

**ANALISIS KEPUASAN KONSUMEN TERHADAP
KUALITAS PRODUK MENGGUNAKAN
ALGORITMA *NAIVE BAYES***

SKRIPSI



**Oleh:
Ahmad Suwandi
160210035**

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN KOMPUTER
UNIVERSITAS PUTERA BATAM
TAHUN 2023**

**ANALISIS KEPUASAN KONSUMEN TERHADAP
KUALITAS PRODUK MENGGUNAKAN
ALGORITMA *NAIVE BAYES***

SKRIPSI

**Untuk memenuhi salah satu syarat
memperoleh gelar sarjana**



Oleh:

Ahmad Suwandi

160210035

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN KOMPUTER
UNIVERSITAS PUTERA BATAM**

2023

SURAT PERNYATAAN ORISINALITAS

Yang bertanda tangan di bawah ini saya:

Nama : Ahmad Suwandi

NPM : 160210035

Fakultas : Teknik dan Komputer

Program Studi : Teknik Informatika

Menyatakan bahwa "Skripsi" yang saya buat dengan judul:

Analisis Kepuasan Konsumen Terhadap Kualitas Produk Menggunakan Algoritma Naive Bayes

Adalah hasil karya sendiri dan bukan "duplikasi" dari karya orang lain. Sepengetahuan saya, di dalam naskah Skripsi ini tidak terdapat karya ilmiah atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis dikutip didalam naskah ini dan disebutkan dalam sumber kutipan dan daftar pustaka.

Apabila ternyata di dalam naskah Skripsi ini dapat dibuktikan terdapat unsur-unsur PLAGIASI, saya bersedia naskah Skripsi ini digugurkan dan gelar akademik yang saya peroleh dibatalkan, serta diproses sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian pernyataan ini saya buat dengan sebenarnya tanpa ada paksaan dari siapapun

Batam, 22 Agustus 2023



Ahmad Suwandi
160210035

**ANALISIS KEPUASAN KONSUMEN TERHADAP
KUALITAS PRODUK MENGGUNAKAN
ALGORITMA *NAIVE BAYES***

SKRIPSI

**Untuk memenuhi salah satu syarat
memperoleh gelar sarjana**

Oleh:

Ahmad Suwandi

160210035

**Telah disetujui oleh pembimbing pada tanggal
seperti yang tertera dibawah ini**

Batam, 22 Agustus 2023


Rahmat Fauzi S.Kom, M.Kom

ABSTRAK

Perkembangan teknologi saat ini telah membawa perubahan pada gaya hidup manusia, seperti gaya hidup, pemilihan barang antik dan peralatan komunikasi seperti smartphone dan perangkat lainnya. Saat ini diketahui bahwa koleksi barang-barang branded juga menjadi tren gaya hidup seperti tas, cincin, pakaian dan fashion lainnya. Batam sebagai kota industri mendorong investor untuk berinvestasi di Batam, karena Batam merupakan kota paling strategis dan merupakan jalur perdagangan internasional yang berbatasan langsung dengan Singapura. Masalah yang sering terjadi di PT Tas Import Sukses dalam melayani konsumen adalah barang yang dikirim tidak sesuai permintaan, sehingga konsumen menduga bahwa produk yang diperdagangkan tidak memiliki kualitas yang baik. Dan jika ada masalah, perusahaan merespons dengan waktu yang sangat lambat. Tujuan dari penelitian ini adalah untuk menganalisis kepuasan konsumen terhadap kualitas produk di PT Tas Import Sukses dengan menggunakan metode Naïve Bayes, untuk mengetahui tingkat kepuasan konsumen terhadap kualitas produk tas. Dan yang terakhir adalah mengetahui tingkat akurasi dalam menentukan kepuasan dan kualitas produk dengan metode Naïve Bayes. Sedangkan hasil penelitian ini adalah kepuasan sangat dipengaruhi oleh tingkat pelayanan, sedangkan urutan kedua adalah kualitas produk, dimana rata-rata tingkat akurasi kepuasan menggunakan algoritma naïve bayes adalah 91%, precision 89% dan recall 88,33%.

Kata Kunci : Naïve Bayes, Akurasi, Kepuasan, Kualitas

ABSTRACT

Current technological developments have brought changes to human lifestyles, such as lifestyle, selection of antic goods and communication equipment such as smartphones and other devices. Nowadays it is known that the collection of branded items is also a lifestyle trend such as bags, rings, clothes and other fashions. Batam as an industrial city encourages investors to invest in Batam, because Batam is the most strategic city and is an international trade route directly adjacent to Singapore. The problem that often occurs at PT Tas Import Sukses in serving consumers is that the goods sent are not on demand, so consumers suspect that the products traded do not have good quality. And if there is a problem, the company responds with a very slow time. The purpose of this study is to analyze consumer satisfaction with the quality of products at PT Tas Import Sukses using the Naïve Bayes method, to determine the level of consumer satisfaction with the quality of bag products. And the last is to find out the level of accuracy in determining product satisfaction and quality with the Naïve Bayes method. While the results of this study are that satisfaction is strongly influenced by the level of service, while the second order is product quality, where the average level of satisfaction accuracy using the naïve bayes algorithm is 91%, precison 89% and recall 88.33%.

Keyword: *Satisfaction, Naïve Bayes, Quality, Service*

KATA PENGANTAR

Segala puji dan syukur kehadirat Allah SWT yang telah melimpahkan segala rahmat dan karuniaNya, sehingga penulis dapat menyelesaikan laporan tugas akhir yang merupakan salah satu persyaratan untuk menyelesaikan program studi strata satu (S1) pada Program Studi Teknik Informatika Universitas Putera Batam.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Karena itu, kritik dan saran akan senantiasa penulis terima dengan senang hati. Dengan segala keterbatasan, penulis menyadari pula bahwa skripsi ini takkan terwujud tanpa bantuan, bimbingan, dan dorongan dari berbagai pihak. Untuk itu, dengan segala kerendahan hati, penulis menyampaikan ucapan terima kasih kepada:

1. Rektor Universitas Putera Batam.
2. Dekan Fakultas Teknik dan Komputer Universitas Putera Batam.
3. Ketua Program Studi Teknik Informatika Universitas Putera Batam.
4. Bapak Rahmat Fauzi, S.Kom., M.Kom selaku dosen pembimbing Skripsi di Universitas Putera Batam.
5. Dosen dan Staff Universitas Putera Batam.
6. Orang tua penulis, serta anggota keluarga yang terus berdoa dan mendukung penulis untuk menyelesaikan tugas akhir skripsi ini
7. PT Tas Import Sukses yang sudah memberikan izin peneliti untuk mengumpulkan data dan melakukan penelitian.
8. Bella karisma sebagai sahabat terbaik yang selalu memberi dukungan selama proses penyelesaian skripsi.
9. Teman kampus yang terus mendukung penulis dalam menyelesaikan skripsi ini

Semoga Allah SWT membalas kebaikan dan selalu mencurahkan taufik serta hidayah Nya. Amin.

Batam, 22 Agustus 2023



Ahmad Suwandi

DAFTAR ISI

HALAMAN SAMPUL	
HALAMAN JUDUL	i
SURAT PERNYATAAN	ii
HALAMAN PENGESAHAN	iii
ABSTRAK	iv
ABSTRACT	v
KATA PENGANTAR	vi
DAFTAR ISI	vii
DAFTAR GAMBAR	ix
DAFTAR TABEL	x
DAFTAR RUMUS	xi
BAB 1 PENDAHULUAN	
1.1 Latar Belakang	1
1.2 Identifikasi Masalah	4
1.3 Batasan Masalah	5
1.4 Rrumusan Masalah	5
1.5 Tujuan Penelitian	6
1.6 Manfaat Peneliti	6
1.6.1 Aspek Teoritis	6
1.6.2 Aspek Praktis	6
BAB II KAJIAN PUSTAKA	
2.1 <i>Knowledge Discovery in Database (KDD)</i>	8
2.2 <i>Machine Learning</i>	9
2.3 Data Mining	10
2.3.1 Fungsi Data Mining	11
2.3.2 Tahapan Data Mining	12
2.4 <i>Data Mining Methods</i>	13
2.4.1 Algoritma <i>C4.5</i>	15
2.4.2 Algoritma <i>K-Means</i>	15
2.4.3 Algoritma <i>K-Nearest Neighbor (K-NN)</i>	16
2.4.4 <i>Naïve Bayes Classifier</i>	16
2.5 <i>Software Pendukung</i>	21
2.5.1 <i>Rapidminer</i>	21
2.5.2 <i>WEKA</i>	22
2.5.3 <i>Microsoft Excel</i>	23
2.6 Kepuasan Konsumen	24
2.7 Penelitian Terdahulu	24
BAB 3 METODE PENELITIAN	
3.1 Desain Penelitian	29
3.2 Teknik Pengumpulan Data	30
3.2.1 Kuesioner	30
3.2.2 Wawancara	31
3.2.3 Studi Literatur	31
3.3 Operasional Variabel	31

3.4	Algoritma <i>Naïve Bayes</i>	33
3.4.1	Data Penelitian	33
3.4.2	Data Trainig dan Testing.....	35
3.4.3	Rumus Pencarian Tingkat Akurasi.....	35
3.4.4	Tool Analisis.....	37
3.5	Lokasi dan Jadwal Penelitian.....	37
BAB IV HASIL DAN PEMBAHASAN		
4.1	Hasil Penelitain	38
4.1.1	Dataset	38
4.1.2	Hasil Pengujian dengan <i>Confusion Matrix</i>	41
4.1.3	Hasil Pengujian Model dengan <i>Rapid Miner</i>	41
4.1.4	Pembuatan Model dengan <i>Weka</i>	50
4.1.5	<i>Feature Selection</i> Penentu Kepuasan Konsumen.....	54
4.2	Hasil Penelitian	55
4.2.1	Menganalisis kepuasan konsumen terhadap kualitas produk.....	55
4.2.2	Tingkat kepuasan konsumen terhadap kualitas produk tas	55
4.2.3	Tingkat Akurasi Dengan Metode <i>Naïve Bayes</i>	56
BAB V KESIMPULAN DAN SARAN		
5.1	Kesimpulan.....	57
5.2	Saran	57
DAFTAR PUSTAKA		
LAMPIRAN		
Lampiran 1. Pendukung Penelitian		
Lampiran 2. Daftar Riwayat Hidup		
Lampiran 3. Surat Keterangan Penelitian		

DAFTAR GAMBAR

Gambar 2.1 Tahapan Proses dalam KDD.....	8
Gambar 2.2 Metode Data Mining.....	13
Gambar 2.3 <i>Flowchart</i> Metode <i>Naïve Bayes</i>	19
Gambar 2.4 Logo <i>Rapidminer</i>	21
Gambar 2.5 Tampilan Menu <i>Rapidminer</i>	21
Gambar 2.6 Tampilan Menu <i>Weka</i>	22
Gambar 2.7 Logo <i>Microsoft Excel</i>	23
Gambar 2.8 Tampilan Menu <i>Ms Excel</i>	24
Gambar 2.9 Kerangka Pemikiran	28
Gambar 3.1 Alur Penelitian	29
Gambar 3.2 Algoritma <i>Naïve bayes</i>	32
Gambar 4.1 Menu Awal <i>RapidMiner</i>	45
Gambar 4.2 Import Dataset	46
Gambar 4.3 Type Dataset <i>RapidMiner</i>	46
Gambar 4.4 Dataset Setelah di Import	47
Gambar 4.5 Model <i>Training</i> dan <i>Testing</i>	47
Gambar 4.6 Model <i>Naïve Bayes</i>	48
Gambar 4.7 Persentasi <i>Dataset</i>	49
Gambar 4.8 Nilai AUC Kepuasan <i>Costumer</i>	49
Gambar 4.9 Tampilan <i>Weka</i>	50
Gambar 4.10 <i>Pre-Processing</i>	51
Gambar 4.11 <i>Import Dataset</i>	51
Gambar 4.12 <i>Cross Validation</i>	52
Gambar 4.13 Pemilihan Model <i>Naïve Bayes</i>	52
Gambar 4.14 Proses <i>Cross Validation</i> untuk <i>Training</i> dan <i>Testing</i>	53
Gambar 4.15 Hasil Analisis Model <i>Naïve Bayes</i>	53
Gambar 4.16 Model <i>Feature Selection</i> dengan <i>Weight Correlation</i>	54

DAFTAR TABEL

Tabel 3.1 Indikator Penelitian	30
Tabel 3.2 Sampel Dataset	34
Tabel 3.3 Skala dan Bobot	34
Tabel 3.4 <i>Confusion Matrix</i>	36
Tabel 3.5 Variabel Penelitian	37
Tabel 3.6 Jadwal Penelitian	37
Tabel 4.1 Dataset Hasil Kuesioner	38
Tabel 4.2 Atribut <i>Class</i>	41
Tabel 4.3 Atribut Jenis Kelamin.....	41
Tabel 4.4 Atribut Pendidikan	42
Tabel 4.5 Atribut Harga.....	42
Tabel 4.6 Atribut Pelayanan	42
Tabel 4.7 Atribut Kualitas	43
Tabel 4.8 Atribut Jaminan	43
Tabel 4.9 Atribut Respon	43
Tabel 4.10 <i>Confusion Matrix</i>	44
Tabel 4.11 Tingkat <i>Precision</i>	45
Tabel 4.12 Tingkat <i>Recall</i>	45
Tabel 4.13 Perbandingan Aplikasi	54
Tabel 4.14 Nilai <i>Weight by Correlation</i>	55

DAFTAR RUMUS

Rumus 2.1 Formula Teorema <i>Bayes</i>	16
Rumus 2.2 Rumus <i>Condition Probability</i>	17
Rumus 2.3 Rumus Teorema <i>Bayes</i>	17
Rumus 2.4 Rumus Densitas <i>Gauss</i>	18
Rumus 2.5 Rumus Menghitung Nilai <i>Mean</i>	20
Rumus 2.6 Rumus Menghitung Standar Deviasi	20
Rumus 3.1 Rumus <i>Confusion Matrix</i>	36
Rumus 4.1 Rumus <i>Accuracy</i>	44
Rumus 4.2 Rumus <i>Precision</i>	44
Rumus 4.3 Rumus <i>Recall</i>	45

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi saat ini telah membawa perubahan akan gaya hidup manusia, seperti gaya hidup, pemilihan barang antik dan peralatan komunikasi seperti *smartphone* dan perangkat lainnya. Saat ini telah diketahui bahwa koleksi barang-barang bermerek juga menjadi tren gaya hidup seperti tas, cincin, baju, dan fassion lainnya. Batam sebagai kota industri mendorong para inverstor untuk menanamkan modal di Batam, karena Batam merupakan kota yang paling strategis dan merupakan jalur perdagangan internasional yang berbatasan langsung dengan negara singapura. Banyak perusahaan industri di Batam yang menjalankan bisnisnya salah satunya PT Tas Import Sukses. Perusahaan ini bergerak dibidang supplier tas import yang telah ada di Batam sejak tahun 2010 yang belokasi di kawasan industri citra buana center park 1 blok k no 3 seraya, Batam. Dengan berkembangnya perusahaan ini, maka perusahaan juga memiliki website resmi yang dapat diakses di www.tas.id.

Website ini menjadi portal resmi bagi konsumen yang ingin membeli tas bermerek seperti JT81800-coffee Tas Handbag Fashion Wanita Elegan Import Terbaru. Harga-harga yang ditawarkan sangat bervariasi, mulai dari harga yang kualitas sedang sampai harga yang kualitas tinggi. Karna tingginya peminat konsumen akan pengguna tas bermerek maka perusahaan tersebut harus memperhatikan kepuasan dari konsumen, terutama produk tas yang ditawarkan.

Karena persaingan yang semakin sengit dan banyaknya pesaing, perusahaan harus terus memperhatikan kebutuhan dan keinginan konsumen dan berusaha memenuhi harapan konsumen dengan memberikan pelayanan yang lebih memuaskan dari pada yang dilakukan oleh pesaing mereka. Agar perusahaan dapat unggul atau bahkan bertahan, perusahaan memerlukan filosofi baru. Hanya perubahan yang dilakukan oleh perusahaan akan memungkinkan mereka bersaing dan menguasai pasar. Kepuasan konsumen adalah tingkat perasaan seseorang (konsumen) setelah membandingkan kinerja atau hasil yang dirasakan dengan yang diharapkan. Kepuasan konsumen didefinisikan sebagai evaluasi purna beli, di mana kinerja produk atau jasa yang dipilih memenuhi atau melebihi harapan (Amanda, Windarto, and Qurniawan 2022).

Permasalahan yang sering terjadi di PT Tas Import Sukses dalam melayani konsumen yaitu produk barang yang dikirimkan tidak sesuai permintaan, sehingga konsumen mencurigai produk yang diperjual belikan tidak memiliki kualitas yang baik. Dan jika ada permasalahan pihak perusahaan merespon dengan waktu yang lambat. Selain itu jaminan terhadap tas atau produk yang diperjual belikan tidak memiliki jaminan yang berbentuk tertulis dan hanya berupa informasi dari website. Dengan permasalahan seperti ini, jika dibiarkan terus menerus terjadi maka dikhawatirkan pelanggan atau konsumen akan beralih ke tempat lain, sehingga mengakibatkan penjualan menurun. Semua ini diawali dengan ketidakpuasan terhadap layanan, harga, kualitas produk tas, Jaminan dan respon keluhan yang menghabiskan waktu yang relatif lama. Untuk mengetahui secara pasti berkaitan dengan kepuasan konsumen maka perlu dilakukan analisis kepada beberapa

konsumen yang sering membeli produk tas. Dengan tujuan untuk melakukan evaluasi terhadap layanan saat ini. Analisis kepuasan menggunakan beberapa metode seperti *Naïve bayes*, *KNN*, *Neural network* dan *SVM*. Akan tetapi pada penelitian ini menggunakan algoritma *naïve bayes*.

Metode *Naive Bayes* diciptakan oleh ilmuwan Inggris Thomas Bayes dan digunakan untuk menentukan tingkat kepuasan konsumen adalah salah satu algoritma pembelajaran induktif yang paling efektif dan efisien untuk pengolahan data dan pembelajaran mesin. Penulis di masa mendatang akan menggunakan metode *Naive Bayes* untuk menentukan tingkat kepuasan konsumen (Syarli and Muin 2016). Metode *Naive Bayes*, yang dikembangkan oleh ilmuwan Inggris Thomas Bayes, digunakan untuk mengklasifikasikan data besar dengan teknik probabilitas dan statistik. Penulis di masa mendatang akan menggunakan metode ini untuk menentukan tingkat kepuasan konsumen. Metode ini memiliki tingkat akurasi yang sangat tinggi ketika diterapkan pada database yang memiliki banyak data (Bambang Hermanto 2019).

Beberapa penelitian yang telah dilakukan terkait dengan analisis kepuasan konsumen seperti penelitian yang dilakukan oleh (Ama et al. 2022), penelitian ini telah menemukan bahwa suatu perusahaan harus dapat menjaga hubungan baik dengan pelanggan agar tidak terjadi sentiment antara pemilik usaha dengan pelanggan atau customer. Kualitas pelayanan yang baik dapat meningkatkan mutu layanan dan image produk yang ditawarkan perusahaan. Ini terjadi di perusahaan Tokopedia yang selalu memberikan servis yang memuaskan kepada mitra seperti UKM yang berjasa terhadap produk yang dijual di tokopedia. Metode yang

digunakan pada penelitian tersebut menggunakan naïve bayes dengan pembobotan TF-IDF yang menghasilkan tingkat akurasi yang baik dalam melakukan klasifikasi. Selain itu penelitian yang dilakukan oleh (Amanda et al. 2022) suatu perusahaan yang bergerak di bidang jasa juga melakukan penelitian berkaitan dengan kepuasan menggunakan pendekatan *naïve bayes*. Karna saat ini pesaing terhadap jasa perawatan kecantikan semakin banyak, sehingga diperlukan kualitas layanan agar dapat bersaing secara terus menerus. Penelitian ini menggunakan pendekatan data mining dengan mengumpulkan data dari kostumer dan menganalisisnya menggunakan algoritma *naïve bayes*, dengan jumlah data sebanyak 75 data. Dan hasilnya memperlihatkan bahwa harga, kualitas, respon dan jaminan dapat meningkatkan kepuasa konsumen dengan tingkat akurasi mencapai 88%.

Karna itu maka penelitian ini mengangkat judul “**Analisis Kepuasan Konsumen Terhadap Kualitas Produk Menggunakan Algoritma Naive Bayes**”, agar perusahaan ini dapat melakukan evaluasi terhadap layanan yang selama ini telah lakukan.

1.2 Identifikasi Masalah

Untuk mengidentifikasi permasalahan maka penulis mengetahuiberdasarkan latar belakang yang ada:

1. Pesaing yang sangat banyak di Kota batam terkait dengan produk Tas Import.
2. Banyaknya keluhan dari konsumen yang berkaitan dengan produk tas dan respon yang kurang cepat.

3. Tempat yang kurang strategis dan akses kendaraan umum tidak bisa di jangkau.
4. Pengantaran barang yang terlambat dan tidak sesuai pesanan.
5. Sistem *website* yang dimiliki kadang tidak dapat diakses sehingga kosumen protes.

1.3 Batasan Masalah

Untuk menghindari pembahasan yang lebih luas terkait dengan analisis kepuasan konsumen terhadap kualitas produk menggunakan algoritma *naive bayes* yang ada di PT Tas Import Sukses, penulis membatasi masalah sebagai berikut:

1. Penelitian ini hanya membahas kepuasan konsumen yang ada di PT Tas Import Sukses Batam.
2. Penelitian ini menggunakan Algoritma *Naïve Bayes* dalam melakukan analisis.
3. Hanya menggunakan 5 variabel yaitu harga, pelayanan, kualitas, Jaminan, Respon.
4. Kriteria konsumen dijabarkan menjadi dua kategori yaitu puas dan tidak puas.

1.4 Rumusan Masalah

Agar meningkatkan dan memperbaiki kinerja pelayanan maka peneliti menemukan permasalahan yang dapat disesuaikan dengan tema yang diambil sebagai bahan skripsi ini adalah:

- 1 Bagaimana menganalisis kepuasan konsumen terhadap kualitas produk yang adai PT Tas Import Sukses menggunakan metode *Naïve Bayes*
- 2 Bagaimana tingkat kepuasan konsumen terhadap kualitas produk tas

- 3 Bagaimana tingkat akurasi dalam menentukan kepuasan dan kualitas produk dengan metode *Naïve Bayes*.

1.5 Tujuan Penelitian

Setelah ditentukan masalah dalam penelitian ini tentunya peneliti harus menentukan tujuan dari penelitian ini guna untuk mengarahkan dan memperjelas penelitian ini yang dimana penelitian ini bertujuan sebagai berikut:

1. Untuk menganalisis kepuasan konsumen terhadap kualitas produk yang ada di PT Tas Import Sukses menggunakan metode *Naïve Bayes*.
2. Untuk mengetahui tingkat kepuasan konsumen terhadap kualitas produk tas.
3. Untuk mengetahui tingkat akurasi dalam menentukan kepuasan dan kualitas produk dengan metode *Naïve Bayes*.

1.6 Manfaat Penelitian

Adapun manfaat penelitian ini dibagi menjadi dua bagian seperti yang dijelaskan berikut ini:

1.6.1 Aspek Teoritis

Berikut uraian secara teoritis dari penelitian ini:

1. Bagi mahasiswa menambah wawasan dan menjadikan penelitian ini sebagai referensi pada penelitian lainnya.
2. Bagi pembaca dapat menambah pengetahuan dan dapat dijadikan bahan acuan untuk melakukan penelitian yang berkaitan dengan kepuasan konsumen terhadap kualitas produk dengan metode *naïve bayes*.

1.6.2 Aspek Praktis

Penelitian ini memiliki manfaat praktis yaitu sebagai berikut:

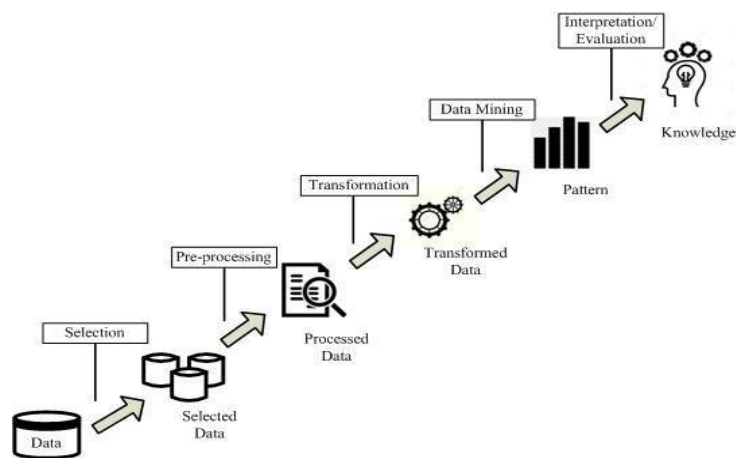
1. Bagi perusahaan penelitian ini dapat mengukur kepuasan konsumen dan dapat membantu perusahaan menawarkan produk secara lebih baik, efektif dan dapat langsung diimplementasikan dan di kembangkan sesuai kebutuhan perusahaan.
2. Bagi peneliti implementasi metode *naïve bayes* dalam melakukan analisis menggunakan pendekatan data *mining* dan *machine learning* dengan algoritma *naïve bayes* dan dapat menambah pengetahuan seta dapat mengembangkan penelitian selanjutnya.

BAB II

KAJIAN PUSTAKA

2.1 *Knowledge Discovery in Database (KDD)*

Proses penemuan pengetahuan dalam *database* (KDD) adalah proses mengidentifikasi pola data. KDD terdiri dari lima tahapan, yang mencakup integrasi, interpretasi, dan visualisasi berbagai pola data sehingga data menjadi mudah dimengerti oleh pengguna. Tahap pertama dari proses ini adalah pemrosesan data *mining*.



Gambar 2.1 Tahapan Proses dalam KDD

Tahapan proses *Knowledge discovery in database* (KDD) sebagai berikut :

1. *Data Selection*

Pada tahap ini, data target yang akan digunakan untuk tahap berikutnya dibuat dengan memilih himpunan data dari sekumpulan data operasional

Himpunan data ini disimpan dalam satu berkas yang berbeda dari data operasional.

2. *Pre-Processing/Cleansing*

Pada tahap ini, data yang telah dipilih akan dibersihkan dari *noise* dan data yang tidak konsisten untuk mencegah duplikasi dan memperbaiki kesalahan. Jika ini tidak dilakukan, pengolahan data akan menjadi kurang akurat.

3. *Transformation*

Proses peringkasan, atau agregasi, adalah proses mengubah atau menggabungkan data ke tempat yang lebih sesuai untuk proses *mining*.

4. *Data Mining*

Pada titik ini, proses pencarian pola atau informasi bermanfaat dalam data yang telah dipilih dilakukan dengan berbagai metode dan pendekatan yang disesuaikan dengan tujuan proses KDD yang dilakukan.

5. *Interpretation/Evaluation*

Dalam Langkah terakhir KDD, informasi yang diperoleh dari proses data mining akan diterjemahkan ke dalam bentuk yang mudah dipahami oleh pengguna. Kemudian, mereka akan mengevaluasi apakah temuan tersebut sejalan atau bertentangan dengan hipotesa atau fakta sebelumnya.

2.2 *Machine Learning*

Machine learning merupakan machine yang mampu belajar sendiri dengan berbagai pengalaman yang dipelajarinya. Suatu komputer dikatakan belajar dari pengalaman yang berhubungan dengan beberapa tugas yang diberikan dan ukuran

performance jika kinerjanya pada tugas yang diberikan sehingga dapat meningkatkan pengalaman.

Berdasarkan dampak dari *machine learning*, algoritma *machine learning* dikelompokkan menjadi enam yaitu *Supervised learning*, *Unsupervised learning*, *Semi supervised learning*, *reinforment learning*, *transduction learning* dan *learning to learn* (Suyanto 2018). Sedangkan algoritma dapat dikelompokkan kedalam dua kategori yaitu diskrit dan kontinu.

2.3 Data Mining

Data mining adalah serangkaian proses untuk mendapatkan nilai tambahan dari kumpulan data menggunakan pengetahuan yang tidak diketahui sebelumnya untuk membantu pengambilan keputusan dan menemukan pola penting dalam data. (Aprizal, Hasriani, and Ningsih 2016).

Selain itu *machine learning* juga memiliki berbagai algoritma seperti *naïve bayes*, *KNN*, *SVM*, *Neural network* dan *Random Forest*. Dan semua algoritma ini dapat digunakan untuk menganalisis data yang biasa digunakan pada *data mining* (Rahman and Sutanto 2023).

Data mining adalah teknik pengolahan data yang menggunakan analisis pengetahuan otomatis untuk menghasilkan informasi yang bermanfaat bagi organisasi atau perusahaan. (Sadewo, Windarto, and Damanik 2019).

Dalam penerapannya *data mining* bisa di implementasikan pada berbagai macam bidang seperti perbankan, asuransi, olahraga, penjualan. Salah satu contoh dari penerapan *data mining* pada bidang penjualan ada pada penelitian yang

dilakukan oleh (Sitohang, 2021) yang menggunakan *data mining* untuk menentukan klasifikasi laptop yang ingin dibeli berdasarkan keinginan konsumen.

Berdasarkan teori di atas dapat di tarik kesimpulan bahwa *data mining* merupakan suatu teknik untuk memproses data yang tersimpan pada *database* dengan tujuan bisa ditemukan informasi dan pengetahuan baru yang belum diketahui sebelumnya.

2.3.1 Fungsi Data Mining

Santosa (Budi Santosa & Ardian uman 2018) menjelaskan bahwa ada tiga fungsi umum dalam pengolahan data, diantaranya adalah sebagai berikut:

1. *Assosiasi*, proses menentukan hubungan antara kumpulan item tertentu dalam suatu waktu tertentu;
2. *Sequence*, proses untuk menemukan aturan yang mengatur kombinasi item yang terjadi dalam lebih dari satu periode waktu.
3. *Clustering*, yang merupakan proses pengelompokan sejumlah data atau obyek ke dalam kelompok data sehingga setiap kelompok mengandung data yang sebanding.
4. Proses penemuan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data dengan tujuan untuk mengidentifikasi kelas objek yang labelnya tidak diketahui dikenal sebagai klasifikasi.
5. Pemetaan data dalam nilai prediksi dikenal sebagai regresi.
6. *Forecasting* adalah proses mengestimasi nilai prediksi berdasarkan pola-pola dalam sekumpulan data.

7. Penyelesaian atau solusi adalah proses menemukan akar masalah dan menyelesaikan masalah tersebut, atau paling tidak sebagai informasi untuk pengambilan keputusan.

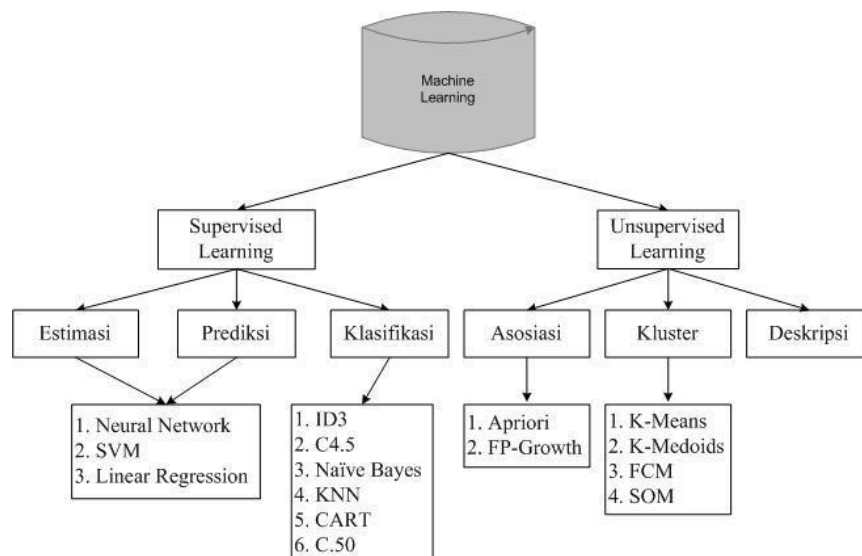
2.3.2 Tahapan *Data Mining*

(Saleh 2015), menjelaskan bahwa *data mining* dibagi menjadi beberapa langkah yaitu:

1. *Pembersihan data (Data Cleaning)*, adalah proses menghilangkan *noise* dari data yang tidak konsisten dan tidak relevan
2. *Data integration*, yaitu penggabungan data dari beberapa database.
3. *Pemilihan Data (Data Selection)*, yaitu proses pengambilan data yang sesuai dari *database* untuk diproses.
4. *Data transformation*, data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam *data mining*.
5. Proses *mining* adalah proses utama ketika teknik digunakan untuk menemukan pengetahuan penting dan tersembunyi dari data dengan menggunakan teknik yang tepat.
6. *Pattern evaluation* proses mengidentifikasi pola-pola menarik ke dalam *knowledge based* yang ditemukan.
7. *Knowledge presentation*, yaitu proses visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk mendapatkan hasil yang diperoleh.

2.4 Data Mining Methods

Penelitian yang dilakukan oleh (Sardiarinto, 2013), metode *data mining* dikelompokkan menjadi dua bagian utama yaitu *supervised model* dan *unsupervised model*.



Gambar 2.2 Metode Data Mining

Sumber: (Sardiarinto, 2018)

1. *Supervised Model*

Pemodelan yang diarahkan dengan tujuan memprediksi kejadian dan memperkirakan nilai atribut angka yang terjadi secara terus menerus dikenal sebagai model supervised. Yang termasuk kedalam metode ini adalah:

a. Klasifikasi

Tujuan dari klasifikasi adalah untuk mengelompokkan hubungan antara variabel kriteria dan variabel target dalam kelompok tertentu. Kelompok atau kelas yang telah diketahui sebelumnya digunakan dalam proses klasifikasi. Contoh algoritma *ID3*, *C4.5*, *Naïve Bayes*.

b. Prediksi

Salah satu model yang paling umum digunakan dalam data mining adalah pembagian prediksi. Prediksi, yang umumnya hampir sama dengan pengelompokan klasifikasi, adalah hasil dari memprediksi nilai output yang akan digunakan di masa depan. Contoh algoritma *Neural Network*, *Linear Regression*, *Support Vector Machine*.

c. Estimasi

Estimasi dikelompokkan seperti klasifikasi; estimasi adalah perkiraan atau prediksi; satu-satunya hal yang membedakan keduanya adalah bagaimana mereka dikelompokkan. Tidak seperti kategori, mengelompokkan berdasarkan estimasi cenderung menggunakan angka. Contoh algoritma *Neural Network*, *Linear Regression*, *Support Vector Machine*.

2. *Unsupervised Model*

Model tidak terarah adalah pemodelan yang terarah, tetapi atribut tidak dipandu oleh target tertentu; bidang input hanya ada dan bidang output tidak ada. Metode yang termasuk model ini, yaitu:

a. Model asosiasi

Asosiasi dapat berupa kelompok, himpunan, gabungan, atau perserikatan. Kemunculan dua atau lebih variabel secara bersamaan adalah proses pengelompokan asosiasi. Nilai kepercayaan dapat digunakan untuk menghitung besarnya kemunculan peluang yang bersamaan pada atribut. Contoh algoritma *Fp-Growth*, *Apriori*.

b. Model *cluster*

Jenis data yang dihasilkan melalui pengklasteran, seperti perolehan pengamatan, perekaman data, dan objek, memiliki nilai yang sama. Contoh algoritma *K-means*, *Fuzzy C-means*, *K-Medoids*, *Self Organization Map (SOM)*.

c. Model deskripsi

Tujuan dari proses deskripsi adalah untuk menemukan pola yang sering terjadi dan kemudian mengubah pola tersebut menjadi aturan baru yang bermanfaat untuk membuat aktivitas lebih mudah.

2.4.1 Algoritma C4.5

Algoritma C4.5 memiliki beberapa pengembangan, seperti mengatasi data terus-menerus, pruning, dan mengatasi nilai yang tidak ada. Ini adalah teknik dari pohon keputusan atau pohon keputusan yang dapat digunakan untuk menghasilkan beberapa aturan. Pohon keputusan, juga dikenal sebagai pohon keputusan, dibuat dengan tujuan untuk meningkatkan nilai prediksi dan keakuratan data yang dilakukan. Ini adalah algoritma C4.5 dari algoritma klasifikasi.

2.4.2 Algoritma *K-Means*

Algoritma ini membagi data menjadi dua atau lebih kelompok sehingga data dengan karakteristik yang sama dapat dimasukkan ke dalam satu kelompok dan data dengan karakteristik yang berbeda dapat dikelompokkan dengan kelompok lain. Tujuan dari pengelompokan data ini adalah untuk mengurangi variasi antar kelompok dan fungsi objek yang diatur.

2.4.3 Algoritma *K-Nearest Neighbor (K-NN)*

Sedangkan (Harrington, 2019) menjelaskan bahwa algoritma K-NN memiliki beberapa kelebihan, termasuk akurasi tinggi, insentif terhadap outlier, dan tidak ada dugaan terhadap data. Namun, algoritma K-NN juga membutuhkan banyak memori, biaya komputasi yang tinggi, dan menentukan nilai k yang ideal.

2.4.4 *Naïve Bayes Classifier*

Untuk menghitung sekumpulan probabilitas, *Naive Bayes* menjumlahkan frekuensi dan nilai dari kumpulan data yang diberikan. (Saleh 2015). Ilmuan Inggris Thomas Bayes pertama kali membuat *Naive Bayes*, yang berfungsi untuk menggunakan pengalaman masa lalu untuk memprediksi peluang di masa depan. Teknik prediksinya berbasis probabilistik sederhana dan berbasis pada penerapan teorema Bayes (juga dikenal sebagai aturan Bayes) dengan asumsi penyederhanaan bahwa jika nilai output diberikan, nilai atribut secara kondisional saling bebas (independen). Dalam konteks ini, konsep independensi berarti bahwa keberadaan atau tidaknya fitur tertentu dalam data tidak terkait dengan keberadaan atau tidaknya fitur lain dalam data yang sama

Persamaan dari teorema *bayes* adalah:

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

Rumus 2.1 Formula
Teorema *Bayes*

Keterangan :

X : Data yang memiliki class yang tidak diketahui.

H : Hipotesis data X merupakan suatu class spesifik.

P(H|X): Probabilitas hipotesis H berdasar kondisi X (*posteriori probability*).

P(H) : Probabilitas hipotesis H (*prior probability*).

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

$H P(X)$: Probabilitas X.

Untuk menjelaskan metode *naive bayes*, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah rekomendasi untuk menentukan kelas yang tepat untuk sampel yang dianalisis. Oleh karena itu, metode *naive bayes* yang disebutkan di atas dimodifikasi sebagai berikut:

$$P(C|F1..Fn) = \frac{P(C)P(F1..Fn|C)}{P(F1..Fn)} \quad \textbf{Rumus 2.2 Rumus Condition Probability}$$

Variabel C menunjukkan kelas, dan variabel F1 hingga Fn menunjukkan karakteristik petunjuk yang diperlukan untuk melakukan klasifikasi. Rumus tersebut menjelaskan bahwa kemungkinan masuknya sampel karakteristik tertentu dalam kelas C (posterior) adalah kemungkinan munculnya kelas C (sebelum masuknya sampel, biasanya disebut prior), dikali dengan kemungkinan kemunculan kelas C (juga disebut kemungkinan), dan dibagi dengan kemungkinan kemunculan kelas C secara keseluruhan (juga disebut bukti). Karena itu, rumus di atas dapat pula ditulis secara sederhana sebagai berikut:

$$\text{Posterior} = \frac{\text{Prior} \times \text{Likelihood}}{\text{Evidence}} \quad \textbf{Rumus 2.3 Rumus Teorema Bayes}$$

Nilai bukti sampel untuk setiap kelas selalu tetap. Nilai posterior sampel ini kemudian dibandingkan dengan nilai posterior kelas lain untuk menentukan kelas mana yang paling sesuai untuk diklasifikasikan. Rumus Bayes dibahas lebih lanjut dengan menjabarkan $(C|F1, \dots, Fn)$ menggunakan aturan perkalian sebagai berikut:

$$\begin{aligned}
P(C|F_1, \dots, F_n) &= P(C)P(F_1, \dots, F_n|C) \\
&= P(C)P(F_1|C)P(F_2, \dots, F_n|C, F_1) \\
&= P(C)P(F_1|C)P(F_2|C, F_1)P(F_3, \dots, F_n|C, F_1, F_2) \\
&= P(C)P(F_1|C)P(F_2|C, F_1)P(F_3|C, F_1, F_2)P(F_4, \dots, F_n|C, F_1, F_2, F_3) \\
&= P(C)P(F_1|C)P(F_2|C, F_1)P(F_3|C, F_1, F_2) \dots P(F_n|C, F_1, F_2, F_3, \dots, F_{n-1})
\end{aligned}$$

Ditunjukkan bahwa faktor-faktor syarat yang mempengaruhi nilai probabilitas menjadi semakin kompleks sebagai akibat dari hasil penjabaran tersebut. Karena faktor-faktor ini hampir tidak dapat dianalisis secara individual, perhitungan menjadi sulit. Di sini, asumsi independensi yang sangat tinggi (naif) digunakan. Ini berarti bahwa masing-masing petunjuk (F_1, F_2, \dots, F_n) saling independen (bebas) satu sama lain. Dengan asumsi tersebut, maka berlaku suatu kesamaan sebagai berikut:

$$P(F_i|F_j) = \frac{P(F_i \cap F_j)}{P(F_j)} = \frac{P(F_i)P(F_j)}{P(F_j)} = P(F_i)$$

Untuk $i \neq j$, sehingga

$$P(F_i|C, F_j) = P(F_i|C) \quad (6)$$

Dalam proses klasifikasi, persamaan di atas berfungsi sebagai representasi dari teorema *naive Bayes*. Rumus Densitas Gauss digunakan untuk klasifikasi dengan data kontinyu:

$$P(X_i=x_i|Y=y_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}}} e^{-\frac{(x_i-\mu_{ij})^2}{2\sigma_{ij}^2}} \quad \textbf{Rumus 2.4 Rumus Densitas Gauss}$$

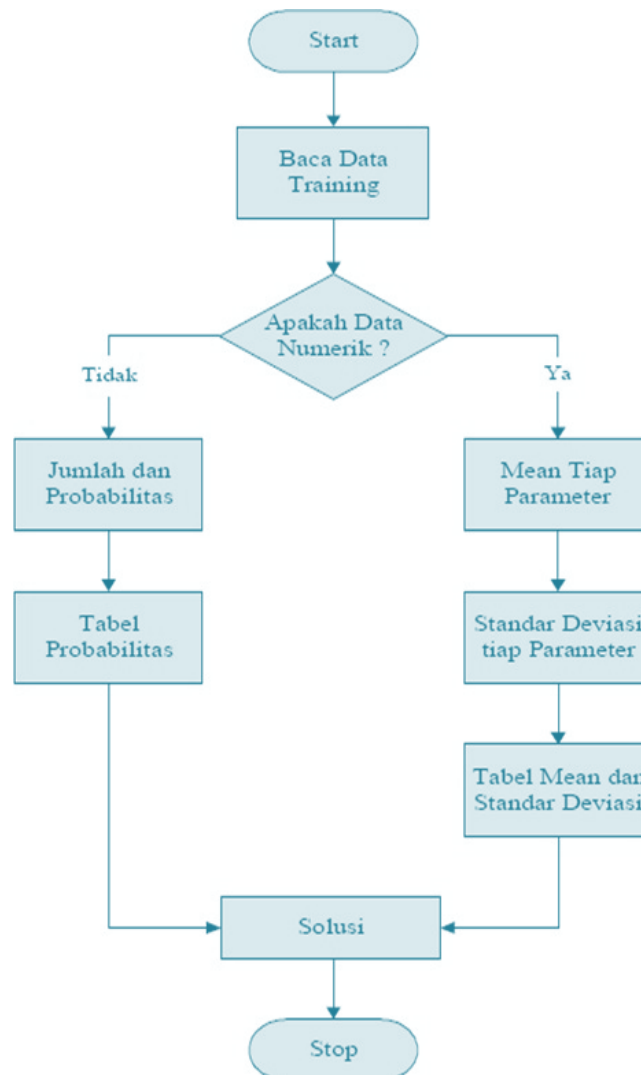
Keterangan :

P : Probability

X_i : Feature ke i

x_i : Nilai Feature ke i

- Y : Kelas yang dicari
 y_i : Sub kelas Y yang dicari
 μ : *mean*, rata – rata dari seluruh atribut
 σ : Deviasi standar merupakan varian dari seluruh atribut.



Gambar 2.3 Flowchart Metode Naive Bayes

Adapun keterangan dari gambar 2 di atas sebagai berikut:

1. Read Data *training*
2. Calculate Jumlah dan probabilitas, namun apabila data numerik maka:

a. Hitung nilai *mean* dan standar deviasi dari masing-masing parameter yang merupakan data numerik.

Berikut ini adalah persamaan yang digunakan untuk menghitung nilai rata, yang dikenal sebagai rata hitung atau *mean*.

$$\mu = \frac{\sum_{i=1}^n X_i}{n} \quad \textbf{Rumus 2.5} \text{ Rumus Menghitung} \\ \text{Nilai Mean}$$

atau

$$\mu = \frac{X_1 + X_2 + X_3 + \dots + X_n}{n}$$

information:

μ : *mean*

x_i : *sample value to -i*

n : *number of samples*

dan persamaan yang digunakan untuk menghitung nilai standar deviasi atau simpangan baku adalah sebagai berikut.

$$\rho = \sqrt{\frac{\sum_{i=1}^n (X_i - \mu)^2}{n-1}} \quad \textbf{Rumus 2.6} \text{ Rumus Menghitung} \\ \text{Standar Deviasi}$$

information:

σ : *standard deviation*

x_i : *value x to -i*

μ : *average count*

n : *number of samples*

b. Cari nilai probabilitik dengan membagi total data dari kategori yang sama dengan total data dalam kategori tersebut.

3. Mencari nilai dalam tabel probabilitas, standar deviasi, dan *mean*.

4. Kemudian Solusi dibuat.

2.5 Software Pendukung

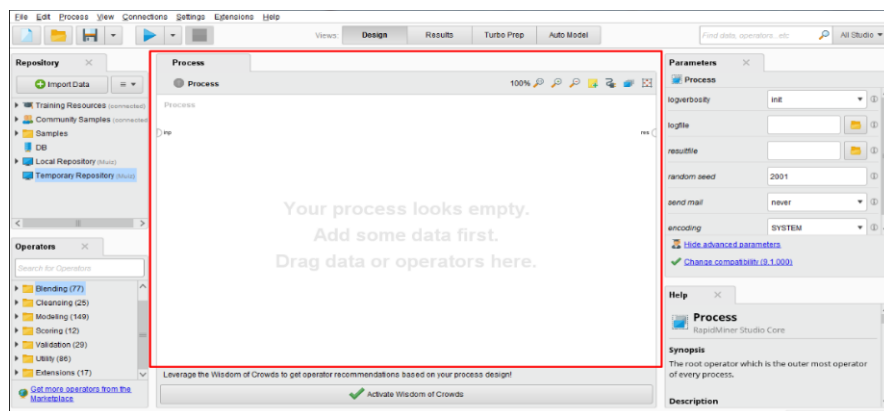
Jika data *mining* dilakukan secara manual, yang akan memakan waktu lebih lama dan kurang akurat, proses pengolahan data akan lebih mudah dan akurat jika dibantu oleh software pengolahan data yang tersedia saat ini. *Rapidminer*, *Microsoft Excel*, dan *WEKA* adalah program yang digunakan dalam penelitian ini.

2.5.1 Rapidminer



Gambar 2.4 Logo *Rapidminer*

Pada tahun 2001, Ralf Klinkenberg, Ingo Mierswa, dan Simon Fischer mengembangkan software bernama RapidMiner di Artificial Intelligence Unit Universitas Dortmund. Sebelumnya bernama YALE (Yale Another Learning Environment), perangkat lunak ini berubah nama menjadi RapidMiner pada tahun 2007.



Gambar 2.5 Tampilan Menu Rapidminer

Perusahaan dengan nama yang sama mengembangkan perangkat lunak data ilmu pengetahuan yang disebut *RapidMiner*. Platform ini menawarkan lingkungan terpadu untuk analisis prediktif, penambangan teks, pembelajaran mesin, dan pembelajaran mendalam. Aplikasinya mencakup berbagai industri, seperti bisnis dan komersial; pendidikan, pelatihan, penelitian, pembuatan *prototype*, dan pengembangan aplikasi. Program ini juga mendukung setiap langkah proses pembelajaran mesin, seperti persiapan data, visualisasi hasil, validasi, dan pengoptimalan. (Nofitri and Irawati 2019a).

2.5.2 WEKA

WEKA merupakan aplikasi untuk *data mining* dan *machine learning*. Aplikasi weka sangat bermanfaat mengolah data yang sangat besar, sehingga dapat menghasilkan hasil prediksi yang akurat pada saat menyelesaikan masalah yang sangat rumit.



Gambar 2.6 Tampilan Menu Weka

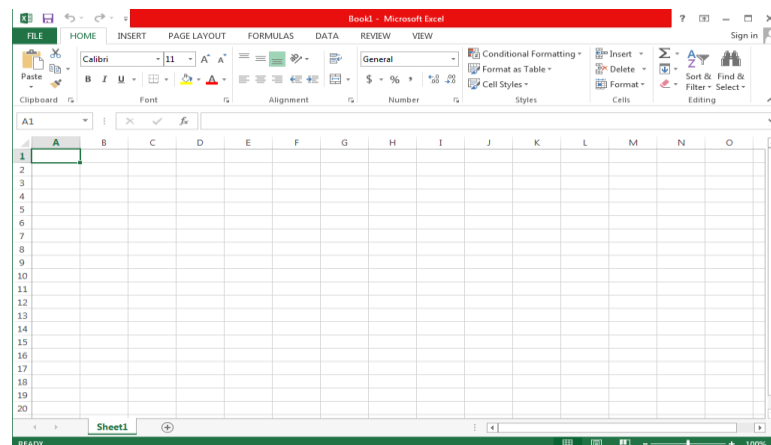
Dari gambar diatas, dapat dijelaskan bahwa aplikasi weka berasal dari New Zealand yang ditemukan oleh seorang peneliti di bidang ilmu komputer. Aplikasi ini sangat bermanfaat menyelesaikan permasalahan data yang berhubungan dengan *Machine Learning* dan *Data Mining*, dengan metode Klasifikasi dan *clustering* dengan menggunakan berbagai macam algoritma.

2.5.3 Microsoft Excel



Gambar 2.7 Logo Microsoft Excel

Microsoft Office Excel adalah program yang dibuat oleh *Microsoft Corporation* dan termasuk dalam paket *Microsoft Office* bersama dengan program lain seperti *MS Word*, *MS PowerPoint*, *MS Access*, dan *MS Outlook*. *MS Excel* adalah program lembar kerja yang populer untuk pengolahan data angka yang memiliki banyak fungsi. Milyaran sel, yang terdiri dari pertemuan baris dan kolom, tersedia untuk digunakan dalam pengolahan data (Patmawati & Santika, 2016). Selain cara pengoperasiannya yang mudah dan ramah pengguna, MS Excel telah diakui oleh banyak profesional karena kemampuan untuk menangani dan memecahkan berbagai masalah, mulai dari penggunaan yang mudah hingga pekerjaan yang rumit dengan angka dan data yang memungkinkan hasil yang optimal.



Gambar 2.8 Tampilan Menu *Ms Excel*

2.6 Kepuasan Konsumen

Ketika ekspektasi pelanggan terhadap suatu produk atau layanan terpenuhi dengan baik, bahkan melebihi harapan mereka, itu disebut kepuasan pelanggan. Menurut Permana (2013), kepuasan konsumen dipengaruhi oleh lima faktor utama: kualitas produk, kualitas pelayanan, harga, faktor emosional, dan kemudahan mendapatkan barang atau jasa tersebut. Ketika pelanggan merasa puas dengan layanan yang mereka terima, mereka akan menjadi lebih setia, yang pada gilirannya akan menghasilkan peningkatan penjualan dan pendapatan bagi perusahaan. Meskipun konsumen kadang-kadang tidak terpuaskan dengan tindakan perusahaan, perusahaan telah berusaha sebaik mungkin untuk membuat konsumen puas.

2.7 Penelitian Terdahulu

Bagian ini membahas penelitian yang berkaitan dengan penggunaan naïve bayes dalam melakukan analisis kepuasan konsumen. Beberapa penelitian yang pernah dilakukan mulai tahun 2018 sampai dengan tahun 2023 dan metode yang digunakan dapat dilihat pada ringkasan penelitian berikut ini:

1. Penelitian yang dilakukan oleh (Amanda et al. 2022) menjelaskan bahwa beberapa aspek kepuasan pelanggan menggunakan metode data mining *Naive Bayes*. Data diperoleh dari kuesioner yang diberikan secara acak kepada seratus pelanggan. Pelayanan, respons, hadiah, promosi, dan kepercayaan adalah faktor kepuasan pelanggan di Store MS Glow. Hasil yang dihasilkan oleh peneliti dan program *Rapid Miner*, yang terdiri dari 75 data instruksi dan 25 data pengujian yang diproses dalam *Rapid Miner 5.3*, menunjukkan hasil pengujian dengan akurasi sebesar 88,00%, dengan 13 pelanggan yang puas dan 12 pelanggan yang tidak puas.
2. Penelitian yang dilakukan oleh (Nasution, Munthe, and Yanris 2022) menjelaskan bahwa produk deterjen yang diproduksi perlu diperhatikan dengan menggunakan metode *naive bayes*, yang mana hasilnya “Dapat diketahui faktor yang digunakan konsumen dalam memilih produk deterjen adalah faktor pewangi, harga, busa yang dihasilkan, dan efeknya pada tangan. Pembelajaran mesin, khususnya *Naive Bayes*, akan digunakan dalam metodologi penelitian ini. Pengklasifikasi *Naive Bayes* adalah metode klasifikasi yang berakar pada teorema Bayes. Matriks kebingungan adalah tabel yang berisi banyak baris data uji yang diprediksi oleh model klasifikasi benar atau salah. Proses penambahan data dapat terdiri dari sejumlah operator bersarang yang dijelaskan dalam file XML dan dibangun menggunakan *RapidMiner*.”
3. Penelitian yang dilakukan oleh (Saritas & Yasar, 2019) menjelaskan tujuan dari penelitian ini adalah untuk memprediksi kemungkinan terkena kanker payudara pada pasien dengan menggunakan data antropometri dan parameter analisis

darah rutin yang dikumpulkan dari pasien. Algoritma *naive bayes* dan JST digunakan untuk memperkirakan diagnosis penyakit dengan menggunakan sembilan data input dan menghasilkan satu output. Dengan menggunakan data ini, mereka dapat menentukan kemungkinan pasien terkena kanker payudara.

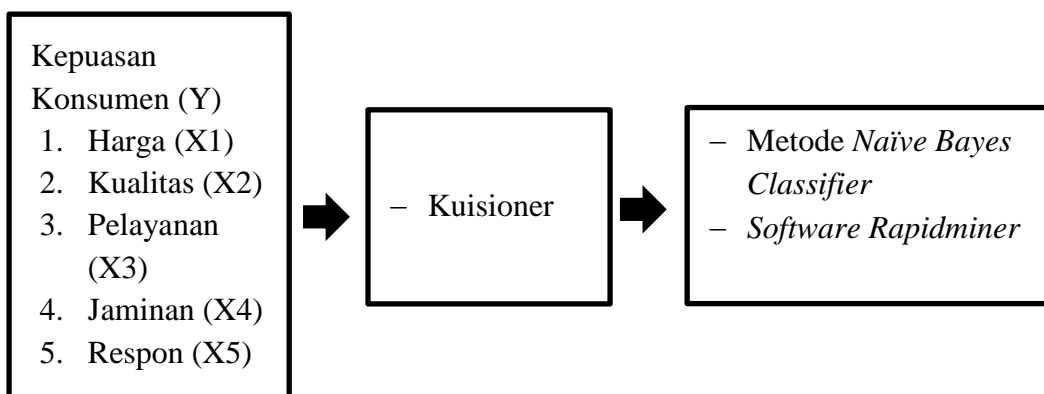
4. Menurut Penelitian (Sitohang 2021) menjelaskan tujuan dari penelitian ini adalah melakukan pengolahan data penjualan laptop di toko *Good Cell* Batam menggunakan algoritma C4.5 agar mempermudah pembeli untuk memilih tipe laptop yang akan dibeli berdasarkan kriteria yang diinginkan pembeli. Dengan menggabungkan algoritma C4.5 dengan aplikasi WEKA sebagai *platform* untuk mengujinya didapatkan hasil bahwa menggunakan data mining bisa mempercepat proses pengolahan data penjualan sehingga bisa efektif untuk meningkatkan pelayanan toko.
5. Menurut (Saleh 2015) penelitian ini bertujuan memprediksi penggunaan listrik melalui pengolahan data. Hasilnya dapat digunakan untuk merencanakan tindakan untuk mengurangi penggunaan listrik di rumah tangga. Dengan menggunakan data pelatihan, metode klasifikasi *naive Bayes* digunakan untuk menentukan kemungkinan setiap kriteria untuk kelas yang berbeda, sehingga nilai kemungkinan dari kriteria tersebut dapat dioptimalkan untuk memprediksi penggunaan listrik. Hasilnya menunjukkan bahwa metode *Naive Bayes* berhasil mengklasifikasikan 47 data dari 60 data yang diuji, yang menunjukkan bahwa metode *Naive Bayes* berhasil mengklasifikasikan 47 data dari 60 data yang diuji.

6. (Djamaludin and Nursikuwagus 2017) melakukan penelitian dengan tujuan untuk mengidentifikasi pola dalam penjualan dan pembelian suatu produk melalui transaksi penjualan. Kemudian, algoritma Apriori digunakan untuk melakukan analisis yang mendukung keputusan. Hasil penelitian menghasilkan *software* pola penjualan dan pembelian konsumen menggunakan algoritma Apriori dan hubungan aturan terakhir untuk bolu cake meses dan keju dengan kepercayaan sebesar 84,62%. Metode pengembangan sistem ini dirancang menggunakan UML dan terdiri dari *usecase diagram*, *class diagram*, dan *activity diagram*.
7. (Hermanto and Romadhoni 2019) menjelaskan agar perusahaan bengkel motor dapat bersaing di pasar, tujuan penelitian ini adalah untuk mengetahui tingkat kepuasan konsumen. Metode *naive bayes*, metode pengumpulan data dengan kuisioner, dan studi literatur digunakan. Software yang digunakan adalah *RapidMiner*. Berdasarkan penelitian yang dilakukan, hasil prediksi menentukan kepuasan konsumen dengan alat *Rapidminer* dan perhitungan manual dengan probabilitas kelas. Setelah pengujian terhadap data pelatihan dan pengujian menggunakan aplikasi *Rapidminer*, didapatkan tingkat akurasi sebesar 90%.
8. (Nofitri & Irawati, 2019) berdasarkan data penjualan, penelitian ini meneliti komponen yang mempengaruhi hasil keuntungan. Penelitian dilakukan melalui integrasi metode *naive bayes* dan alat bantu *software rapidminer*. Metode ini akan menghasilkan informasi dalam bentuk pola hubungan atau hubungan antara elemen satu dengan lainnya yang mempengaruhi tercapainya laba. Penelitian ini dapat digunakan sebagai dasar untuk memprediksi bagaimana

keuntungan perusahaan dagang akan berkembang dan membantu pelaku usaha meningkatkan keuntungan mereka.

2.8 Kerangka Pemikiran

Kerangka berpikir adalah model konseptual yang menjelaskan hubungan antara teori dengan faktor-faktor yang diidentifikasi sebagai masalah (Sugiyono, 2015).



Gambar 2.9 Kerangka Pemikiran

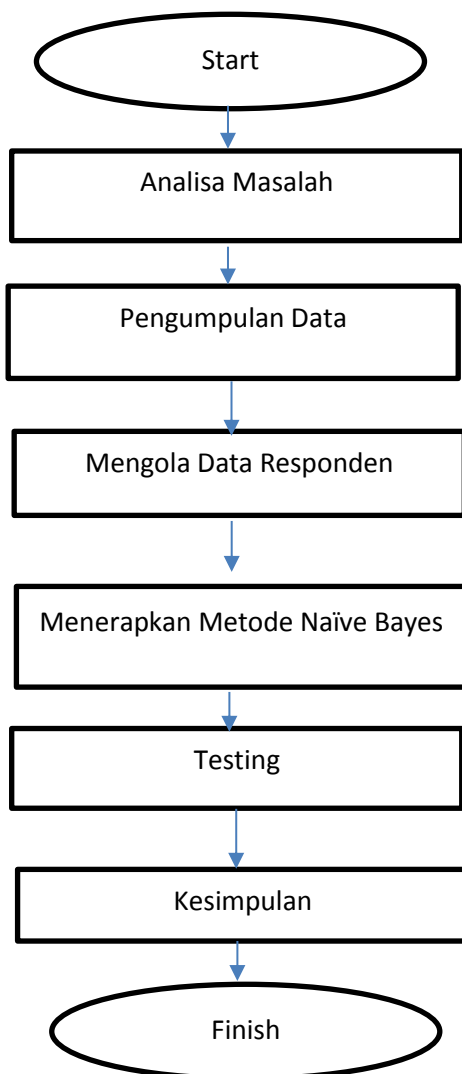
Gambar 2.9 di atas menunjukkan bahwa penelitian ini dimulai dengan tujuan mengidentifikasi kepuasan konsumen PT Tas Import Sukses dengan menggunakan algoritma klasifikasi *naïve bayes*. Ada dua variabel utama yaitu variabel dependen dan independen. Variabel dependen terdiri dari Kepuasan *customer* dan variabel independen terdiri dari harga, kualitas, pelayanan, Jaminan dan respon. Untuk mendapatkan data tersebut maka diperlukan kuisisioner yang disebarkan ke *customer* lalu di olah menggunakan aplikasi *rapid miner* dengan menerapkan algoritma *naïve bayes*.

BAB III

METODE PENELITIAN

3.1 Desain Penelitian

Untuk membuat penelitian berjalan sesuai dengan harapan peneliti, desain penelitian harus jelas, dimulai dengan identifikasi masalah dan berakhir pada kesimpulan. Berikut merupakan desain penelitian yang dilakukan.



Gambar 3.1 Alur Penelitian

Dari gambar 3.1 dijelaskan bahwa alur penelitian dimulai dari Analisa masalah yang sedang di hadapi oleh perusahaan, kemudian dilanjutkan dengan mengumpulkan data dugaan penyebab terjadinya ketidakpuasan konsumen, setelah data diperoleh maka dilanjutkan dengan melakukan analisis dengan metode *naïve bayes* dengan mendapatkan tingkat akurasi dalam menentukan kepuasan konsumen dan Langkah terakhir dari penelitian ini adalah melakukan pengujian pada metode yang di terapkan, lalu Langkah terakhir yaitu menarik kesimpulan dan sekaligus memberikan rekomendasi kepada pemilik perusahaan penyebab tertinggi terjadinya ketidak puasan konsumen terhadap produk tas yang ada di PT Tas Import Sukses Batam.

3.2 Teknik Pengumpulan Data

Teknik pengumpulan pada penelitian ini dimulai dengan menyebarkan kuesioner kepada konsumen baik secara online atau offline dan datang langsung keperusahaan yang dapat dijabarkan sebagai berikut:

3.2.1 Kuesioner

Penelitian ini menggunakan kuesioner dalam mengumpulkan data secara langsung dan mengumpulkan data historis tentang kondisi perusahaan. Indikator data yang akan dikumpulan yaitu :

Tabel 3.1 Indikator Penelitian

No	Indikator	Pernyataan
1	Harga	Harga terjangkau
2	Pelayanan	Pelayanan saat ini telah sesuai harapan
3	Kualitas	Kualitas Produk sesuai harapan
4	Jaminan	Produk yang diperjual belikan terjamin

5	Respon	Respon keluhan ditanggapi dengan cepat
---	--------	--

Untuk mendapatkan data-data indikator diatas maka penelitian ini menggunakan kuesioner yang disebarakan kepada konsumen baik secara offline maupun online dengan link kuesioner terlampir. Jumlah responden yang terlibat dalam penelitian ini adalah sebanyak 100 orang dan data ini merupakan bahan awal untuk menentukan kepuasan konsumen.

3.2.2 Wawancara

Proses wawancara pada penelitian ini dilakukan kepada karyawan yang ada di PT Tas Import Sukses Batam pada bagian penjualan dan produksi serta bagian layanan, agar ditemukan permasalahan atau penyebab terjadinya ketidak puasan konsumen.

3.2.3 Studi Literatur

Selain data yang dikumpulkan secara langsung, penelitian ini juga membutuhkan data dari buku, jurnal, dan artikel yang berkaitan dengan topik penelitian. Penelitian juga membutuhkan pengetahuan tentang naïve bayes. Selain itu, peneliti juga membutuhkan referensi dari penelitian sebelumnya yang melakukan penelitian serupa.

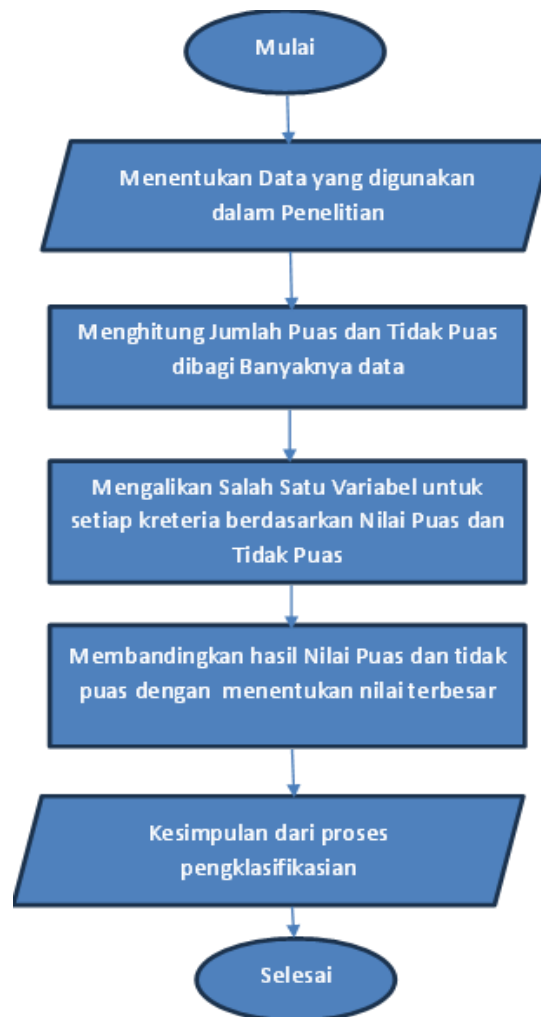
3.3 Operasional Variabel

Variabel penelitian adalah suatu atribut atau nilai yang berbentuk segalanya dan memiliki variasi tersendiri yang ditentukan oleh peneliti untuk dipahami dan ditarik kesimpulannya (Sugiyono, 2015).

1. Variabel dependen terdiri dari Kepuasan *customer*, variabel ini dijadikan sebagai tujuan dari penelitian yang dilakukan.
2. variabel independen terdiri dari harga, kualitas, pelayanan, Jaminan dan respon. Variabel ini di gunakan sebagai poin-poin dasar dalam pembuatan kuisisioner yang digunakan untuk mendapatkan hasil dari penelitian.

3.4 Algoritma *Naïve Bayes*

Untuk dapat memahami alur algoritma *naïve bayes* maka penelitian ini menggunakan model kerangka analisis sebagai berikut:



Gambar 3.2 Algoritma *Naïve bayes*

Dari gambar 3.2 dijelaskan bahwa untuk menentukan tingkat kepuasan konsumen PT Tas Import Sukses Batam, metode *Naive Bayes* digunakan untuk melakukan beberapa penyelesaian berikut:

1. Menemukan data yang akan digunakan untuk proses pengklasifikasian tingkat kepuasan konsumen PT Tas Import Sukses Batam terhadap layanan. dimana data dikumpulkan melalui survei yang diberikan kepada konsumen .
2. Menghitung jumlah kepuasan dan ketidakpuasan untuk setiap kriteria berdasarkan data yang diperoleh, kemudian membagi kepuasan dan ketidakpuasan dengan jumlah data yang lebih besar.
3. Menghitung jumlah kepuasan dan ketidakpuasan untuk setiap kriteria berdasarkan data yang diperoleh, kemudian menghitung probabilitas kepuasan dan ketidakpuasan untuk setiap kriteria.
4. Menghitung perkalian salah satu variabel yang digunakan untuk mengevaluasi tingkat kepuasan pelanggan terhadap pelayan.
5. Membandingkan nilai puas dan tidak puas dengan melihat nilai terbesar dari nilai yang telah didapatkan.
6. Memberikan kesimpulan dari data yang diklasifikasikan.

3.4.1 Data Penelitian

Sampel Data penelitian yang didapatkan dari hasil penyebaran kuesioner adalah sebagai berikut:

Tabel 3.2 Sampel *Dataset*

No	Usia	Jenis Kelamin	Harga	Pelayanan	Kualitas	Jaminan	Respon	Class Kepuasan
1	2	1	4	3	4	5	5	Puas
2	1	2	5	5	4	4	3	Puas
3	1	1	3	2	4	4	3	Tidak Puas
4	1	1	3	2	3	3	4	Tidak Puas
5	2	2	4	3	4	4	5	Puas
6	3	2	2	2	2	2	3	Tidak Puas
7	1	1	3	3	3	4	4	Puas
8	3	1	2	3	2	4	3	Tidak Puas
9	3	1	3	3	3	3	4	Puas
10	4	2	3	4	3	4	4	Puas
11	1	1	5	4	3	5	5	Puas
12	2	2	4	4	4	5	4	Puas
13	3	1	3	1	3	3	4	Tidak Puas
14	2	2	2	2	3	4	3	Tidak Puas
15	2	2	3	3	3	4	3	Puas
16	1	1	4	4	4	3	4	Puas
17	2	1	4	3	4	2	4	Tidak Puas
18	3	2	5	5	3	4	4	Puas
....
100	3	1	3	1	3	3	4	Tidak Puas

Nilai numerik mewakili keterangan berikut ini:

Tabel 3.3 Skala dan Bobot

Keterangan	Bobot
Sangat Setuju	4
Setuju	3
Tidak Setuju	2
Sangat tidak Setuju	1

Rule yang digunakan untuk menentukan class kepuasan yaitu

```
"=IF(Harga<=2,"Tidak Puas",IF(Pelayanan<=2,"Tidak
Puas",IF(Kualitas<=2,"Tidak Puas",IF(Jaminan<=2,"Tidak
Puas",IF(Respon<=2,"Tidak Puas","Puas")))))))"
```

3.4.2 Data *Training* dan *Testing*

Untuk mendapatkan tingkat akurasi terkait dengan kepuasan konsumen maka diperlukan *training* dan *testing* data. Pada penelitian ini menggunakan mode *training cross validation* yaitu model *k-Fold Cross-Validation* untuk memastikan bahwa model pembelajaran yang dibuat, yang menggunakan data pelatihan, dapat menggeneralisasi data segar yang belum pernah dilatih sebelumnya. Teknik ini menghindari kondisi *overfit* atau meminimalkan bias (perbedaan antara nilai aktual dan keluaran model) dan variasi (perbedaan antara keluaran model dalam data pelatihan dan pengujian). Metode *k-fold cross-validation* membagi kumpulan data D secara acak menjadi k subset independen (biasanya disebut folds): f1, f2, ... fk, sehingga setiap fold berisi 1 / k data. Selanjutnya, buat k data set D1, D2, ... Dk, masing-masing berisi (k-1) fold untuk data training dan satu fold untuk data testing (Wayan Wardani et al. 2022).

3.4.3 Rumus Pencarian Tingkat Akurasi

Akurasi adalah seberapa jauh nilai sebenarnya dan prediksi berbeda. Untuk mengukur akurasi model, digunakan *matrix confusion* yang menitik beratkan pada kelasnya. Sebuah array yang digunakan untuk mencatat hasil kerja klasifikasi disebut *confusion matrix*. Pada langkah ini, *matrix confusion* menguji data dan menggunakan model untuk menemukan tingkat akurasi terbaik.

Tabel 3.4 *Confusion Matrix*

Clasifikasi Naïve Bayes	Puas (+)	Tidak Puas (-)
Puas (+)	True Positif (TP)	False Negatif (FN)
Tidak Puas	False Positif (FP)	True Negatif (TN)

Untuk menentukan tingkat akurasi pada naïve bayes, maka untuk hasil akurasi yang terbaik memerlukan data minimal 100 data. Teori Naïve Bayes merupakan teori probabilitas untuk melakukan klasifikasi. Rumus confusion matrik yang digunakan adalah :

$$\text{Akurasi} : \frac{TP+TN}{TP+FN+FS+TN} \quad \text{Rumus 3.1 Rumus Confusion Matrix}$$

Keterangan :

- a. Akurasi adalah jumlah prediksi yang benar yang menyatakan puas.
- b. TP (*True positive*) adalah jumlah responen yang diklasifikasikan sebagai positif oleh classifier.
- c. N (*True negative*) adalah jumlah Tidak puas yang diklasifikasikan sebagai negatif oleh clasifier.
- d. FP (*False positive*) adalah jumlah record negatif yang diklasifikasikan sebagai positif oleh classifier
- e. FN (*False negative*) adalah jumlah puas tapi tidak puas yang diklasifikasikan sebagai negatif oleh *classifier*

Sedangkan rumus *Naive Bayes* yang digunakan dalam penelitian adalah

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

Tabel 3.5 Variabel Penelitian

Variabel	Keterangan
X	Data yang memiliki kelas yang tidak diketahui
Y	Hipotesis data X adalah kelas tertentu
$P(Y X)$	Hipotesis Y, berdasarkan kondisi X
$P(Y)$	Potensi hipotesis Y
$P(X Y)$	Potensi X didasarkan pada kondisi pada saat hipotesis Y
$P(X)$	Probabilitas (X)

3.4.4 Tool Analisis

Untuk memudahkan dalam proses analisis, maka penelitian ini menggunakan tiga *software* bantu yaitu *excel*, *rapidminer* dan program *WEKA*. Tujuan menggunakan aplikasi ini agar memudahkan dalam proses analisis dan perhitungan data-data yang ada menggunakan algoritma *naïve bayes*.

3.5 Lokasi dan Jadwal Penelitian

Penelitian ini dilakukan di PT Tas Import Sukses Batam selama enam bulan- dari Maret hingga Agustus 2023.

Tabel 3.6 Jadwal Penelitian

Kegiatan	Maret 2023	April 2023	Mei 2023	Juni 2023	Juli 2023	Agustus 2023
Identifikasi Masalah						
Penentuan Judul						
Penelusuran Pustaka						
Penulisan Skripsi						
Pengumpulan Data						
Persiapan <i>analisis</i>						
Analisa temuan						
Pengujian Hotspot						
Laporan Akhir						