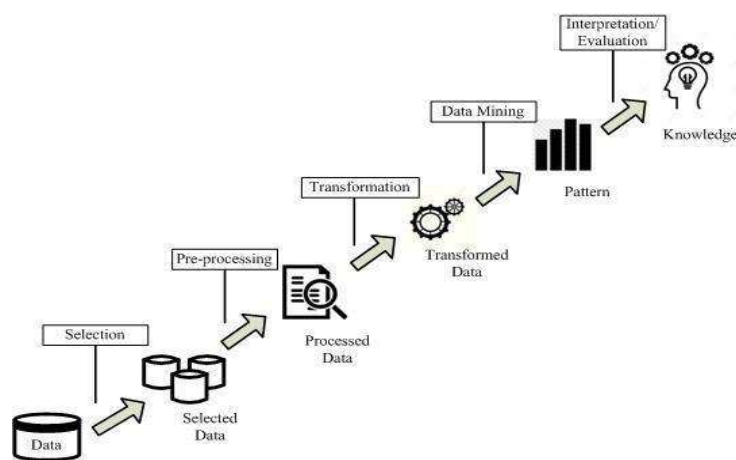


## BAB II

### 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.

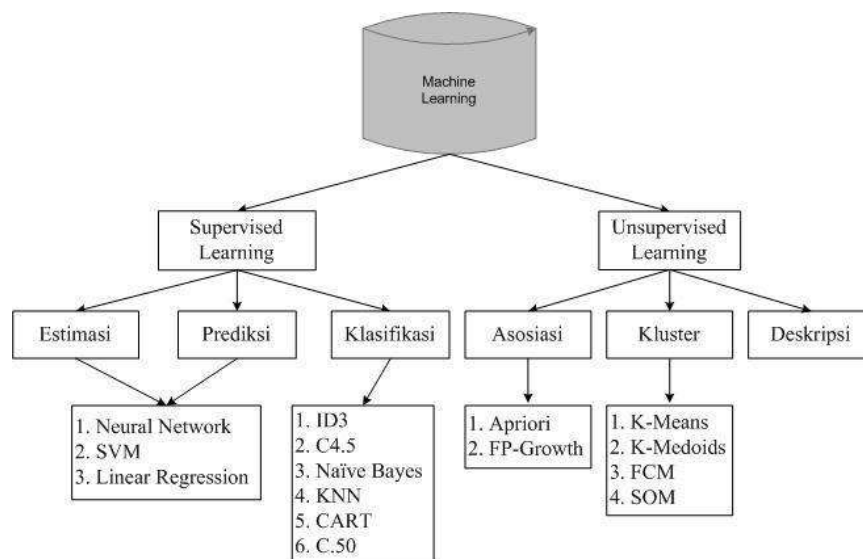
### **2.3.2 Tahapan *Data Mining***

(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:

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

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:

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

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:

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

Keterangan :

$P$  : Probability

$X_i$  : Feature ke  $i$

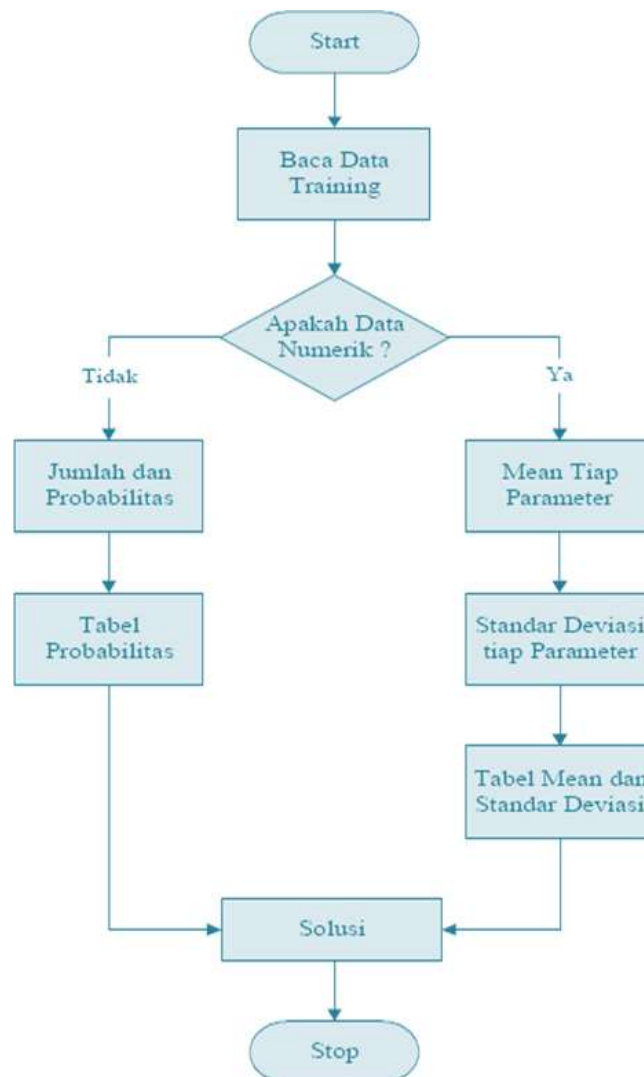
$x_i$  : Nilai Feature ke  $i$

$Y$  : Kelas yang dicari

$y_i$  : Sub kelas  $Y$  yang dicari

$\mu$  : *mean*, rata – rata dari seluruh atribut

$\sigma$  : Deviasi standar merupakan varian dari seluruh atribut.



**Gambar 2.3** Flowchart Metode Naïve 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:

$\mu$  : *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:

$\sigma$  : *standard deviation*

$x_i$  : *value x to -i*

$\mu$  : *average count*

$n$  : *number of samples*

b. Cari nilai probabilitas 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

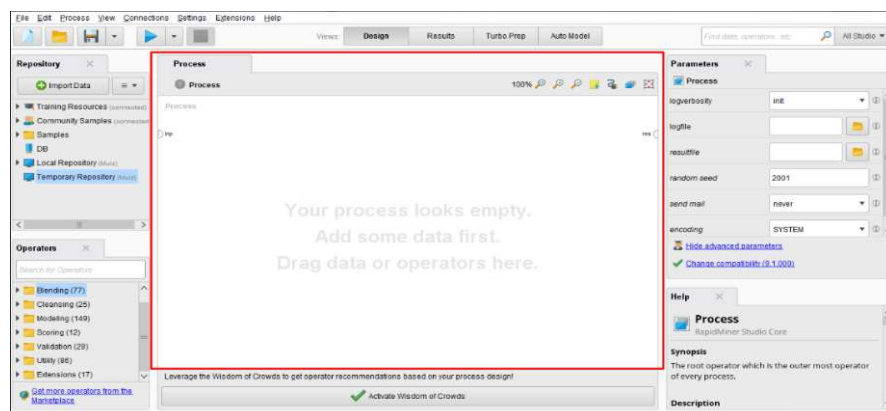
Jika data *mining* dilakukan secara manual, yang akan memakan waktu lebih lama dan kurang akurat, proses pengolahan data akan lebih mudah dan akurat jika dibantu oleh software pengolahan data yang tersedia saat ini. *Rapidminer*, *Microsoft Excel*, dan *WEKA* adalah program yang digunakan dalam penelitian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

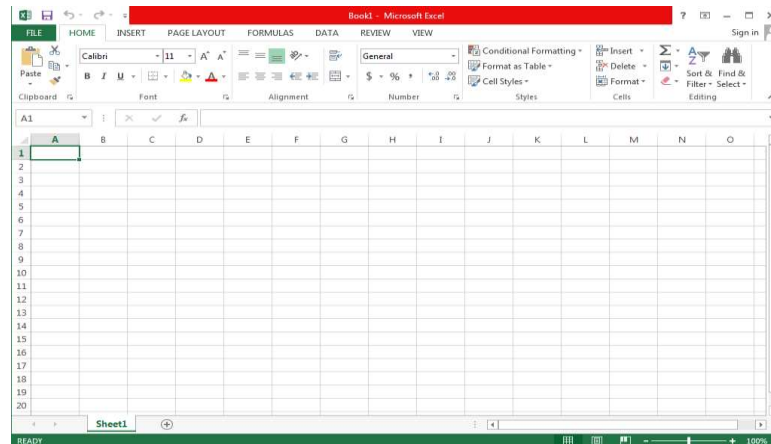
### 2.5.3 Microsoft Excel



**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 *Naïve 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 detergen yang diproduksi perlu diperhatikan dengan menggunakan metode *naïve 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.

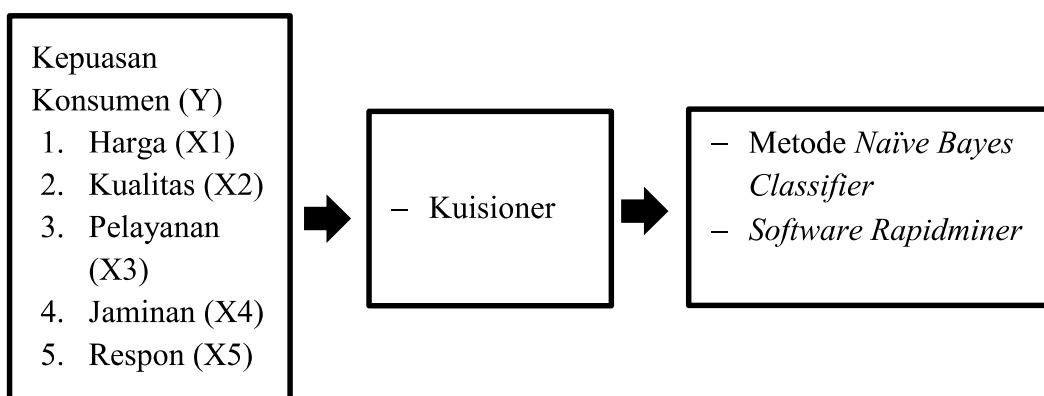
4. Menurut Penelitian (Sitohang 2021) menjelaskan tujuan dari penelitian ini adalah melakukan pengolahan data penjualan laptop di toko *Good Cell* Batam menggunakan algoritma C4.5 agar mempermudah pembeli untuk memilih tipe laptop yang akan dibeli berdasarkan kriteria yang diinginkan pembeli. Dengan 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 kuisisioner, 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 disebar ke *costumer* lalu di olah menggunakan aplikasi *rapid miner* dengan menerapkan algoritma *naïve bayes*.