosiding Seminar Ilmiah Sistem Informasi Dan Teknologi Informasi*, X(1), 186–195.
- Novianti, D. (2019). Implementasi Algoritma Naïve Bayes Pada Data Set Hepatitis Menggunakan Rapid Miner. *Paradigma - Jurnal Komputer Dan Informatika*, 21(1), 49–54. <https://doi.org/10.31294/p.v21i1.4979>
- Prasetyo, P. A. M., & Hermawan, A. (2023). Analisis sentimen twitter terhadap pemilihan presiden menggunakan algoritma naïve bayes. *INFOTECH : Jurnal Informatika & Teknologi*, 4(2), 224–233. <https://doi.org/10.37373/infotech.v4i2.863>
- Rahmat, A. ., Ladjamuddin, M. ., & Awaludin, T. . (2023). Perbandingan Algoritma Decision Tree, Random Forest Dan Naive Bayes Pada Prediksi Penilaian Kepuasan Penumpang Maskapai Pesawat Menggunakan Dataset Kaggle. *Jurnal Rekayasa Informasi*, 12(2), 150–159. www.kaggle.com,
- Rooroh, C., Moniharapon, S., & Loindong, S. (2020). Pengaruh Suasana Cafe, Kualitas Pelayanan Dan Harga Terhadap Kepuasan Pelanggan (Studi Pada Cafe Casa De Wanea Manado). *Jurnal EMBA: Jurnal Riset Ekonomi, Manajemen, Bisnis Dan Akuntansi*, 8(4), 130–139.
- Salsabilah, B. N., Kadek, I., & Nuryana, D. (2023). Komparasi Algoritma Naive Bayes dan K Nearest Neighbor dalam Kepuasan Pengguna Fitur Tiktok Shop. *Journal of Emerging Information System and Business Intelligence (JEISBI)*, 4(3), 31–39. <https://ejournal.unesa.ac.id/index.php/JEISBI/article/view/54211%0Ahttps://ejournal.unesa.ac.id>
- Saphari, A. (2023). Sentiment Analisis Pada Moda Transportasi Bus Untuk Meningkatkan Kualitas Layanan Pada Perusahaan Otobus Terendah. *Jurnal Teknik Dan Komputer (JITKOM)*, 07(01), 1–79. <https://doi.org/10.22441/jitkom.v7i1.010>
- Sariatin, S., & Ekawati, C. (2023). Faktor-faktor yang Mempengaruhi Kepuasan Pelanggan. *Jurnal Informatika Ekonomi Bisnis*, 5, 1258–1261. <https://doi.org/10.37034/infkeb.v5i4.772>
- Setiawan, B. P., & Frianto, A. (2021). Pengaruh Harga dan Kualitas Pelayanan

- Terhadap Kepuasan Pelanggan (Studi Kasus Perusahaan Jasa Ekspedisi Krian). *BIMA: Journal of Business and Innovation Management*, 3(3), 352–366. <https://doi.org/10.33752/bima.v3i3.5493>
- Styawati, S., Hendrastuty, N., & Isnain, A. R. (2021). Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine. *Jurnal Informatika: Jurnal Pengembangan IT*, 6(3), 150–155. <https://doi.org/10.30591/jpit.v6i3.2870>
- Suriani, N., Risnita, & Jailani, M. S. (2023). Konsep Populasi dan Sampling Serta Pemilihan Partisipan Ditinjau Dari Penelitian Ilmiah Pendidikan. *Jurnal IHSAN: Jurnal Pendidikan Islam*, 1(2), 24–36. <https://doi.org/10.61104/ihsan.v1i2.55>
- Tirta Komara, A., Yura Roslina, N., Jatmika, L., & Pasundan, S. (2021). Pengaruh Kualitas Pelayanan Dan Harga Terhadap Kepuasan Konsumen (Studi pada salah satu perusahaan jasa transportasi di Kota Bandung). *Acman: Accounting and Management Journal*, 1(2), 104–114.
- Wiyanto. (2020). Analisa Tingkat Kepuasan Pelanggan Terhadap Pelayanan Perusahaan Otobus XYZ Menggunakan Metode Naive Bayes. *Jurnal Pelita Teknologi*, 15(1), 56–67.
- Yuliarina, A. N., & Hendry, H. (2022). Comparison of Prediction Analysis of Gofood Service Users Using the Knn & Naive Bayes Algorithm With Rapidminer Software. *Jurnal Teknik Informatika (Jutif)*, 3(4), 847–856. <https://doi.org/10.20884/1.jutif.2022.3.4.294>
- Zai, C. (2022). Implementasi Data Mining Sebagai Pengolahan Data. *Jurnal Teknoinfo*, 16(1), 46. <http://portaldata.org/index.php/portaldata/article/view/107>

DAFTAR RIWAYAT HIDUP

I. Biodata Mahasiswa

NIM : 15200304
Nama Lengkap : Dimas Miftakhul Fakri
Tempat/ Tanggal Lahir : Bekasi, 25 Juli 2002
Alamat Lengkap : Taman Tridaya Indah 1 Blok A.02 No.7, RT 002, RW 009, Kelurahan Tridayasakti, Kecamatan Tambun Selatan. Kab. Bekasi, Provinsi Jawa Barat, Kode Pos 17510

II. Pendidikan

a. Formal

1. SD Negeri Tridayasakti 01, lulus tahun 2014
2. SMP Negeri 5 Tambun Selatan, lulus tahun 2017
3. SMK 10 November Bekasi, lulus tahun 2020
4. S1 Ilmu Komputer Universitas Bina Sarana Informatika (Sekarang)

III. Riwayat Pekerjaan

1. Digital Marketing di Kantor Taman BAR, Okt 2017 – Jan 2018
2. Digital Journalism Internship Program, Kampus Merdeka di PT Trans Digital Media (detikcom), Feb 2023 – Jun 2023
3. Digital Marketing intern di PT. Sahabat Safar Haramain, Okt 2023 – Januari 2024

Jakarta, 28 Juni 2024



Dimas Miftakhul Fakri

	LEMBAR KONSULTASI SKRIPSI
	UNIVERSITAS BINA SARANA INFORMATIKA

NIM : 15200304
 Nama Lengkap : Dimas Miftakhul Fakri
 Dosen Pembimbing : Waeisul Bismi, M.Kom.
 Judul Skripsi : Komparasi Algoritma Machine Learning Untuk Memprediksi Keuasan Penggunaan Jasa Transportasi Bus Rosalia Indah

No	Tanggal Bimbingan	Pokok Bahasan	Paraf Dosen Pembimbing
1.	1 April 2024	Sosialisasi dan Pengajuan Judul	
2.	24 April 2024	ACC judul dan Pengajuan Bab 1	
3.	8 Mei 2024	Revisi Bab 1 dan Pengajuan Bab 2	
4.	30 Mei 2024	ACC bab 1 dan Revisi Bab 2	
5.	5 Juni 2024	ACC Bab 2 dan Pengajuan Bab 3 dan Bab 4	
6.	19 juni 2024	Revisi Bab 3 dan Bab 4	
7.	27 juni 2024	Pengajuan Bab 5 dan ACC Bab 3 dan Bab 4	
8.	28 juni 2024	Pengajuan Power Point dan ACC Keseluruhan	

Catatan untuk Dosen Pembimbing.

Bimbingan Skripsi

Dimulai pada tanggal : 1 April 2024

Diakhiri pada tanggal : 28 Juni 2024

Jumlah pertemuan bimbingan : 8 Kali Bimbingan

Disetujui oleh,



(Waeisul Bismi, M.Kom)

**SURAT PERNYATAAN KEBENARAN/KEABSAHAN DATA
HASIL RISET UNTUK KARYA ILMIAH**

Yang bertanda tangan di bawah ini, saya:

Nama : Dimas Miftakhul Fakri
NIM : 15200304
Jenjang : Sarjana (S1)
Program Studi : Informatika
Fakultas : Teknik dan Informatika
Perguruan Tinggi : Universitas Bina Sarana Informatika

Dengan ini menyatakan bahwa data dan atau informasi yang saya gunakan dalam penulisan karya ilmiah dengan judul **“Komparasi Algoritma Machine Learning Dalam Memprediksi Kepuasan Penggunaan Jasa Transportasi Bus Rosalia Indah”** merupakan data dan atau informasi yang saya peroleh melalui hasil penelitian sendiri dan tidak didasarkan pada data atau informasi hasil riset dari perusahaan/instansi/lembaga manapun.

Saya bersedia untuk bertanggung jawab secara pribadi, tanpa melibatkan pihak **Universitas Bina Sarana Informatika**, atas materi/isi karya ilmiah tersebut, termasuk bertanggung jawab atas dampak atau kerugian yang timbul dalam bentuk akibat tindakan yang berkaitan dengan data dan atau informasi yang terdapat pada karya ilmiah saya ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta
Pada tanggal : 28 Juni 2024

Mengetahui,

Yang Menyatakan

Dosen Pembimbing

Wacisul Bismi, M.Kom



Dimas Miftakhul Fakri

LAMPIRAN-LAMPIRAN

Lampiran 1. Cek Plagiat

TURNITIN_15200304_DIMAS MIFTAKHUL FAKRI.pdf

ORIGINALITY REPORT

20%	19%	9%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	repository.bsi.ac.id Internet Source	2%
2	ejournal.istn.ac.id Internet Source	1%
3	repository.uin-suska.ac.id Internet Source	1%
4	download.garuda.kemdikbud.go.id Internet Source	1%
5	journal.amikveteran.ac.id Internet Source	1%
6	journal.sekawan-org.id Internet Source	1%
7	docplayer.info Internet Source	1%
8	ejournal.itn.ac.id Internet Source	1%
9	www.researchgate.net Internet Source	<1%

10	e-journals.unmul.ac.id Internet Source	<1 %
11	id.123dok.com Internet Source	<1 %
12	ilmuwanmuda.wordpress.com Internet Source	<1 %
13	e-journal.upr.ac.id Internet Source	<1 %
14	123dok.com Internet Source	<1 %
15	Submitted to kopusat.turnitin@gmail.com Student Paper	<1 %
16	pt.scribd.com Internet Source	<1 %
17	skripsi.tunasbangsa.ac.id Internet Source	<1 %
18	ejournal.almaata.ac.id Internet Source	<1 %
19	repository.its.ac.id Internet Source	<1 %
20	lppm.unpam.ac.id Internet Source	<1 %
21	Nailufar Farha Afifah, Apriade Voutama. "Analisis Sentimen Isu Perselingkuhan pada	<1 %

Postingan Autbase Twitter @tanyarlfe
Menggunakan Metode Naïve Bayes",
Bianglala Informatika, 2023

Publication

22	Submitted to UIN Sultan Syarif Kasim Riau Student Paper	<1 %
23	Submitted to Fakultas Ekonomi Universitas Indonesia Student Paper	<1 %
24	repository.ub.ac.id Internet Source	<1 %
25	jurnal.stkipggritulungagung.ac.id Internet Source	<1 %
26	id.scribd.com Internet Source	<1 %
27	es.scribd.com Internet Source	<1 %
28	www.scribd.com Internet Source	<1 %
29	Submitted to President University Student Paper	<1 %
30	eprints.iain-surakarta.ac.id Internet Source	<1 %
31	Lusiana Efrizoni, Sarjon Defit, Muhammad Tajuddin, Anthony Anggrawan. "Komparasi	<1 %

Ekstraksi Fitur dalam Klasifikasi Teks
Multilabel Menggunakan Algoritma Machine
Learning", MATRIK : Jurnal Manajemen, Teknik
Informatika dan Rekayasa Komputer, 2022
Publication

-
- | | | |
|-----------|--|----------------|
| 32 | Surohman Surohman, Sopian Aji, Rousyati Rousyati, Fanny Fatma Wati. "Analisa Sentimen Terhadap Review Fintech Dengan Metode Naive Bayes Classifier Dan K- Nearest Neighbor", EVOLUSI : Jurnal Sains dan Manajemen, 2020
Publication | <1 % |
|-----------|--|----------------|
-
- | | | |
|-----------|--|----------------|
| 33 | Submitted to Universitas Diponegoro
Student Paper | <1 % |
|-----------|--|----------------|
-
- | | | |
|-----------|---|----------------|
| 34 | Submitted to Universitas Trunojoyo
Student Paper | <1 % |
|-----------|---|----------------|
-
- | | | |
|-----------|--|----------------|
| 35 | repository.unisma.ac.id
Internet Source | <1 % |
|-----------|--|----------------|
-
- | | | |
|-----------|---|----------------|
| 36 | etheses.uin-malang.ac.id
Internet Source | <1 % |
|-----------|---|----------------|
-
- | | | |
|-----------|---|----------------|
| 37 | Submitted to Universiti Malaysia Sabah
Student Paper | <1 % |
|-----------|---|----------------|
-
- | | | |
|-----------|------------------------------------|----------------|
| 38 | vdocuments.site
Internet Source | <1 % |
|-----------|------------------------------------|----------------|
-
- | | | |
|-----------|--|----------------|
| 39 | Submitted to Sriwijaya University
Student Paper | <1 % |
|-----------|--|----------------|
-

40	Submitted to Universitas Negeri Padang Student Paper	<1 %
41	e-journal.uajy.ac.id Internet Source	<1 %
42	johannessimatupang.wordpress.com Internet Source	<1 %
43	ojs.stmik-banjarbaru.ac.id Internet Source	<1 %
44	Budiman Budiman, Zulmeida Silvana Anggraeni, Chairul Habibi, Nur Alamsyah. "Analisis Sentimen Publik pada Media Sosial Twitter Terhadap Tiket.com Menggunakan Algoritma Klasifikasi", Jurnal Informatika, 2024 Publication	<1 %
45	Pungkas Subarkah, Enggar Pri Pambudi, Septi Oktaviani Nur Hidayah. "Perbandingan Metode Klasifikasi Data Mining untuk Nasabah Bank Telemarketing", MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer, 2020 Publication	<1 %
46	journal.literasisains.id Internet Source	<1 %
47	jurnal.polibatam.ac.id Internet Source	<1 %

48	www.scilit.net Internet Source	<1 %
49	Anggita Safitri Febriarini, Erna Zuni Astuti. "Penerapan Algoritma C4.5 untuk Prediksi Kepuasan Penumpang Bus Rapid Transit (BRT) Trans Semarang", Eksplora Informatika, 2019 Publication	<1 %
50	adoc.pub Internet Source	<1 %
51	dspace.umkt.ac.id Internet Source	<1 %
52	ejournal.unesa.ac.id Internet Source	<1 %
53	idoc.pub Internet Source	<1 %
54	journal.lppmunindra.ac.id Internet Source	<1 %
55	repositori.uin-alauddin.ac.id Internet Source	<1 %
56	repository.stiedewantara.ac.id Internet Source	<1 %
57	tempatberwisatamurah.com Internet Source	<1 %

58	Salman Alfaris, Kusnawi. "Komparasi Metode KNN dan Naive Bayes Terhadap Analisis Sentimen Pengguna Aplikasi Shopee", Indonesian Journal of Computer Science, 2023 Publication	<1 %
59	Submitted to Universitas Gunadarma Student Paper	<1 %
60	cakrawala.imwi.ac.id Internet Source	<1 %
61	ejournal.sidyanusa.org Internet Source	<1 %
62	ejurnal.unisan.ac.id Internet Source	<1 %
63	garuda.kemdikbud.go.id Internet Source	<1 %
64	journal.ibrahimy.ac.id Internet Source	<1 %
65	media.neliti.com Internet Source	<1 %
66	pustaka.unama.ac.id Internet Source	<1 %
67	repository.uph.edu Internet Source	<1 %
68	repository.iainpurwokerto.ac.id Internet Source	<1 %

69 Wulan Sri Lestari. "Optimasi Model Prediksi Kesuksesan Startup Menggunakan StandartScaler Tranform", Seminar Nasional Teknologi & Sains, 2024 <1%
Publication

70 Yusuf Fadlila Rachman, Kusrini Kusrini, Hanif Al Fatta. "Klasifikasi Citra Digitalretina Penderita Diabetes Retinopati Menggunakan Metode Euclidean", DoubleClick: Journal of Computer and Information Technology, 2020 <1%
Publication

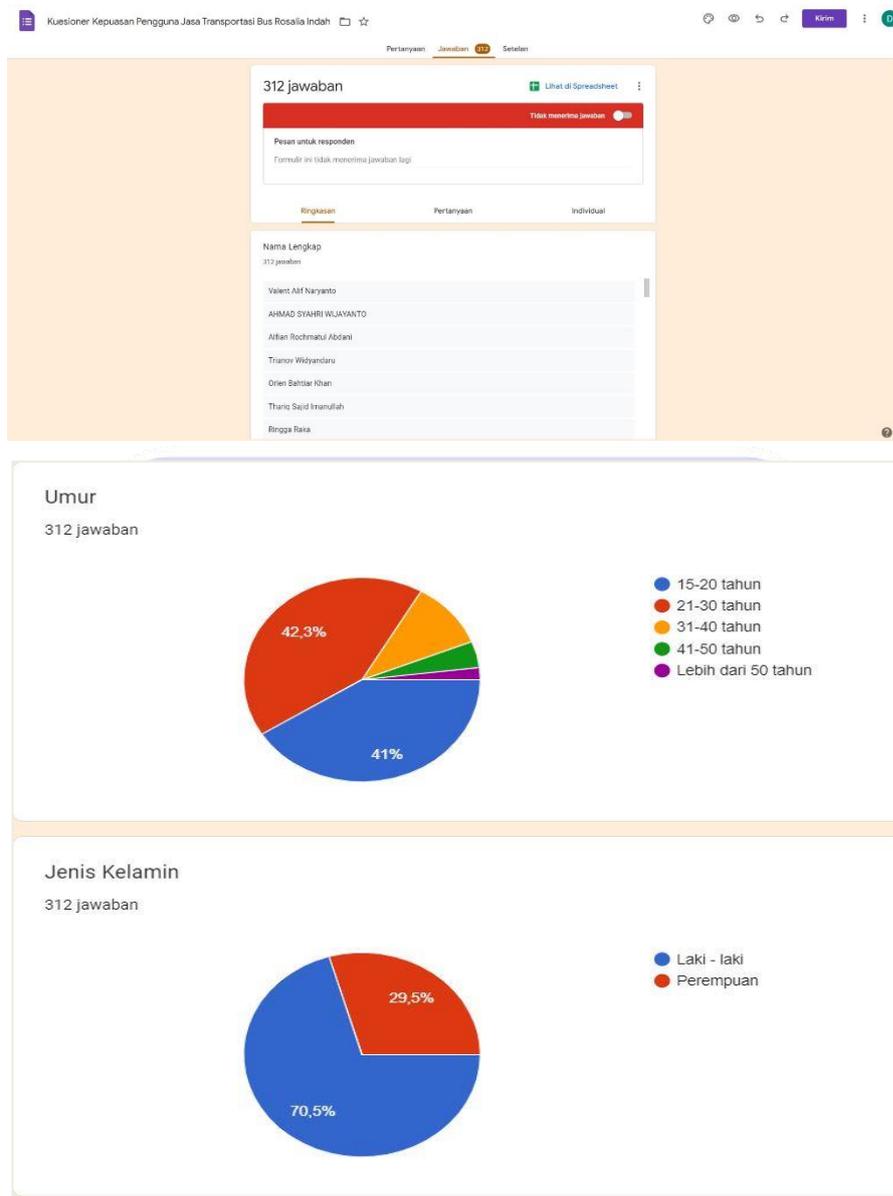
Exclude quotes Off

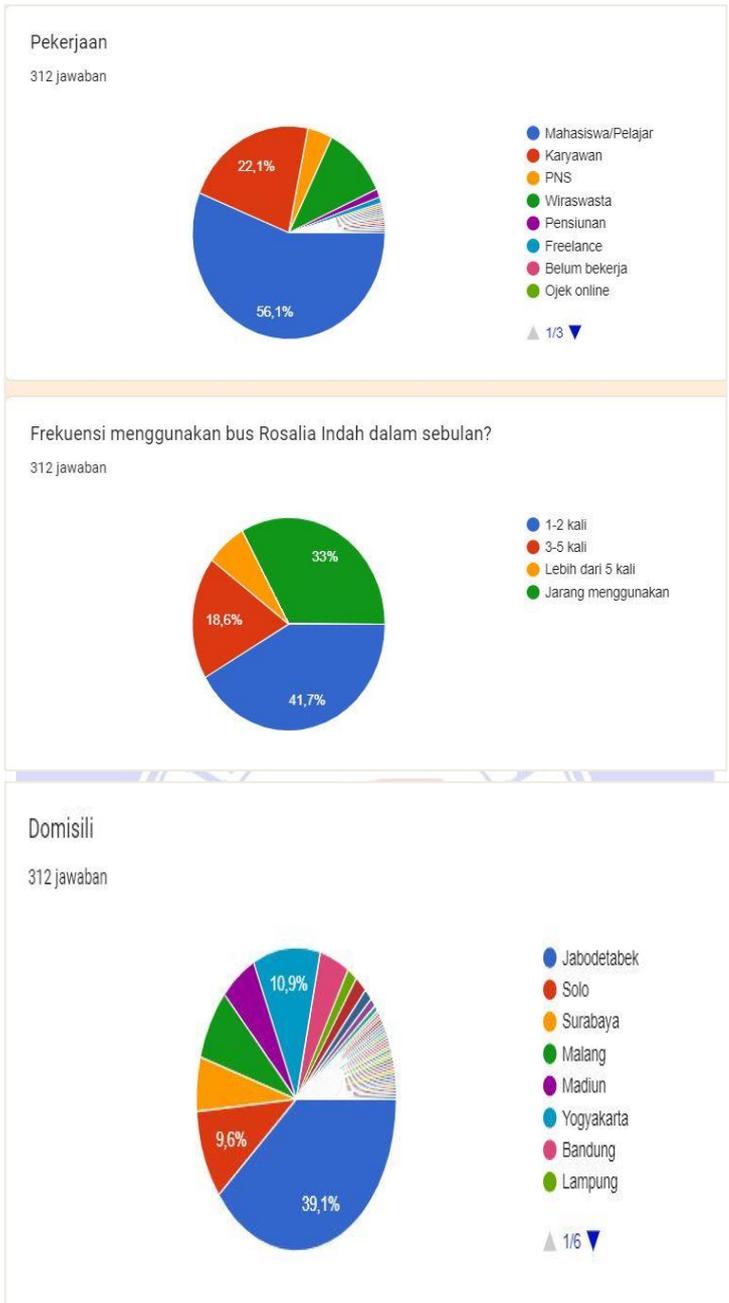
Exclude matches Off

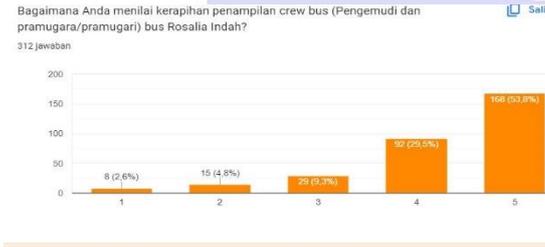
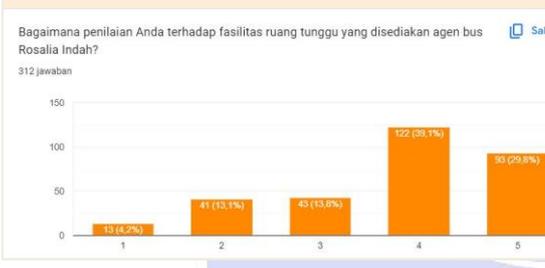
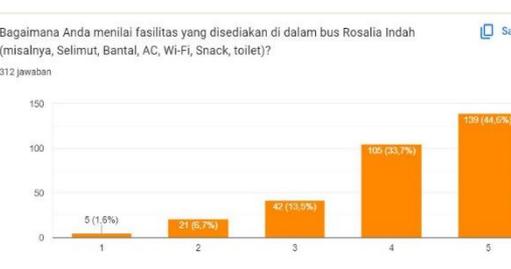
Exclude bibliography Off

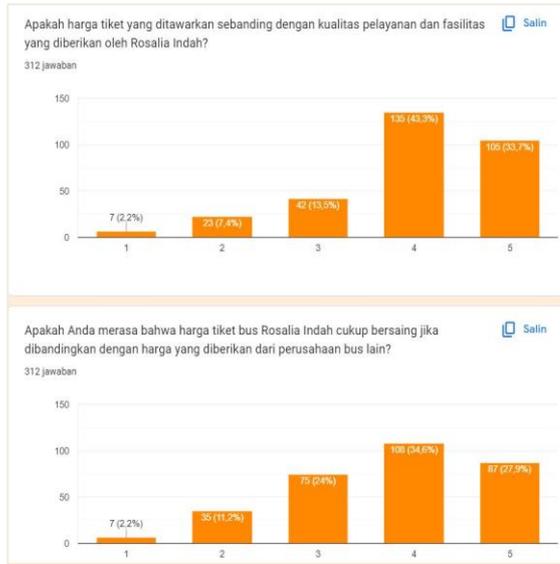


Lampiran 2. Kuesioner



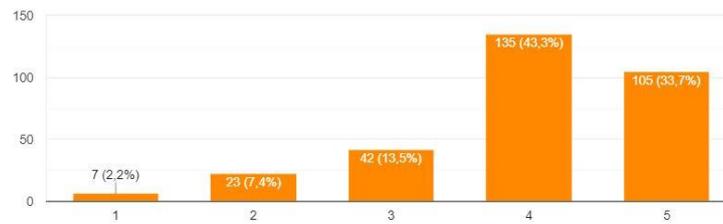






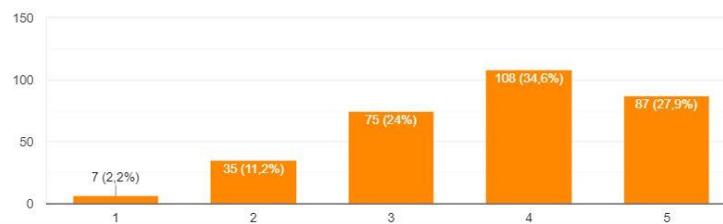
Apakah harga tiket yang ditawarkan sebanding dengan kualitas pelayanan dan fasilitas yang diberikan oleh Rosalia Indah? [Salin](#)

312 jawaban



Apakah Anda merasa bahwa harga tiket bus Rosalia Indah cukup bersaing jika dibandingkan dengan harga yang diberikan dari perusahaan bus lain? [Salin](#)

312 jawaban

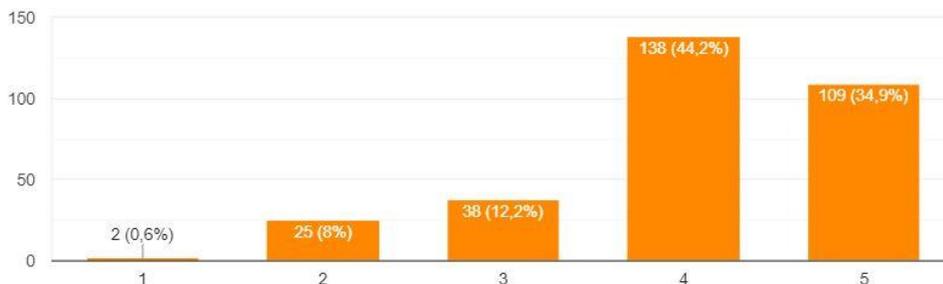


Kepuasan Pengguna

Bagaimana penilaian Anda terhadap pelayanan yang diberikan oleh pihak bus Rosalia Indah?



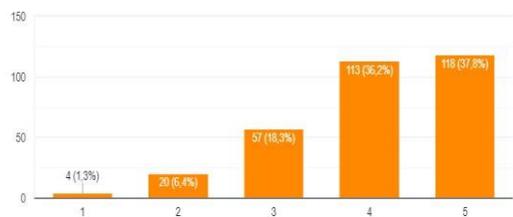
312 jawaban



Saya akan memberikan informasi yang positif tentang bus Rosalia Indah?



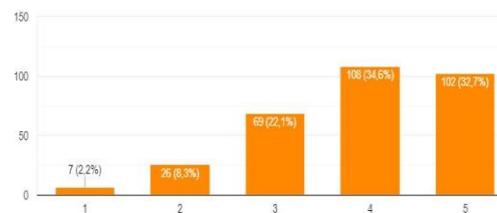
312 jawaban



Seberapa besar Anda akan menjadikan bus Rosalia Indah sebagai Transportasi utama untuk berpergian ke luar kota?



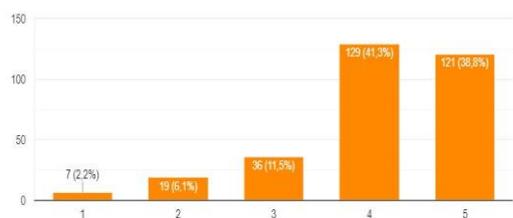
312 jawaban



Seberapa puas Anda dengan keseluruhan layanan yang diberikan oleh bus Rosalia Indah?



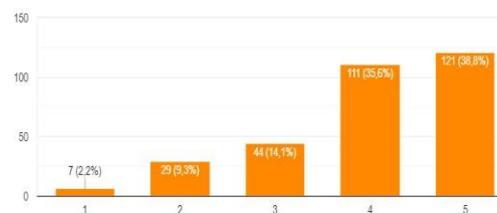
312 jawaban



Seberapa besar Anda akan merekomendasikan bus Rosalia Indah kepada orang terdekat Anda?



312 jawaban



Lampiran 3. Uji Validitas dan Reliabilitas

		Correlations			
		A01	A02	A03	Total
A01	Pearson Correlation	1	.590**	.520**	.824**
	Sig. (2-tailed)		<.001	<.001	<.001
	N	312	312	312	312
A02	Pearson Correlation	.590**	1	.532**	.868**
	Sig. (2-tailed)	<.001		<.001	<.001
	N	312	312	312	312
A03	Pearson Correlation	.520**	.532**	1	.812**
	Sig. (2-tailed)	<.001	<.001		<.001
	N	312	312	312	312
Total	Pearson Correlation	.824**	.868**	.812**	1
	Sig. (2-tailed)	<.001	<.001	<.001	
	N	312	312	312	312

** . Correlation is significant at the 0.01 level (2-tailed).

Reliability

Scale: ALL VARIABLES

Case Processing Summary

		N	%
Cases	Valid	312	100.0
	Excluded ^a	0	.0
	Total	312	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.779	3

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
A01	4.9455	6.978	.636	.689
A02	4.9135	5.430	.642	.683
A03	4.7372	6.786	.590	.729

		Correlations													
		B01	B02	B03	B04	B05	B06	B07	B08	B09	B10	B11	B12	B13	Total
B01	Pearson Correlation	1	.633**	.636**	.653**	.634**	.496**	.640**	.571**	.590**	.643**	.584**	.629**	.595**	.803**
	Sig. (2-tailed)		<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
	N	312	312	312	312	312	312	312	312	312	312	312	312	312	312
B02	Pearson Correlation	.633**	1	.590**	.662**	.538**	.571**	.625**	.705**	.661**	.614**	.694**	.638**	.509**	.813**
	Sig. (2-tailed)	<.001		<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
	N	312	312	312	312	312	312	312	312	312	312	312	312	312	312
B03	Pearson Correlation	.636**	.590**	1	.588**	.616**	.440**	.577**	.434**	.553**	.574**	.494**	.567**	.607**	.747**
	Sig. (2-tailed)	<.001	<.001		<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
	N	312	312	312	312	312	312	312	312	312	312	312	312	312	312
B04	Pearson Correlation	.653**	.662**	.598**	1	.532**	.593**	.651**	.661**	.668**	.625**	.710**	.666**	.571**	.826**
	Sig. (2-tailed)	<.001	<.001	<.001		<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
	N	312	312	312	312	312	312	312	312	312	312	312	312	312	312
B05	Pearson Correlation	.634**	.538**	.616**	.532**	1	.393**	.594**	.438**	.468**	.613**	.520**	.627**	.636**	.744**
	Sig. (2-tailed)	<.001	<.001	<.001	<.001		<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
	N	312	312	312	312	312	312	312	312	312	312	312	312	312	312
B06	Pearson Correlation	.496**	.571**	.440**	.593**	.393**	1	.605**	.631**	.637**	.561**	.650**	.564**	.422**	.726**
	Sig. (2-tailed)	<.001	<.001	<.001	<.001	<.001		<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
	N	312	312	312	312	312	312	312	312	312	312	312	312	312	312
B07	Pearson Correlation	.640**	.625**	.577**	.651**	.594**	.605**	1	.595**	.621**	.637**	.658**	.663**	.585**	.816**
	Sig. (2-tailed)	<.001	<.001	<.001	<.001	<.001	<.001		<.001	<.001	<.001	<.001	<.001	<.001	<.001
	N	312	312	312	312	312	312	312	312	312	312	312	312	312	312
B08	Pearson Correlation	.571**	.705**	.434**	.661**	.438**	.631**	.595**	1	.778**	.629**	.760**	.634**	.497**	.799**
	Sig. (2-tailed)	<.001	<.001	<.001	<.001	<.001	<.001	<.001		<.001	<.001	<.001	<.001	<.001	<.001
	N	312	312	312	312	312	312	312	312	312	312	312	312	312	312
B09	Pearson Correlation	.590**	.661**	.553**	.668**	.468**	.637**	.621**	.778**	1	.633**	.737**	.665**	.517**	.819**
	Sig. (2-tailed)	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001		<.001	<.001	<.001	<.001	<.001
	N	312	312	312	312	312	312	312	312	312	312	312	312	312	312
B10	Pearson Correlation	.643**	.614**	.574**	.625**	.613**	.637**	.629**	.633**	.633**	1	.686**	.756**	.587**	.827**
	Sig. (2-tailed)	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001		<.001	<.001	<.001	<.001
	N	312	312	312	312	312	312	312	312	312	312	312	312	312	312
B11	Pearson Correlation	.584**	.694**	.494**	.710**	.520**	.650**	.658**	.760**	.737**	.686**	1	.693**	.513**	.836**
	Sig. (2-tailed)	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001		<.001	<.001	<.001
	N	312	312	312	312	312	312	312	312	312	312	312	312	312	312
B12	Pearson Correlation	.629**	.638**	.567**	.666**	.627**	.564**	.663**	.634**	.665**	.756**	.693**	1	.637**	.845**
	Sig. (2-tailed)	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001		<.001	<.001
	N	312	312	312	312	312	312	312	312	312	312	312	312	312	312
B13	Pearson Correlation	.595**	.509**	.607**	.571**	.636**	.422**	.585**	.497**	.517**	.587**	.513**	.637**	1	.749**
	Sig. (2-tailed)	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001		<.001
	N	312	312	312	312	312	312	312	312	312	312	312	312	312	312
Total	Pearson Correlation	.803**	.813**	.747**	.826**	.744**	.726**	.816**	.799**	.819**	.827**	.836**	.845**	.749**	1
	Sig. (2-tailed)	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
	N	312	312	312	312	312	312	312	312	312	312	312	312	312	312

** Correlation is significant at the 0.01 level (2-tailed).

Reliability

Scale: ALL VARIABLES

Case Processing Summary			
		N	%
Cases	Valid	312	100.0
	Excluded ^a	0	.0
	Total	312	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.951	13

Item-Total Statistics				
	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
B01	47.9391	94.437	.766	.947
B02	47.6987	94.179	.777	.947
B03	48.0962	94.267	.696	.949
B04	47.7500	94.297	.793	.946
B05	48.0385	93.664	.689	.949
B06	47.6827	96.037	.677	.949
B07	47.8269	93.970	.780	.946
B08	47.5385	94.442	.761	.947
B09	47.6859	94.583	.785	.946
B10	47.8301	93.518	.793	.946
B11	47.6058	94.072	.804	.946
B12	47.8462	93.076	.813	.945
B13	48.1923	93.956	.697	.949

		Correlations				
		C01	C02	C03	C04	Total
C01	Pearson Correlation	1	.794**	.597**	.662**	.897**
	Sig. (2-tailed)		<.001	<.001	<.001	<.001
	N	312	312	312	312	312
C02	Pearson Correlation	.794**	1	.554**	.649**	.884**
	Sig. (2-tailed)	<.001		<.001	<.001	<.001
	N	312	312	312	312	312
C03	Pearson Correlation	.597**	.554**	1	.536**	.783**
	Sig. (2-tailed)	<.001	<.001		<.001	<.001
	N	312	312	312	312	312
C04	Pearson Correlation	.662**	.649**	.536**	1	.840**
	Sig. (2-tailed)	<.001	<.001	<.001		<.001
	N	312	312	312	312	312
Total	Pearson Correlation	.897**	.884**	.783**	.840**	1
	Sig. (2-tailed)	<.001	<.001	<.001	<.001	
	N	312	312	312	312	312

** . Correlation is significant at the 0.01 level (2-tailed).

Reliability

Scale: ALL VARIABLES

Case Processing Summary

		N	%
Cases	Valid	312	100.0
	Excluded ^a	0	.0
	Total	312	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.873	4

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
C01	11.3494	6.961	.808	.806
C02	11.2885	6.843	.779	.817
C03	10.9071	7.892	.628	.875
C04	11.1474	7.219	.705	.848

		Correlations					
		D01	D02	D03	D04	D05	Total
D01	Pearson Correlation	1	.675**	.713**	.778**	.719**	.882**
	Sig. (2-tailed)		<.001	<.001	<.001	<.001	<.001
	N	312	312	312	312	312	312
D02	Pearson Correlation	.675**	1	.739**	.674**	.596**	.846**
	Sig. (2-tailed)	<.001		<.001	<.001	<.001	<.001
	N	312	312	312	312	312	312
D03	Pearson Correlation	.713**	.739**	1	.779**	.706**	.903**
	Sig. (2-tailed)	<.001	<.001		<.001	<.001	<.001
	N	312	312	312	312	312	312
D04	Pearson Correlation	.778**	.674**	.779**	1	.710**	.898**
	Sig. (2-tailed)	<.001	<.001	<.001		<.001	<.001
	N	312	312	312	312	312	312
D05	Pearson Correlation	.719**	.596**	.706**	.710**	1	.850**
	Sig. (2-tailed)	<.001	<.001	<.001	<.001		<.001
	N	312	312	312	312	312	312
Total	Pearson Correlation	.882**	.846**	.903**	.898**	.850**	1
	Sig. (2-tailed)	<.001	<.001	<.001	<.001	<.001	
	N	312	312	312	312	312	312

** . Correlation is significant at the 0.01 level (2-tailed).

Reliability

Scale: ALL VARIABLES

Case Processing Summary

		N	%
Cases	Valid	312	100.0
	Excluded ^a	0	.0
	Total	312	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.923	5

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
D01	15.9455	12.624	.818	.903
D02	16.1250	12.296	.751	.916
D03	16.0000	11.672	.839	.898
D04	15.9679	12.218	.838	.899
D05	15.9359	12.517	.764	.913