

BAB II

TINJAUAN PUSTAKA

2.1 Data Mining

2.1.1 Pengertian Data Mining

Data mining adalah suatu proses pengumpulan informasi dan data yang penting dalam jumlah yang besar. Data mining digunakan untuk melakukan analisis dan prediksi untuk kedepannya, dan memberikan informasi berguna bagi keperluan bisnis. Data mining juga disebut sebagai serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual dari suatu kumpulan data (Pramudiono, 2007). Sedangkan menurut Santoso (2007), data mining adalah proses yang bertujuan untuk menemukan suatu informasi atau pengetahuan yang berguna dari data yang berskala besar.

Data Mining berguna selama tahap perencanaan, memberikan informasi yang akurat untuk membuat prediksi dan mengurangi biaya berdasarkan tren masa lalu dan kondisi saat ini. Data Mining memungkinkan perusahaan menggunakan alokasi sumber daya secara lebih efisien dengan mengurangi biaya melalui otomatisasi pengambilan keputusan.

2.1.2 Teknik Dasar Data Mining

2.1.2.1 Association

Association adalah metode berbasis aturan yang digunakan untuk menemukan asosiasi dan hubungan variabel dalam satu set data. Biasanya analisis ini terdiri dari pernyataan “if atau then” sederhana.

2.1.2.2 Classification

Selanjutnya classification, ia adalah metode yang paling umum digunakan dalam data mining. Classification adalah tindakan untuk memprediksi kelas suatu objek.

2.1.2.3 Regression

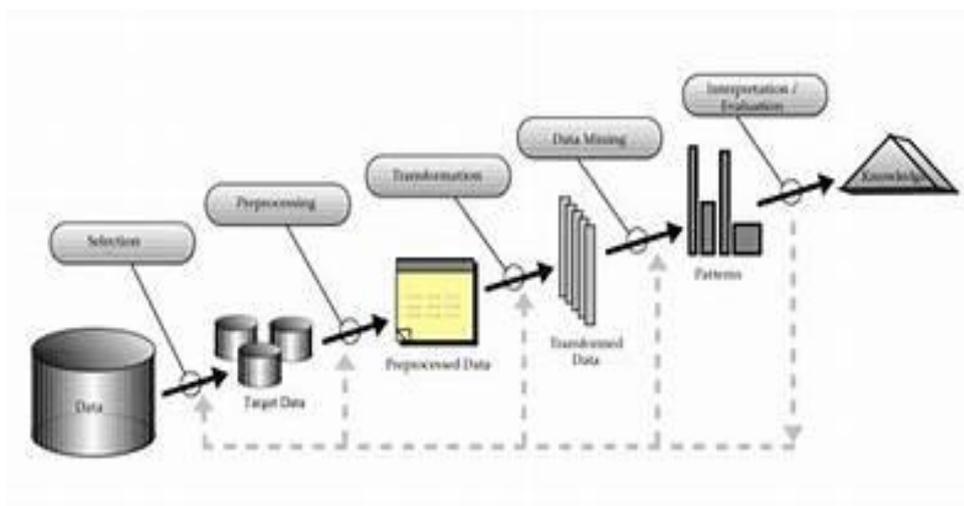
Regression adalah teknik yang menjelaskan variabel dependen melalui proses analisis variabel independen.

2.1.2.4 Clustering

Clustering digunakan dalam membagi kumpulan data menjadi beberapa kelompok berdasarkan kemiripan atribut yang dimiliki.

2.1.3 Tahapan Dalam Data Mining

Tahapan yang dilakukan pada proses data mining diawali dari seleksi data dari data sumber ke data target, tahap preprocessing untuk memperbaiki kualitas data, transformasi, data mining serta tahap interpretasi dan evaluasi yang menghasilkan output berupa pengetahuan baru yang diharapkan memberikan kontribusi yang lebih baik (Lestari, 2022). Secara detail dijelaskan sebagai berikut:



Gambar 2.1 Tahap-tahap Data Mining

Tahap-tahap data mining adalah sebagai berikut:

1. Data selection

Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang digunakan untuk proses data mining, disimpan dalam suatu berkas, terpisah dari basis data operasional.

2. Pre-processing /cleaning.

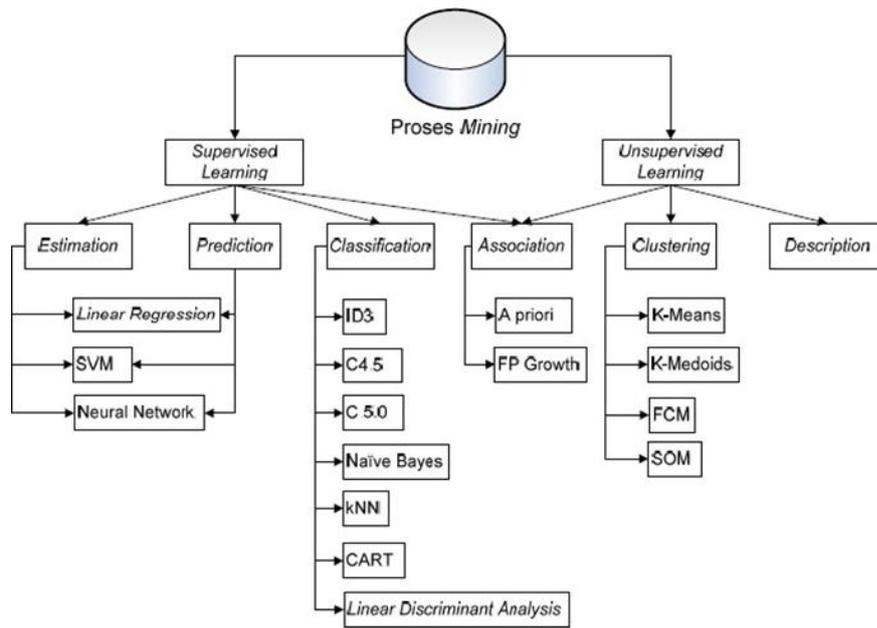
Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses cleaning pada data yang menjadi fokus KDD. Proses cleaning mencakup antara lain membuang duplikasi data, memeriksa data yang konsisten, dan memperbaiki kesalahan pada data.

3. Transformation

Proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Proses coding dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

4. Data mining

Proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.



Gambar 2.2 Beberapa metode Data Mining

5. Interpretation/evaluation

Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut interpretation.

2.2 Algoritma Naive Bayes

2.2.1 Pengertian Algoritma Naive Bayes

Naive bayes merupakan metode pengklasifikasian paling populer digunakan dengan tingkat keakuratan yang baik. Banyak penelitian tentang pengklasifikasian yang telah dilakukan dengan menggunakan algoritma ini. Berbeda dengan metode pengklasifikasian dengan logistic regression ordinal maupun nominal, pada algoritma naive bayes pengklasifikasian tidak membutuhkan adanya pemodelan maupun uji statistik .

Naive bayes merupakan metode pengklasifikasian berdasarkan probabilitas sederhana dan dirancang agar dapat dipergunakan dengan asumsi antar variabel penjelas saling bebas (independen). Pada algoritma ini pembelajaran lebih ditekankan pada pengestimasiian probabilitas. Keuntungan algoritma naive bayes adalah tingkat nilai error yang didapat lebih rendah ketika dataset berjumlah besar, selain itu akurasi naive bayes dan kecepatannya lebih tinggi pada saat diaplikasikan ke dalam dataset yang jumlahnya lebih besar.

2.2.2 Tipe Algoritma Naive Bayes

2.2.2.1 Bernoulli Naive Bayes

Dimana prediktornya adalah variabel boolean. Oleh karena itu satu-satunya yang ada hanya benar atau salah. Biasanya algoritma ini digunakan ketika data sesuai dengan distribusi bernoulli multivariat.

2.2.2.2 Naive Bayes Multinomial

Dimana algoritma ini sering digunakan untuk memecahkan masalah klasifikasi dokumen. Contohnya, jika ingin menentukan apakah suatu dokumen termasuk dalam suatu kategori dan algoritma ini digunakan untuk memilahnya. Naive bayes menggunakan frekuensi kata-kata sekarang sebagai fitur.

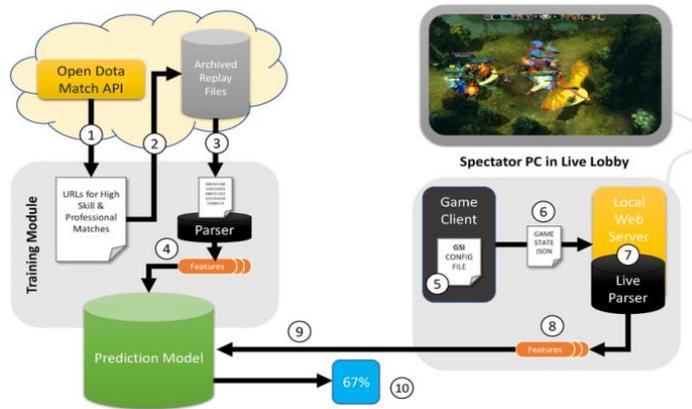
2.2.2.3 Gaussian Naive Bayes

Digunakan jika prediktor tidak diskrit namun memiliki nilai kontinu dan prediktor tersebut diasumsikan sebagai sampel dari distribusi gaussian.

2.2.3 Manfaat Algoritma Naive Bayes

Algoritma naive bayes memiliki banyak manfaat sehingga banyak digunakan di berbagai aspek kehidupan, antara lain sebagai berikut:

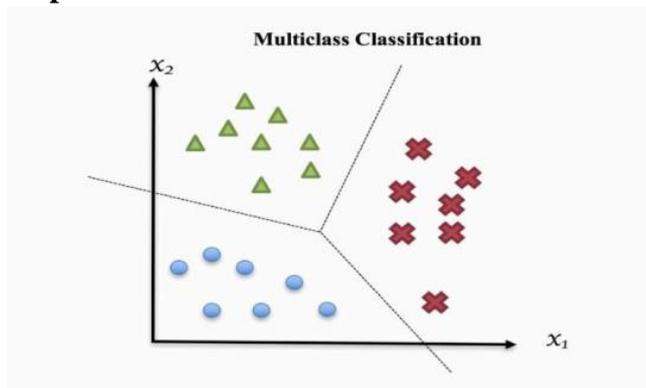
1. Real time prediction.



Gambar 2.3 Real Time Prediction

Naive bayes adalah algoritma pengklasifikasi pembelajaran yang bersemangat dan pasti cepat. Dengan demikian, algoritma ini dapat digunakan untuk membuat *real time prediction*.

2. Multiclass prediction



Gambar 2.4 Multiclass Classification

Algoritma ini juga terkenal dengan fitur *multi class prediction*. Algoritma ini dapat memprediksi probabilitas beberapa kelas variabel target.

2.2.4 Rumus Algoritma Naive Bayes

Untuk bisa lebih memahami algoritma ini, berikut rumus umum Teorema Bayes yang menjadi dasar Naive bayes.

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

Gambar 2.7 Rumus Umum Naive Bayes

Keterangan :

X = Sampel data yang memiliki *class* (label) yang tidak diketahui.

C = Hipotesis bahwa **X** adalah data *class* (label).

P(C) = Probabilitas hipotesis **C**.

P(X) = Peluang dari data sampel yang diamati (probabilitas **C**).

P(X|C) = Probabilitas berdasarkan kondisi pada hipotesis.

Adapun alur dari metode naive bayes sebagai berikut.

1. Menghitung nilai peluang kasus baru dari setiap hipotesa dengan *class* (label) yang ada di **P(C_i)**.
2. Menghitung nilai akumulasi peluang dari setiap kelas **P(X|C_i)**.
3. Menghitung nilai **P(X|C_i) x P(C_i)**.
4. Menentukan *class* dari kasus baru tersebut.

Namun jika atribut ke-*i* bersifat kontinu, maka **P(X_i|C)** diestimasi dengan fungsi densitas Gaussian.

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

μ = mean, σ = deviasi standar.

Gambar 2.8 Rumus Fungsi Densitas Gaussian

Rumus Teorema Bayes diatas tadi menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam *class* C (posterior) adalah peluang munculnya kelas C (sebelum masuknya sampel tersebut, sering kali disebut *prior*), dikali dengan peluang kemunculan karakteristik-karakteristik sampel pada *class* C (disebut juga *likelihood*) kemudian dibagi dengan peluang kemunculan karakteristik sampel secara global (disebut juga *evidence*). Oleh karena itu rumus dapat pula ditulis sebagai berikut.

$$\textit{Posterior} = \frac{\textit{Prior} \times \textit{Likelihood}}{\textit{Evidence}}$$

Gambar 2.9 Rumus Posterior

Nilai *evidence* selalu tetap untuk setiap *class* pada satu sampel. Nilai dari *posterior* tersebut nantinya akan dibandingkan dengan nilai *posterior* class lainnya untuk menentukan *class* apa suatu sampel akan diklasifikasikan.

2.2.5 Cara Kerja Algoritma Naive Bayes

Cara kerja Naive Bayes Classifier melalui dua tahapan, yaitu :

- Learning (Pembelajaran) *Naive Bayes* adalah suatu metode yang termasuk ke dalam supervised learning, maka akan dibutuhkan pengetahuan awal untuk dapat mengambil keputusan.

Langkah-langkah :

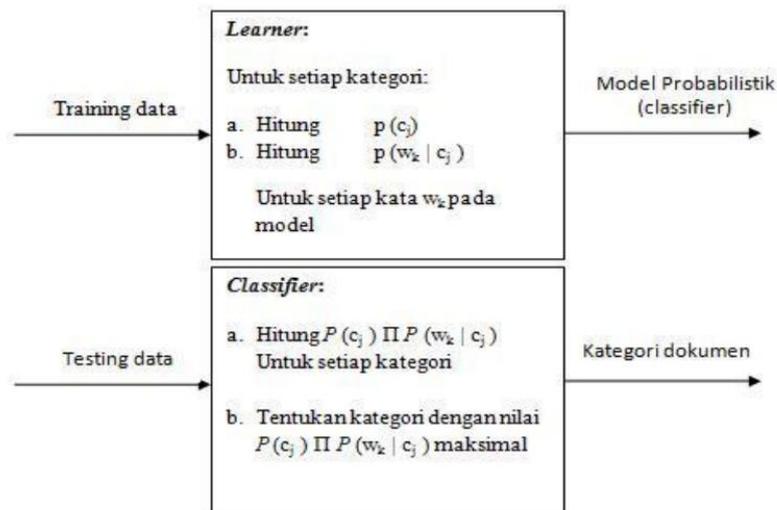
Step 1 : Bentuk vocabulary pada setiap dokumen data training

Step 2 : Hitung probabilitas pada setiap kategori

- Classify (Pengklasifikasian). Langkah-langkahnya adalah :

Step 1 : Hitung $P(v_j) \prod P(w_k | v_j)$ untuk setiap kategori.

Step 2 : Tentukan kategori dengan nilai $P(v_j) \prod P(w_k | v_j)$ maksimal.



Gambar 2.10 Tahapan Proses Klasifikasi Dokumen Naive Bayes Classifier

2.2.6 Kelebihan Algoritma Naive Bayes

1. Hanya memerlukan sejumlah kecil data pelatihan untuk mengestimasi parameter (rata-rata dan variasi dari variabel) yang dibutuhkan untuk klasifikasi.
2. Menangani nilai yang hilang dengan mengabaikan instansi selama perhitungan estimasi peluang.
3. Cepat dan efisiensi ruang.
4. Kokoh terhadap atribut yang tidak relevan.

2.2.7 Kelemahan Algoritma Naive Bayes

1. Tidak berlaku jika probabilitas kondisionalnya adalah nol, apabila nol maka probabilitas prediksi akan bernilai nol juga.
2. Mengasumsikan variabel bebas.
3. Kelemahan lainnya terletak pada masalah probabilitas. Naive Bayes memiliki masalah probabilitas nol, terutama saat Anda menemukan kata-kata dalam data pengujian untuk kelas tertentu yang tidak ada dalam data pelatihan. Kemungkinan besar Anda akan berakhir dengan probabilitas “*Zero Frequency*”. Meski begitu, probabilitas nol ini dapat diatasi dengan *smoothing* (teknik penghalusan). Tambahkan faktor penghalusan pada pembilang dan penyebut setiap probabilitas untuk menghindari munculnya nilai nol.

2.3 Algoritma K-Nearest Neighbors

2.3.1 Pengertian Algoritma K-Nearest Neighbors

Algoritma K-Nearest Neighbors (K-NN) adalah algoritma machine learning yang bersifat non-parametric dan lazy learning. Metode yang bersifat non-parametric memiliki makna bahwa metode tersebut tidak membuat asumsi apapun tentang distribusi data yang mendasarinya. Dengan kata lain, tidak ada jumlah parameter atau estimasi parameter yang tetap dalam model, terlepas dari data tersebut berukuran kecil atau besar.

K-nearest neighbor adalah salah satu algoritma yang paling mudah dipahami dan dipakai. Biasanya, *K-nearest neighbor* digunakan untuk mengumpulkan data dari jejaring sosial. Konsep utama dari algoritma *K-nearest neighbor* adalah mengambil kumpulan data yang berbeda. Pengaplikasian *K-nearest neighbor* memang cukup sederhana dan mudah.

K-nearest neighbor merupakan algoritma yang digunakan untuk mengidentifikasi adanya persamaan antara data baru dan lama. Kemudian, algoritma satu ini akan memasukkan data baru tersebut dalam kategori yang paling mirip dengan kategori yang telah ada

sebelumnya. Dengan kata lain, *K-nearest neighbor* menyimpan seluruh data lama dan mengklasifikasikan *data point* baru berdasarkan kemiripan.

2.3.2 Tujuan Algoritma K-Nearest Neighbors

Tujuan algoritma K-Nearest Neighbors (K-NN) adalah untuk mengidentifikasi tetangga terdekat dari titik kueri yang diberikan, sehingga kita dapat menetapkan label kelas ke titik tersebut.

2.3.3 Cara Kerja Algoritma K-Nearest Neighbors

K-NN bekerja berdasarkan prinsip bahwa setiap titik data yang berdekatan satu sama lain akan berada di kelas yang sama. Dengan kata lain, K-NN mengklasifikasikan titik data baru berdasarkan kemiripan. Berikut adalah langkah-langkah utama dalam cara kerja algoritma KNN:

1. Pilih nilai K

Langkah pertama adalah memilih nilai K, yang menentukan jumlah tetangga yang akan diperiksa dalam menentukan klasifikasi suatu titik data. Misalnya, jika K sama dengan 3, algoritma akan mempertimbangkan tiga tetangga terdekat.

Cara Menentukan nilai K pada Algoritma K-Nearest Neighbor :

Untuk menentukan nilai K yang benar, Anda bisa menjalankan algoritma *K-nearest neighbor regression* beberapa kali dengan memakai nilai K yang berbeda-beda. Hal ini bertujuan untuk mengurangi jumlah kesalahan yang Anda temui, sambil mempertahankan algoritma untuk memprediksi dengan akurat ketika data diberikan. Langkah-langkahnya adalah sebagai berikut:

1. Pertama, Anda bisa menurunkan nilai K menjadi 1. Bisa diprediksi bahwa pada tahap ini, nilai menjadi kurang stabil. Bayangkan jika $K=1$ dan Anda memiliki titik kueri yang dikelilingi beberapa warna

sama dan satu warna yang berbeda. Namun, warna berbeda tersebut adalah yang terdekat. Jadi, karena $K=1$, maka titik kueri adalah warna yang berbeda tersebut.

2. Kedua, kalau Anda menaikkan nilai K , prediksi yang dilakukan bisa lebih stabil karena ada suara mayoritas. Maka, lebih memungkinkan bagi Anda membuat prediksi yang lebih akurat. Ketika Anda mengambil suara terbanyak, Anda biasanya akan membuat K dengan angka ganjil.

2. Hitung jarak

Kemudian, algoritma menghitung jarak antara titik data baru yang akan diklasifikasikan dengan semua titik data yang ada dalam set pelatihan. Perhitungan jarak ini dapat menggunakan berbagai metrik, seperti Euclidean distance, Hamming distance, Manhattan distance, dan Minkowski distance, tergantung pada jenis data dan masalah yang dihadapi.

3. Temukan tetangga terdekat

Setelah menghitung jarak, KNN akan mengidentifikasi K tetangga terdekat yang memiliki jarak terpendek dari titik data baru.

4. Tentukan kelas mayoritas

Dari tetangga-tetangga yang telah diidentifikasi, KNN akan menghitung jumlah titik data di setiap kelas atau kategori. Kelas yang memiliki jumlah tetangga terbanyak akan dianggap sebagai klasifikasi untuk titik data baru tersebut.

5. Klasifikasikan titik data baru

Akhirnya, KNN akan memetakan titik data baru ke dalam kelas yang telah ditentukan sebagai mayoritas oleh tetangga-tetangga terdekatnya. Dengan demikian, titik data baru akan mendapatkan label atau kategori yang sesuai.

2.3.4 Kelebihan Algoritma K-Nearest Neighbors

1. Kemudahan implementasi

KNN dikenal karena kesederhanaannya. Algoritma ini mudah untuk dipahami dan diterapkan, sehingga sangat cocok untuk pemula dalam dunia data science dan machine learning.

2. Kemampuan beradaptasi

KNN dapat dengan mudah menyesuaikan diri dengan perubahan dalam dataset. Ketika sampel pelatihan baru ditambahkan, KNN akan langsung mempertimbangkan data baru ini karena semua data pelatihan disimpan dalam memori.

3. Hyperparameter yang sedikit

KNN hanya memiliki dua hyperparameter utama: nilai K (jumlah tetangga terdekat yang akan diperhitungkan) dan metrik jarak yang digunakan. Hal ini membuat proses penentuan parameter model menjadi lebih sederhana dan kurang kompleks dibandingkan dengan beberapa algoritma lainnya.

2.3.5 Kelemahan Algoritma K-Nearest Neighbors

1. Tidak cocok untuk dataset berukuran besar

Salah satu kelemahan utama KNN adalah ketidakcocokannya untuk dataset berukuran besar. Algoritma ini memerlukan perhitungan jarak antara titik baru dengan semua titik dalam dataset, sehingga biaya komputasi menjadi sangat besar dan dapat mengurangi kinerja algoritma secara signifikan pada dataset yang besar.

2. Tidak cocok untuk dimensi tinggi

KNN tidak efektif pada data dengan dimensi tinggi. Ketika jumlah dimensi meningkat, algoritma akan menghadapi masalah perhitungan jarak yang semakin rumit dan memerlukan lebih banyak data untuk melakukan perhitungan yang akurat.

3. Penskalaan fitur diperlukan

Sebelum menerapkan KNN pada dataset, penting untuk melakukan penskalaan fitur. Tanpa penskalaan yang benar (standarisasi dan normalisasi), KNN dapat menghasilkan prediksi yang salah karena beberapa fitur memiliki skala yang dominan.

4. Sensitif terhadap noise, missing value, dan outlier

KNN cenderung sensitif terhadap noise dalam dataset. Ini berarti kita perlu melakukan pemrosesan data yang cermat, termasuk mengatasi nilai yang hilang dan mengidentifikasi serta mengatasi outlier, sebelum menggunakan KNN.

2.4 Confusion Matrix

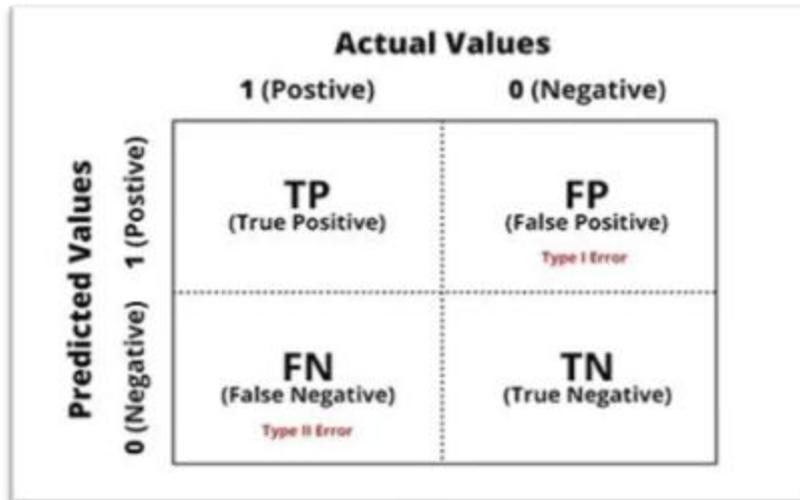
2.4.1 Pengertian Confusion Matrix

Confusion Matrix merupakan sebuah teknik yang digunakan dalam data mining dan *machine learning* untuk menghitung seberapa baik sebuah model dapat memprediksi label dari sebuah data. Teknik ini sering digunakan dalam evaluasi model *classification* yang mana model harus memprediksi label dari sebuah data berdasarkan atribut-atribut yang ada.

Confusion Matrix adalah sebuah tabel yang menggambarkan seberapa sering model memprediksi label yang benar dan salah. Setiap baris dari tabel tersebut mewakili sebuah label aktual, sedangkan setiap kolom mewakili label yang diprediksi oleh model.

2.4.2 Hasil dari Confusion Matrix

Confusion Matrix membantu mengukur kinerja dimana output algoritma dapat berada dalam dua kategori atau lebih biasa disebut **Positive** atau **Negative**, **Ya** atau **Tidak**. Setiap tabel terdiri dari empat sel, masing-masing mewakili kombinasi unik dari nilai prediksi atau aktual. Empat hasil potensial tersebut adalah.



Gambar 2.11 Hasil Confusion Matrix

Keterangan :

1. True Positive (TP)

Merupakan data positif yang diprediksi benar. Contohnya, pasien menderita hipertensi (*class 1*) dan model membuat prediksi bahwa pasien tersebut hipertensi (*class 1*).

2. True Negative (TN)

Merupakan data negatif yang diprediksi benar. Contoh, pasien tidak menderita hipertensi (*class 2*) dan model membuat prediksi bahwa pasien tersebut tidak menderita hipertensi (*class 2*).

3. False Positive (FP) – Type I Error

Merupakan data negatif namun diprediksi sebagai data positif. Contoh, pasien tidak menderita hipertensi (*class 2*) tetapi model memprediksi pasien tersebut menderita hipertensi (*class 1*).

4. False Negative (FN) – Type II Error

Merupakan data positif namun diprediksi sebagai data negatif. Contohnya, pasien menderita hipertensi (*class 1*) tetapi model memprediksi pasien tersebut tidak menderita hipertensi (*class 2*).

2.4.3 Cara Menggunakan Confusion Matrix

Berikut ini cara menghitung metode evaluasi menggunakan Confusion Matrix

1. Accuracy

Accuracy merupakan metode pengujian berdasarkan tingkat kedekatan antara nilai prediksi dengan nilai aktual. Dengan mengetahui jumlah data yang diklasifikasikan secara benar maka dapat diketahui akurasi hasil prediksi.

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

Gambar 2.12 Rumus Accuracy

2. Precision

Precision merupakan metode pengujian dengan melakukan perbandingan jumlah informasi relevan yang didapatkan sistem dengan jumlah seluruh informasi yang terambil oleh sistem baik yang relevan maupun tidak.

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

Gambar 2.13 Rumus Precision

3. Recall

Recall merupakan metode pengujian yang membandingkan jumlah informasi relevan yang didapatkan sistem dengan jumlah seluruh informasi relevan yang ada dalam koleksi informasi (baik yang terambil atau tidak terambil oleh sistem).

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

Gambar 2.14 Rumus Recall

4. F-Measure atau F1-Score

F-Measure atau disebut juga dengan F1-Score menggambarkan perbandingan rata-rata *precision* dan *recall* yang dibobotkan. *Accuracy* tepat kita gunakan sebagai acuan performansi algoritma jika data set kita memiliki jumlah data *false* negatif dan *false* positif yang sangat mendekati (*symmetric*). Namun jika jumlahnya tidak mendekati, maka sebaiknya kita menggunakan F1-Score.

$$F1 = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}$$

Gambar 2.15 Rumus F-Measure

2.5 Standar Nasional Perguruan Tinggi (SN-Dikti)

Predikat kelulusan mahasiswa telah diatur pada Pasal 27 Permendikbud 3 Tahun 2022 Tentang Standar Nasional Pendidikan Tinggi.

1. Mahasiswa program diploma dan program sarjana dinyatakan lulus apabila telah menempuh seluruh beban belajar yang ditetapkan dan memiliki capaian Pembelajaran lulusan yang ditargetkan oleh Program Studi dengan Indeks Prestasi Kumulatif (IPK) lebih besar atau sama dengan 2,00 (dua koma nol nol).
2. Kelulusan mahasiswa dari program diploma dan program sarjana dapat diberikan predikat memuaskan, sangat memuaskan, atau pujian dengan kriteria:
 - a. mahasiswa dinyatakan lulus dengan predikat memuaskan apabila mencapai Indeks Prestasi Kumulatif (IPK) 2,76 (dua koma tujuh enam) sampai dengan 3,00 (tiga koma nol nol);
 - b. mahasiswa dinyatakan lulus dengan predikat sangat memuaskan apabila mencapai Indeks Prestasi Kumulatif (IPK) 3,01 (tiga koma nol satu) sampai dengan 3,50 (tiga koma lima nol); atau mahasiswa dinyatakan lulus dengan predikat pujian apabila mencapai Indeks Prestasi Kumulatif (IPK) lebih dari 3,50 (tiga koma nol).
3. Kelulusan mahasiswa dari program profesi, program spesialis, program magister, program magister terapan, program doktor, program doctor terapan, dapat diberikan predikat dengan kriteria :
 - a. mahasiswa dinyatakan lulus dengan predikat memuaskan apabila mencapai Indeks Prestasi Kumulatif (IPK) 3,00 (tiga koma nol nol) sampai dengan 3,50 (tiga koma lima nol);

b. mahasiswa dinyatakan lulus dengan predikat sangat memuaskan apabila mencapai Indeks Prestasi Kumulatif (IPK) 3,51 (tiga koma lima satu) sampai dengan 3,75 (tiga koma tujuh lima); atau

c. mahasiswa dinyatakan lulus dengan predikat pujian apabila mencapai Indeks Prestasi Kumulatif (IPK) lebih dari 3,75 (tiga koma tujuh lima).

Tabel 2.1 Predikat Kelulusan SN-Dikti

Program	IPK	Predikat Lulusan
Diploma dan Sarjana		
Mahasiswa program diploma dan program sarjana dinyatakan lulus apabila telah menempuh seluruh beban belajar yang ditetapkan dan memiliki capaian pembelajaran lulusan yang ditargetkan oleh program studi dengan indeks prestasi kumulatif (IPK) lebih besar atau sama dengan 2,00 (dua koma nol)		
	2,76-3,00	Memuaskan
	3,01-3,50	Sangat Memuaskan
	>3,50	Pujian
Magister		
Mahasiswa program magister, dinyatakan lulus apabila telah menempuh seluruh beban belajar yang ditetapkan dan memiliki capaian pembelajaran lulusan yang ditargetkan oleh program studi dengan indeks prestasi kumulatif (IPK) lebih besar atau sama dengan 3,00 (tiga koma nol).		
	3,00-3,50	Memuaskan
	3,51-3,75	Sangat Memuaskan
	>3,75	Pujian
Mahasiswa yang dinyatakan lulus berhak memperoleh ijazah, gelar atau sebutan, dan surat keterangan pendamping ijazah sesuai dengan peraturan perundangan.		

2.6 Penelitian Terkait

Tabel 2.2 Penelitian Terkait

No.	Judul Penelitian	Metode	Kesimpulan
1	Analisis Metode K-Nearest Neighbors (K-NN) Dan Naive Bayes Dalam Memprediksi Kelulusan Mahasiswa	K-Nearest Neighbor dan Naive Bayes	Berdasarkan perhitungan akurasi menggunakan RapidMiner Metode K-NN yaitu sebesar 96,18% dan metode Naive Bayes sebesar 91,94%. Dapat dikatakan bahwa baik metode Naive Bayes ataupun metode K-Nearest Neighbor memiliki peluang yang kecil untuk melakukan kesalahan dalam proses prediksi, Disimpulkan bahwa pengklasifikasian menggunakan metode K-Nearest Neighbor lebih baik dibandingkan dengan metode Naive Bayes dalam proses prediksi kelulusan mahasiswa.
2	Metode Naive Bayes Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi)	Naive Bayes	Naive Bayes dapat melakukan pengklasifikasian dengan metode probabilitas dan statistik, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya. Nilai Presentase keakuratan menunjukkan keefektifan dataset yang diterapkan ke dalam metode Naive Bayes Clasification.

No.	Judul Penelitian	Metode	Kesimpulan
3	Analisis Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Data Mining Naïve Bayes : Systematic Review	Desicion Tree Dan Artificial Neural Network	Metode data mining naïve bayes memprediksi kelulusan mahasiswa tepat waktu dengan memperhitungkan atribut – atribut dari database perguruan tinggi yang digunakan. Tingkat akurasi menghasilkan akurasi di atas 90% walaupun dengan jumlah atribut dan aplikasi data mining yang berbeda. Atribut yang dapat menentukan prediksi adalah atribut IPK (Indeks Prestasi Kumulatif).
4	Model Algoritma K-Nearest Neighbor (K-Nn) Untuk Prediksi Kelulusan Mahasiswa	K-Nearest Neighbor	Tingkat akurasi pengujian model kelulusan mahasiswa dengan menggunakan algoritma K-Nearest Neighbor (K-NN) dipengaruhi oleh jumlah klustering data. Akurasi dan nilai AUC paling tinggi adalah dengan mengklaster data k=5.
5	Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Desicion Tree Dan Artificial Neural Network	Desicion Tree Dan Artificial Neural Network	Berdasarkan hasil pengujian metode decision tree memiliki akurasi sebesar 74,51% dan artificial neural network sebesar 79,74%. c. Metode artificial neural network memiliki akurasi lebih tinggi jika dibandingkan dengan decision tree karena data yang digunakan dalam penelitian ini berupa data label.

No.	Judul Penelitian	Metode	Kesimpulan
6	Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Naive Bayes : Studi Kasus UIN Syarif Hidayatullah Jakarta	Naive Bayes	Akurasi pengujian data yang diperoleh dalam penelitian ini sebesar 80,72% dari 1162 data yang digunakan untuk data training dan 587 data untuk testing.
7	Aplikasi Prediksi Kelulusan Mahasiswa Berbasis K-Nearest Neighbor (k-nn)	K-Nearest Neighbor	Pengujian dengan K-Fold Cross Validation didapatkan akurasi 80% ketika k-fold ke-4 dan 61% ketika nilai K=1. Pengujian dengan Confusion Matrix didapatkan akurasi tertinggi sebesar 98% pada K=1 untuk klasifikasi "Tepat Waktu" dan 98% pada K=2 untuk klasifikasi "Tidak Tepat Waktu".