BAB II
LANDASAN TEORI

2.1 Tinjauan Pustaka

Beberapa penelitian terdahulu yang menjadi tinjauan pusataka pada penelitian ini adalah sebagai berikut:

Tabel 2.1 Tabel Penelitian Terdahulu

No	Nama	Tahun	Judul	Hasil dan Kesimpulan
1	Herry Derajad Wijaya, Saruni Dwiasnati	2020	Implementasi Data Mining dengan Algoritma Naïve Bayes pada Penjualan Obat	Sistem Klasifikasi Penjualan Obat ini digunakan untuk menentukan merk obat apa yang LAKU atau TIDAK LAKUnya pada sebuah apotek yang ada di Jakarta dengan meggunakan algoritma Naïve Bayes. Penelitian ini menggunakan 150 dataset. Penelitian ini dilakukan karena adanya masalah untuk mengetahui produk pada bulan sebelumnya agar dapat menentukan stok produk mana yang mesti di perbanyak lagi, dan stok barang mana yang harus lebih dipersedikitkan. Pengujian pada data

No	Nama	Tahun	Judul	Hasil dan Kesimpulan
				rekapitulasi penjualan Obat dengan proses mining Algoritma Naïve Bayes menghasilkan tingkat accuracy dengan nilai 88.00%, dimana dalam pengujian model data, keseluruhan data set digunakan sebagai data testing. Penelitian ini dilakukan menggunakan tools Rapidminner pada dataset penjualan obat dengan menggunakan algoritma Naïve Bayes. (Wijaya, H. D., & Dwiasnati, 2020)
2	Simon Prananta Barus	2021	Implementation of Naïve Bayes Classifier based Machine Learning to Predict and Classify New Students at Matana University	Algoritma Naïve Bayes Classifier digunakan untuk memprediksi dan mengklasifikasikan mahasiswa baru di Universitas Matana. Data yang digunakan adalah data calon siswa yang terdaftar di Universitas Matana. Bahasa pemrograman Python digunakan untuk

No	Nama	Tahun	Judul	Hasil dan Kesimpulan
				pembelajaran mesin. Hasil aplikasi memiliki akurasi 0,73 (73%) dan sangat terbatas. Membantu kepala departemen uiversitas dalam pengembangan strategi pemasaran.(Barus, 2021)
3	Nafizatus Salmi dan Zuherman Rustam	2019	Model Pengklasifikasi Naïve Bayes untuk Memprediksi Kanker Usus Besar	Model Naïve Bayes Classifier merupakan metode klasifikasi untuk menunjukkan bahwa model memiliki akurasi yang tinggi, recall yang tinggi, baik dalam mengklasifikasikan data pasien kanker usus besar maupun tidak. Naïve Bayes Classifier adalah teknik prediksi berdasarkan probabilitas sederhana dan penerapan teorema Bayes dengan asumsi independensi yang kuat. Oleh karena itu, model ini dapat mengklasifikasikan dengan akurasi yang lebih tinggi dengan kompleksitas

No	Nama	Tahun	Judul	Hasil dan Kesimpulan
				yang lebih sedikit. Secara khusus, mencapai akurasi klasifikasi hingga 95,24%, sehingga model ini dapat menjadi alat analisis yang efektif. (Salmi, 2019)
4	Sri Wahyuni, Murni Marbun	2020	Implementasi Data Mining Dalam Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma Naïve Bayes	Program ini dikembangkan menggunakan algoritma Naive Bayes, yang bekerja berdasarkan jarak terpendek antara dua objek dengan menentukan nilai k. Data yang digunakan sebagai variabel adalah data akademik siswa, IPK 4 semester pertama, rata- rata nilai UN semasa SMA. Hasil survey menunjukkan bahwa aplikasi data mining yang dirancang dapat memprediksi waktu belajar mahasiswa Panka Bodi Development University. (Wahyuni, S., & Marbun, 2020)
5	Jian Zheng, Wei Yang1,	2022	Research on complaint	Teorema klasifikasi Bayesian untuk

No	Nama	Tahun	Judul	Hasil dan Kesimpulan
	Changchu n Wang, Dandan Jiang, Dan Wang, Qi Yang dan Yijiao		prediction based on feature weighted Naïve Bayes	mempelajari pola prediksi keluhan pelanggan listrik. Dengan menerima data pengaduan dari berbagai daerah di Provinsi Liaoning. Hasil penelitian menunjukkan bahwa model jaringan Bayesian berbobot fungsional yang dipelajari dalam penelitian adalah 92,05% ini lebih akurat dalam memprediksi keluhan pengguna listrik 10,56% lebih baik dari pada model Bayesian dasar. Prospek aplikasi bagus untuk memprediksi keluhan dari pengguna listrik. (Zheng, 2022)
6	Arya Mukti Saputra	2020	Sistem Prediksi Persediaan Obat Pada Apotek Menggunakan Metode Naive Bayes (Studi Kasus : Apotek Seger Waras,	Dengan adanya sistem ini akan dapat diketahui apakah obat yang akan dibeli akan laku atau tidak untuk tujuan dijual kembali, dan pengelola apotek akan dapat memilih produk yang akan dibeli dengan lebih selektif. Perancangan

No	Nama	Tahun	Judul	Hasil dan Kesimpulan
			Cianjur)	sistem ini menggunakan bahasa pemrograman PHP, MySQL sebagai database server, dan Microsoft Visio sebagai pendukung perancangan sistem. Informasi yang ditampilkan dalam sistem ini adalah informasi outbound, informasi inbound, dan pemberitahuan tanggal kadaluwarsa obat. (Saputra, 2020)
7	Tugiman, Lily Damayanti, Alexius Hendra Gunawan, Samuel Ryon Elkana	2022	Prediksi Penggunaan Obat Peserta Jaminan Kesehatan Nasional Menggunakan Algoritma Naïve Bayes Classifier	Sistem dapat memprediksi permintaan obat berdasarkan 5 (lima) tahun penjualan obat kepada pasien JKN. Data obat yang digunakan sebagai sampel uji kemudian diolah menggunakan pengujian algoritma pengklasifikasi Naive Bayes. Perangkat lunak menggunakan model User Acceptance Testing (UAT). Berdasarkan pengujian UAT, sistem dapat diterima dengan baik

No	Nama	Tahun	Judul	Hasil dan Kesimpulan
				oleh pengguna dengan skor 78,64% (kriteria baik). (Damayanti, L., 2022)
8	MT Sembiring , R H Tambunan	2021	Analysis of graduation prediction on time based on student academic performance using the Naïve Bayes Algorithm with data mining implementation (Case study: Department of Industrial Engineering USU)	Algoritma Naïve Bayes digunakan sebagai metode untuk mengklasifikasikan prestasi akademik mahasiswa teknik industri USU. Karakteristik yang digunakan antara lain NIM, Nama, Jenis Kelamin, Daerah Asal, Sekolah Asal, Penerimaan dan IPK. Dari hasil pengujian 173 sampel data dengan algoritma Naïve Bayes, sampel yang terbentuk memiliki akurasi pencocokan sebesar 70,83%. Artinya berguna untuk memprediksi kelulusan siswa. (Sembiring, M. T., & Tambunan, 2021)
9	Sri Wahyuni dan Murni	2020	Implementation of Data Mining In Predicting the	Aplikasi dikembangkan menggunakan algoritma Naive Bayes berbasis jarak

No	Nama	Tahun	Judul	Hasil dan Kesimpulan
	Marbun		Study Period of Student Using the Naïve Bayes Algorithm	terpendek antara dua objek dengan menentukan nilai k. Data yang digunakan sebagai Variabelnya adalah prestasi akademik siswa, IPK mahasiswa baru, rata-rata nilai UN SMA, dari semua jurusan SMA. Hasil penelitian menunjukkan bahwa aplikasi data mining dikembangkan untuk memprediksi waktu belajar bagi mahasiswa di Universitas Pembangunan Panca Budi. (Wahyuni, S., & Marbun, 2020).
10	TS Jaya dan M Yusman	2021	Predicting the Quality of Pineapple Using the Naive Bayes Classifier Method	Pengklasifikasi Naive-Bayes adalah teknik prediksi berdasarkan kriteria probabilitas sederhana dan penerapan teorema Bayes di bawah asumsi independensi yang kuat. Oleh karena itu, model ini efisien sebagai alat analisis, karena memiliki akurasi klasifikasi tinggi dengan kompleksitas

No	Nama	Tahun	Judul	Hasil dan Kesimpulan
				rendah dan dapat mencapai akurasi klasifikasi hingga 75%.(Jaya, T. S., & Yusman, 2020)
11	L Melian dan A Nursikuwagus	2018	Prediction Student Eligibility in Vocation School with Naïve-Bayes Decision Algorithm.	Variabel dihitung menggunakan algoritma Naïve Bayes Decisions. Yaitu variabel hasil ujian akhir, sertifikasi, psikologi, wawancara, dan . kompetensi. Langkah pertama adalah memasukkan setiap variabel ke nilai dalam rentang waktu tertentu. Langkah selanjutnya adalah menghitung nilai probabilitas untuk setiap variabel. Sebuah survei terhadap 270 siswa menghasilkan 199 tes validasi yang sesuai dengan situasi dunia nyata. 96,1% presisi, 99,3% recall. Hasil ini berarti bahwa algoritma dapat menerima keputusan dengan akurasi 7,87%. Nilai inilah yang

No	Nama	Tahun	Judul	Hasil dan Kesimpulan
				menentukan diterima atau tidaknya di sekolah tersebut. (Melian, L., & Nursikuwagus, 2018)
12	Isron Al Miraz Siregar, Elvianna, Nurul Saepul	2018	Sistem Informasi Persediaan Obat dengan Metode Naïve Bayes Pada RSUD Tanjungpinang	Sistem peramalan data obat berdasarkan penjualan obat menggunakan metode Nave Bayes. Pengembangan sistem informasi dalam pengembangan perangkat lunak menggunakan metode Waterfall. Mengembangkan aplikasi dengan menggunakan netbeans 7.3.1 dan MySQL sebagai database. Program tersebut telah diuji dengan menggunakan metode Black Box. Dengan adanya aplikasi ini dapat membantu apotek untuk mengumpulkan data obat secara cepat dan akurat serta dapat membantu memprediksi data obat dengan metode naive bayes (Siregar, I. A. M., & Saepul, 2018)

No	Nama	Tahun	Judul	Hasil dan Kesimpulan
13	Tugiman, dkk	2022	Prediksi Penggunaan Obat Peserta Jaminan Kesehatan Nasional Menggunakan Algoritma Naïve Bayes Classifier	Memprediksi kebutuhan obat pasien peserta JKN selama 5 (lima) tahun. Data obat yang dijadikan sampel penelitian kemudian diolah menggunakan algoritma adalah Naive Bayes Classifier. Berdasarkan pengujian menggunakan metode UAT, sistem dapat diterima dengan baik oleh pengguna dengan skor 78,64% (kriteria baik). (Damayanti, L., 2022)
14	Furqan, M., Wibowo, S.H.,Kom,S.,Kom, M.,& Abdullah,D	2016	Sistem Persediaan Obat Pada Puskesmas Menggunakan Metode Naïve Bayes	Aplikasi Prediksi Persediaan Obat yang dibangun menggunakan bahasa pemrograman Visual Basic 6.0 dengan menggunakan algoritma naïve bayes. (Furqan, M., Wibowo, S. H., Kom, S., Kom, M., & Abdullah, 2016)
15	Vishal Nehete,	2020	Emergency Drug Procurement	Banyak hal yang tidak semuanya terselesaikan melalui uji klinis sebelum

No	Nama	Tahun	Judul	Hasil dan Kesimpulan
	Hitesh Prajapati, Aditi Mahale, Simran Pandita, Shruti Chaudhari		using Data Mining	obat diberikan untuk digunakan pada Rumah Sakit. Untungnya, database ini memberikan kesempatan untuk mensurvei dampak pengobatan dari banyaknya pasien. Namun komponen yang tidak mendukung seperti asosiatif resep, sosio ekonomi, restoratif yang toleran narasi, dan tujuan merekomendasikan suatu obat. Dalam pemeriksaan data tersebut. dampak pengobatan menggunakan tiga set informasi tes. Petunjuk bagi petugas medis kepada klien di mana pun resep dapat ditebus dan diakses.(Nehete, V., Mahale, A., Prajapati, H., Pandita, S., & Chaudhari, 2020)
16	Esha Sahaa, and Pradip Kumar Raya	2019	Modelling and Analysis of Inventory Management Systems in Healthcare: A	Artikel ini mengkategorikan pendekatan pemodelan dan teknik solusi yang ada untuk sistem inventaris dalam layanan kesehatan.

No	Nama	Tahun	Judul	Hasil dan Kesimpulan
			Review and Reflections	Sebagai hasil dari tinjauan literatur, disajikan dalam bentuk kerangka penelitian integratif dan arah penelitian masa depan yang dapat diterapkan pada situasi saat ini.(Saha, E., & Ray, 2019b)
17	Tony Antoniou PhD, Muhammad Mamdani PharmD MPH	2021	Evaluation of machine learning solutions in medicine	Kecerdasan buatan menggambarkan perbedaan antara pengembangan mesin dengan validasi algoritme pembelajaran mesin dan penggunaannya dalam praktik klinis. Oleh karena itu, diperlukan tambahan hasil klinis dan studi implementasi untuk mewujudkan potensi mesin pembelajaran di bidang medis.(Antoniou, T., & Mamdani, 2021)
18	Tarek Abu Zwaida , Chuan Pham and Yvan Beauregard	2021	Optimization of Inventory Management to Prevent Drug Shortages in the	Penelitian ini mempelajari masalah optimalisasi pengisian ulang obat, yang merupakan model umum dalam manajemen persediaan obat di rumah

No	Nama	Tahun	Judul	Hasil dan Kesimpulan
			Hospital Supply Chain	sakit. Peneliti kemudian mempertimbangkan model pembelajaran penguatan mendalam (DRL) yang mengatasi masalah ini dalam konteks solusi online yang secara otomatis dapat membuat permintaan obat untuk mencegah kekurangan obat. Selanjutnya, peneliti menyajikan hasil numerik untuk memverifikasi kinerja dari algoritma yang diusulkan. (Abu Zwaida, T., Pham, C., & Beauregard, 2021)

Berdasarkan beberapa penelitian diatas dapat diketahui bahwa algortima *naïve bayes* digunakan untuk berbagai kasus penelitian yang pada akhir dari tujuannya algoritma Naïve Bayes ini dapat memprediksi dan mengklasifikasi data dengan akurasi tinggi untuk berbagai permasalahan. Pada penelitian ini untuk mempermudah mobilitas pemakaian sistem persediaan obat maka peneliti mengusulkan menggunakan sistem prediksi berbasis website. Dimana website ini nantinya akan dibangun menggunakan bahasa pemrograman Phyton. Proses pemodelan data pada penelitian ini

menggunakan dua jenis Algoritma *Naïve Bayes* yaitu Algoritma *Gaussian Naive Bayes* dan *Multinomial Naive Bayes.*

2.2 Dasar Teori

2.2.1 Pengertian Data Mining

Definisi data mining secara sederhana adalah istilah yang digunakan untuk menjelaskan proses pencarian atau penambangan knowledge dari data yang sangat besar. Menurut analogi, orang mungkin berpikir bahwa istilah data mining adalah sesuatu yang tidak tepat; menambang emas dari bebatuan atau lumpur diacu sebagai 'penambangan emas' dan bukannya penambangan 'batu' atau 'lumpur'. Jadi, data mining barangkali lebih cocok diberi nama 'knowledge mining' atau 'knowledge discovery'. Meskipun ada ketidakcocokan antara makna dan istilah, data mining telah menjadi pilihan bagi komunitas ilmu ini. Banyak nama- nama lain yang ter-asosiasi dengan data mining antara lain 'knowledge extraction', 'pattern analysis', 'data archaeology', 'information harvesting', 'pattern searching', dan 'data dredging'. (Albert Verasius Dian Sano, S.T., 2019)).

Secara teknis, data minig adalah proses yang memanfaatkan teknik-teknik statistik, matematika, dan kecerdasan buatan untuk mengekstrak dan mengidentifikasi informasi dan knowledge selanjutnya (atau pola-pola) yang berasal dari sekumpulan data yang sangat besar. Berbagai macam pola tersebut bisa dalam bentuk aturan bisnis, kesamaan-kesamaan, korelasi, trend, atau model- model prediksi. Kebanyakan literatur mendefinisikan data mining sebagai "proses yang rumit untuk mengidentifikasi pola-pola yang valid, baru, memiliki potensi bermanfaat, dan bisa dipahami, terhadap data yang disimpan di dalam database yang terstruktur", dimana data diorganisir dalam baris-baris yang terstruktur menurut kategori, ordinal/berurutan, dan variable-variabel yang berkesinambungan.

Menurut Gartner Group, data mining adalah proses menemukan

hubungan baru dengan makna, nilai, dan karakteristik dengan mengatur sejumlah besar data yang disimpan dalam media penyimpanan menggunakan teknologi identifikasi standar seperti sistem statistik dan matematika. Data mining adalah kombinasi dari beberapa disiplin ilmu yang menggabungkan proses dari pembelajaran mesin, identifikasi proses, statistik, database dan visi untuk memecahkan masalah pengambilan informasi dalam database besar. Data mining menurut David Hand, Heikki Mannila dan MIT's Padhraic Smyth adalah analisis data (biasanya big data) untuk menemukan hubungan yang jelas dan menyimpulkan dengan cara yang sekarang dipahami dan berguna bagi pemilik data.(Idris, 2019)

Data mining adalah proses penggalian informasi yang berguna dari database yang besar. Data mining juga dapat didefinisikan sebagai ekstraksi informasi dari database yang besar untuk membantu dalam pengambilan keputusan .Data mining atau knowledge discovery in database (KDD) adalah proses mengambil dan menggunakan data untuk menemukan nilai atau hubungan dari kumpulan data besar. Hasil dari proses pengumpulan data tersebut dapat dijadikan sebagai penelitian untuk pengambilan keputusan di masa yang akan datang . Pengertian umum dari data mining adalah proses menemukan informasi tersembunyi yang tidak diketahui melalui big data dalam database, atau media penyimpanan lainnya. Data mining digunakan untuk menganalisis nilai tambah secara manual ke suatu bentuk informasi yang belum diketahui dari database. Informasi tersebut diperoleh dengan mengekstraksi dan mengidentifikasi nilai-nilai yang relevan atau menarik dari data yang ada di database.(Harahap, 2019)

Berdasarkan dari beberapa pengertian diatas, dapat disimpulkan bahwa data mining merupakan teknik untuk menggali informasi yang tersembunyi dalam suatu basis data sehingga ditemukan pola menarik yang sebelumnya tidak diketahui.

Untuk penelitian ini hanya menggunakan 5 tahapan di data

mining sebagai berikut:

1. Seleksi Data (*Data Selection*) Data yang ada pada database sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari database.
2. Transformasi Data (*Data Transformation*) Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam data mining. Beberapa metode data mining membutuhkan format data yang khusus sebelum bisa diaplikasikan. Sebagai contoh beberapa metode standar seperti analisis asosiasi dan clustering hanya bisa menerima input data kategorikal. Karenanya data berupa angka numerik yang berlanjut perlu dibagi-bagi menjadi beberapa interval. Proses ini sering disebut transformasi data.
3. Proses Mining Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.
4. Evaluasi Pola (*Pattern Evaluation*) Berfungsi untuk mengidentifikasi pola-pola menarik ke dalam knowledge based yang ditemukan. Dalam tahap ini hasil dari teknik data mining berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai hipotesa yang ada memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai hipotesa ada beberapa alternatif yang dapat diambil seperti menjadikannya umpan balik untuk memperbaiki proses data mining, mencoba metode data mining lain yang lebih sesuai, atau menerima hasil ini sebagai suatu hasil yang di luar dugaan yang mungkin bermanfaat.
5. Presentasi Pengetahuan (*Knowledge*) Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna. Tahap terakhir dari proses data mining adalah memformulasikan keputusan atau aksi dari hasil analisis yang didapat. Ada kalanya hal ini harus melibatkan orang-orang yang tidak memahami data mining. Oleh

sebab itu presentasi hasil data mining dalam bentuk pengetahuan yang bisa dipahami semua orang adalah satu tahapan yang diperlukan dalam proses data mining. Dalam presentasi ini, visualisasi juga bisa membantu mengkomunikasikan hasil data mining.

2.2.2 Pengertian Klasifikasi

Dalam data mining terdapat beberapa teknik yang memiliki fungsi berbeda-beda yaitu prediksi, estimasi, deskripsi, klasifikasi, pengklasteran dan asosiasi. Klasifikasi merupakan metode yang digunakan untuk menemukan model atau fungsi yang digambarkan dengan perbedaan kelas data atau konsep yang berfungsi untuk memprediksi kelas dari objek yang label sudah diketahui (Annur, 2018). Proses penemuan model (atau fungsi) yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui Algoritma klasifikasi yang banyak digunakan secara luas, yaitu Decision classification trees, Bayesian classifiers Naïve Bayes classifiers, Neural networks, Analisa Statistik, Algoritma Genetika, Rough sets, k-nearest neighbor, Metode Rule Based, Memory based reasoning, dan Support vector machines (Annur, 2018). Dalam klasifikasi akan terjadi 2 proses, proses pertama adalah learning (fase training), dimana algoritma klasifikasi diciptakan untuk menganalisa data training kemudian direpresentasikan ke dalam bentuk rule klasifikasi. Proses kedua adalah klasifikasi, dimana data testing akan dipakai untuk memperkirakan akurasi dari rule klasifikasi (Gorunescu, 2011).

Terdapat komponen yang mendasari proses klasifikasi sebagai berikut (Gorunescu, 2011):

1. Class Variabel dependen atau variabel yang memiliki ketergantungan (terikat) berupa data kategorikal yang

merepresentasikan suatu label yang melekat pada objek.

2. Predictor Variabel dependen variabel yang memiliki ketergantungan (terikat) yang direpresentasikan oleh (data) atribut tertentu.
3. Training dataset Sekumpulan data yang mengandung nilai dari kedua komponen sebelumnya yang dimanfaatkan untuk pemilihan kelas yang tepat berdasarkan predictor.
4. Testing dataset Sekumpulan data baru yang akan melalui proses klasifikasi menggunakan model yang telah ditentukan. Hasil dari klasifikasi tersebut yang kemudian akan dievaluasi keakuratannya

2.2.3 Pengertian Algoritma NaïveBayes

Deng dkk, berpendapat bahwa pengklasifikasi Naïve Bayes dikenal sebagai Desain Bayesian sederhana dan merupakan model probabilitas yang penting dan sukses dalam pelaksanaannya. Terlepas dari gagasan independensi yang hebat, Naïve Bayes Classifier telah terbukti efektif dalam pengklasifikasian pengeditan teks, pengujian medis, dan tata kelola kinerja TI (Moonallika, 2020)

Naïve Bayes merupakan sebuah metode klasifikasi yang berakar pada teorema bayes. Metode pengklasifikasian dengan menggunakan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai teorema bayes. Kegunaan dari algoritma naïve bayes ialah untuk mengklasifikasikan dokumen teks seperti teks berita maupun teks akademis, sebagai metode machine learning yang menggunakan probabilitas, untuk membuat diagnosis medis secara otomatis, dan mendetekasi atau menyaring spam.

Algoritma ini bekerja berdasarkan prinsip probabilitas bersyarat, seperti yang diberikan oleh Teorema Bayes. Teorema Bayes menemukan probabilitas atau kemungkinan suatu peristiwa

akan terjadi dengan memberikan probabilitas peristiwa lain yang telah terjadi. Dalam istilah yang lebih sederhana, Teorema Bayes adalah metode untuk menemukan probabilitas ketika kita mengetahui probabilitas tertentu lainnya. Bentuk umum atau persamaan dari teorema Bayes (Bustami, 2013)adalah :

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \quad (2.1)$$

Keterangan :

X : Data dengan class yang belum diketahui

H: Hipotesa data X merupakan suatu cass spesifik

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

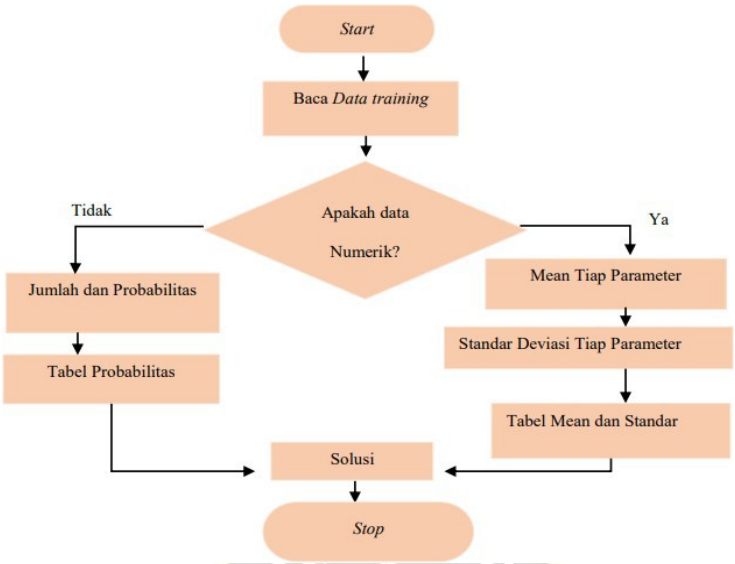
P(H) : Probabilitas hipotesis H (prior probability)

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

P(X) : Probabilitas X

Langkah-langkah algoritma *Naïve Bayes* menurut (Bustami, 2014) sebagai berikut:

- Siapkan dataset.
- Hitung jumlah kelas pada data training.
- Hitung jumlah kasus yang sama dengan kelas yang sama.
- Kalikan semua hasil sesuai dengan data testing yang akan dicari kelasnya.
- Bandingkan hasil perkelas, nilai tertinggi ditetapkan sebagai kelas baru.



Gambar 2.2 Alur Proses algoritma Naïve Bayes, sumber (Bustami., 2014).

Berikut ini merupakan contoh data training sebanyak 14 data dengan output main sepak bola atau tidak. Setiap data ditandai dengan atribut cuaca, temperatur, kelembaban, dan angin yang dapat dilihat pada Tabel 2.2

Tabel 2.2 Contoh Data Training

Cuaca x1	**Temperaturx2**	**Kelembaban x3**	**Angin x4**	**Main atau Tidak**
Cerah	Panas	Tinggi	Kecil	Tidak
Cerah	Panas	Tinggi	Besar	Tidak
Mendung	Panas	Tinggi	Kecil	Ya
Hujan	Sedang	Tinggi	Kecil	Ya
Hujan	Dingin	Normal	Kecil	Ya
Hujan	Dingin	normal	Besar	Tidak
Mendung	Dingin	normal	Besar	Ya
Cerah	Sedang	tinggi	Kecil	Tidak
Cerah	Dingin	normal	Kecil	Ya
Hujan	Sedang	normal	Kecil	Ya

Cerah	Sedang	normal	Besar	Ya
Mendung	Sedang	tinggi	Besar	Ya
Mendung	Panas	normal	Kecil	Ya
Hujan	Sedang	tinggi	Besar	Tidak

Berikut ini merupakan contoh perhitungan metode *Naïve Bayes* yang menggunakan rumus pada persamaan dibawah ini untuk menentukan kelas dari data baru berikut:

(*Cuaca= cerah, Temperatur= dingin, Kelembaban= tinggi, Angin= besar)*

Langkah – langkah perhitungan yang dikerjakan sebagai berikut:

- Mencari *probabilitas prior*

Probabilitas prior ini menyatakan berapa peluang munculnya keputusan main sepak bola *P(C1)* dan peluang keputusan tidak main sepak bola *P(C2).* Misalkan *N* adalah jumlah total keputusan main sepak bola dan tidak, *N1* menyatakan jumlah keputusan main sepak bola dan *N2* menyatakan jumlah keputusan tidak main sepak bola. Berdasarkan pada Tabel 2.2 jumlah total keputusan main sepak bola dan tidak sebanyak 14, jumlah keputusan main sepak bola sebanyak 9, dan jumlah keputusan tidak main sepak bola sebanyak 5. Didapatkan rumus sebagai berikut :

$$P(C_1) = \frac{N_1}{N} \qquad P(C_2) = \frac{N_2}{N} \tag{2.2}$$

$$P(C_1) = \frac{9}{14} \qquad P(C_2) = \frac{5}{14}$$

$$P(C_1) = 0{,}64 \quad P(C_2) = 0{,}36$$

a. Mencari *Probabilitas bersyarat (likelihoad)*

Probabilitas bersyarat (likelihoad), P(X|Ci) ini menyatakan peluang munculnya *X* jika diketahui *Ci* . Misalkan :

X:Cuaca=cerah,Temperatur=dingin,Kelembaban=tinggi,Angin=

besar)

Ci:*C1* adalah main sepak bola dan *C2* adalah tidak main sepak bola.

Sehingga *likelihoad* dapat dihitung dengan mengalikan hasil dari masing– masing nilai probabilitas per-atribut yang berdasarkan pada *Ci*.

Langkah – langkah menghitung *likelihood* :

- Menghitung probabilitas per-atribut

Misalkan menghitung probabilitas Angin = besar dengan Ci adalah *C1* yaitu main sepak bola. Berdasarkan Table 2.2 banyaknya Angin = besar yang keputusannya adalah main sepak bola ada 3 dan banyaknya kelas dengan keputusan main sepak bola ada 9. Maka probabilitasnya adalah :

$$P(\text{angin} = \text{Besar} \mid \text{Main}) = \frac{\textit{jumlah angin besar dengan keputusan main}}{\textit{jumlah seluruh keputusan main}} \quad (2.3)$$

$$= 3/9$$

$$= 0{,}33$$

Melakukan perhitungan dengan cara yang sama pada probabilitas per-atribut yang lain, sehingga diperoleh hasil :

P(Cuaca=Cerah|Main)	=	2/9	=	0,22
P(Cuaca=Cerah|Tidak)	=	3/5	=	0,60
P(Temperatur=dingin | Main)	=	3/9	=	0,33
P(Temperatur=dingin | Tidak)	=	1/5	=	0,20
P(kelembaban=tinggi |Main)	=	3/9	=	0,33
P(kelembaban=tinggi |tidak)	=	4/5	=	0,80
P(Angin=Besar |Tidak)	=	3/5	=	0,60

Mengalikan semua probabilitas per-atribut yang berdasarkan pada *Ci* yaitu *C1* adalah main yang disebut *Likelihoad Ya* atau *C2* adalah tidak main yang disebut *Likelihoad Tidak*.

- *Likelihoad Ya*

  P(X|Main)

  =P(Cuaca=Cerah|Main)*P(Temperatur=dingin|Main)*P(kelembaban=tinggi|Main) * P(Angin=Besar |Main)

  =0,22 * 0,33* 0,33 * 0,33

  = 0.0080

- *Likelihoad Tidak*

  P(X|Tidak)

  =P(Cuaca=Cerah|Tidak)*P(Temperatur=dingin|Tidak*P(kelembaban=tinggi |Tidak) * P(Angin=Besar |Tidak)

  =0.60 * 0,20 * 0,80 * 0,60

  = 0.0576

b. Mengalikan *Likelihoad* dengan *Prior*

   Perkalian *Likelihoad* dengan *Prior* merupakan hal yang penting untuk menemukan *posterior*

   - Perkalian *Likelioad* dengan *Prior* pada keputusan Main (2.4)

     = P(X|Ci) * P(Ci)

     = *P(X|Main) * P(Main)*

     =0.0080 * 0,64

     =0.00512

   - Perkalian *Likelioad* dengan *Prior* Tidak Main

     = P(X|Ci) * P(Ci)

     *= P(X|Tidak) * P(Tidak)*

     =0.0576 * 0,36

=0.020736

c. Menghitung *probabilitas posterior*

Probabilitas posterior, P(Ci|X) ini menyatakan probabilitas keluarnya hasil *Ci* jika diketahui nilai *X* tertentu. *Probabilitas posterior, P(Ci|X)* dicari dengan menggunakan rumus pada persamaan (2.5)

- Probabilitas *Posterior* Main (2.5)

$$P(C_i|X) = \frac{p(X|C_i)P(C_i)}{p(X)}$$

$$posterior = \frac{likelihood \times prior}{evidence}$$

$$posterior = \frac{p(X|Main) \times P(Main)}{P(X|Main) \times P(Main) + P(X|Tidak) \times P(Tidak)}$$

$$posterior = \frac{0{,}00512}{0{,}00512 + 0{,}020736} = 0{,}198019802$$

- Probabilitas *Posterior* Tidak Main (2.6)

$$P(C_i|X) = \frac{p(X|C_i)P(C_i)}{p(X)}$$

$$posterior = \frac{likelihood \times prior}{evidence}$$

$$posterior = \frac{p(X|Tidak) \times P(Tidak)}{P(X|Tidak) \times P(Tidak) + P(X|Main) \times P(Main)}$$

$$posterior = \frac{0{,}020736}{0{,}020736 + 0{,}00512} = 0{,}801980198$$

Nilai *probabilitas posterior* Tidak Main sebesar 0,801980198 lebih besar dari nilai *probabilitas posterior* Main yang bernilai 0,198019802 dan nilai *probabilitas posterior* Tidak Main mendekati nilai 1, sehingga dengan *Naïve Bayes* prediksi dapat disimpulkan Tidak

Main untuk data *input* ini yang berdasarkan pada estimasi probabilitas yang dipelajari dari *data training*.

2.2.4 Teorema Bayes

Bayes merupakan teknik prediksi berbasis probabilistik sederhana yang berdasar pada penerapan teorema Bayes (atau aturan Bayes) dengan asumsi independensi (ketidaktergantungan) yang kuat (naif). Dengan kata lain, dalam Naïve Bayes, model yang digunakan adalah "model fitur independen" (Prasetyo, 2021). Metode Naive Bayes Classifier menggunakan konsep probabilitas yang bertujuan untuk melakukan klasifikasi data pada class tertentu, metode Naive Bayes Classifier merupakan penyederhanaan dari teorema bayes. Prediksi bayes didasarkan pada teorema bayes (Uriawan & Yusfira, 2018). Ide dasar dari aturan Bayes adalah bahwa hasil dari hipotesis atau peristiwa (H) dapat diperkirakan berdasarkan pada beberapa bukti (E) yang diamati. Ada beberapa hal penting dari aturan Bayes tersebut, yaitu:

1. Sebuah probabilitas awal/priori H atau P(H) adalah probabilitas dari suatu hipotesis sebelum bukti diamati.
2. Sebuah probabilitas akhir H atau P(H|E) adalah probabilitas dari suatu hipotesis setelah bukti diamati

Teorema Bayes adalah konsep dalam statistika dan teori probabilitas yang menjelaskan hubungan antara probabilitas awal, peluang, dan probabilitas posterior. Teorema ini memiliki dua penafsiran, yaitu penafsiran Bayes dan penafsiran frekuentis. Konsep teorema Bayes adalah:

- **Prior probability**: Probabilitas awal atau keyakinan awal sebelum mendapatkan data tambahan.
- **Likelihood**: Distribusi probabilitas dari data yang diberikan parameter.
- **Posterior probability**: Distribusi probabilitas yang diperbarui setelah menggabungkan informasi dari prior dan likelihood.

Teorema Bayes dinyatakan secara matematis dalam persamaan berikut:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (2.7)$$

Rumus teorema bayes Dimana P(B) tidak sama dengan 0

- Pada dasarnya, kita mencoba mencari peluang kejadian A, apabila kejadian B bernilai benar. Kejadian B juga disebut sebagai bukti.
- P(A) adalah apriori dari A (probabilitas sebelumnya, yaitu probabilitas peristiwa sebelum bukti terlihat). Bukti adalah nilai atribut dari instance yangtidak diketahui (peristiwa B).
- P(A|B) adalah probabilitas posteriori dari B, yaitu probabilitas kejadian setelah bukti terlihat.
- Ciri utama dari algoritma Naive Bayes Classifier adalah adanya asumsi yg sangat kuat (naif) akan independensi dari masing- masing kondisi / kejadian.

2.2.5 Naïve Bayes Untuk Klasifikasi

Naive Bayesian Classifier merupakan klasifikasi dengan model statistik untuk menghitung peluang dari suatu kelas yang memiliki masing – masing kelompok atribut yang ada, dan menentukan kelas mana yang paling optimal. Pada metode ini semua atribut akan memberikan kontribusinya dalam pengambilan keputusan, dengan bobot atribut yang sama penting dan setiap atribut saling bebas satu sama lain. Dalam sebuah dataset yang besar, pemilihan data training secara random akan menyebabkan kemungkinan adanya nilai nol dalam model probabilitas. Nilai nol ini akan menyebabkan *Naive Bayes Classifier* tidak dapat mengklasifikasi sebuah data inputan. Oleh karena itu diperlukan suatu metode smoothing yang dapat menghindari adanya nilai nol dalam model probabilitas. *Laplacian Smoothing* merupakan metode smoothing yang biasa digunakan dalam *Naive Bayes Classifier*. Laplacian Smoothing biasa dikenal dengan nama add one smoothing, karena dalam perhitungannya, setiap variabel pada masing – masing parameter ditambahkan 1.

NBC adalah metode klasifikasi yang berdasarkan probabilitas dan Teorema Bayesian dengan asumsi bahwa setiap variable X bersifat bebas (independence). Dengan kata lain, NBC mengansumsikan bahwa keberadaan sebuah atribut (variable) tidak ada kaitannya dengan keberadaan atribut (variable) yang lain. Metode NBC menempuh dua tahap dalam proses klasifikasi teks,yaitu tahap pelatihan dan tahap klasifikasi (Indriani, 2014).

Kaitan antara Naïve Bayes dengan klasifikasi, korelasi hipotesis, dan bukti dengan klasifikasi adalah bahwa hipotesis dalam teorema Bayes merupakan label kelas yang menjadi target pemetaan dalam klasifikasi, sedangkan bukti merupakan fitur-fitur yang menjadi masukan dalam model klasifikasi. Jika X adalah vektor masukan yang berisi fitur dan Y adalah label kelas, Naïve Bayes dituliskan dengan P(Y|X). Notasi tersebut berarti probabilitas label kelas Y didapatkan setelah fitur-fitur X diamati. Notasi ini disebut juga probabilitas akhir (posterior probability) untuk Y, sedangkan P(Y) disebut probabilitas awal (prior probability) Y.

Selama proses pelatihan harus dilakukan pembelajaran probabilitas akhir (P(Y|X) pada model untuk setiap kombinasi X dan Y berdasarkan informasi yang didapat dari data latih. Dengan membangun model tersebut, suatu data uji X' dapat diklasifikasikan dengan mencari nilai Y' dengan memaksimalkan nilai P(Y'|X') yang didapat.

2.2.6 Karakteristik Naïve Bayes

Klasifikasi dengan Naïve Bayes bekerja berdasarkan teori probabilitas yang memandang semua fitur dari data sebagai bukti dalam probabilitas. Hal ini memberikan karakteristik Naïve Bayes sebagai berikut (Laroussi, H. M., & AL, 2015).

1. Metode Naïve Bayes teguh (robust) terhadap data-data yang terisolasi yang biasanya merupakan data dengan karakteristik berbeda (outlier). Naïve Bayes juga bisa menangani nilai atribut

yang salah dengan mengabaikan data latih selama proses pembangunan model dan prediksi.

2. Tangguh menghadapi atribut yang tidak relevan.
3. Atribut yang mempunyai korelasi bisa mendegradasi kinerja klasifikasi Naïve Bayes karena asumsi independensi atribut tersebut sudah tidak ada (Laroussi, H. M., & AL, 2015).

2.2.7 Gaussian Naïve Bayes dan Multinomial Naïve Bayes

2.2.7.1 Pengetian Gaussian Naïve Bayes dan Multinomial Naïve Bayes

- Gaussian Naive Bayes (GNB) adalah salah satu varian dari algoritma Naive Bayes yang digunakan dalam pembelajaran mesin untuk masalah klasifikasi dan pemodelan probabilitas. GNB berdasarkan pada Teorema Bayes dan asumsi bahwa setiap fitur dalam dataset mengikuti distribusi Gaussian (normal). Berikut adalah beberapa poin penting yang menjelaskan Gaussian Naive Bayes:
  1. Asumsi Distribusi Gaussian: GNB mengasumsikan bahwa setiap fitur dalam dataset terdistribusi secara Gaussian. Ini berarti bahwa nilai-nilai fitur mengikuti distribusi normal, yang memiliki kurva lonceng (bell curve) dengan nilai rata-rata (mean) dan simpangan baku (standard deviation) sebagai parameter.
  2. Naive Bayes Assumption: GNB juga mengikuti asumsi dasar dari algoritma Naive Bayes, yaitu asumsi naif bahwa semua fitur adalah independen satu sama lain, yang berarti tidak ada korelasi antara fitur-fitur ini. Meskipun asumsi ini sering kali tidak benar dalam dunia nyata, Naive Bayes tetap efektif dan mudah diimplementasikan.
  3. Perhitungan Probabilitas: GNB menggunakan probabilitas dalam pembuatan keputusan klasifikasi. Ini berarti bahwa ia menghitung probabilitas bahwa sebuah instance/data tertentu masuk ke dalam setiap kelas yang mungkin, dan kemudian

memilih kelas dengan probabilitas tertinggi sebagai hasil klasifikasinya.

4. Klasifikasi Sederhana: GNB cocok untuk data yang bersifat kontinu, seperti data numerik yang mengukur sesuatu. Meskipun dapat digunakan untuk klasifikasi, GNB cenderung sederhana dan dapat menghasilkan hasil yang baik pada dataset yang cocok dengan asumsinya.
5. Kasus Penggunaan Umum: GNB sering digunakan dalam aplikasi seperti pengenalan teks, klasifikasi dokumen, dan pemrosesan bahasa alami, meskipun dapat digunakan dalam berbagai domain lainnya.

GNB merupakan algoritma yang cepat dan memiliki tingkat akurasi yang baik ketika asumsinya memadai untuk dataset yang diberikan. Namun, perlu diperhatikan bahwa jika data tidak benar-benar mengikuti distribusi Gaussian atau jika ada keterkaitan yang kuat antara fitur-fitur, maka hasilnya mungkin tidak sesuai dengan kenyataan.

- Multinomial Naive Bayes adalah salah satu varian dari algoritma Naive Bayes yang digunakan dalam pembelajaran mesin untuk masalah klasifikasi, terutama di bidang pemrosesan bahasa alami dan pengolahan teks. Algoritma ini dirancang khusus untuk mengatasi data yang direpresentasikan sebagai data frekuensi, seperti dokumen teks di mana setiap fitur mewakili jumlah kemunculan kata dalam dokumen. Berikut adalah penjelasan mengenai Multinomial Naive Bayes:

  1. **Variabel Kategorikal:** Multinomial Naive Bayes cocok untuk kasus di mana variabel yang diamati adalah diskrit dan kategorikal, terutama dalam konteks pengolahan teks. Ini termasuk data yang dijelaskan sebagai frekuensi kata-kata, token, atau kategori lain dalam dokumen.
  2. **Frekuensi Kemunculan:** Algoritma ini bekerja dengan

menghitung frekuensi kemunculan setiap kata atau token dalam dokumen atau dataset. Ini membuatnya sangat sesuai untuk analisis teks, seperti klasifikasi teks, analisis sentimen, pengelompokan dokumen, dan lain-lain.

3. **Asumsi Naive Bayes:** Multinomial Naive Bayes juga mengikuti asumsi dasar algoritma Naive Bayes, yaitu asumsi bahwa setiap fitur adalah independen. Dalam konteks teks, ini berarti bahwa munculnya kata-kata atau token dalam suatu dokumen tidak dipengaruhi oleh munculnya kata-kata atau token lainnya. Meskipun asumsi ini sering kali tidak sepenuhnya benar, Multinomial Naive Bayes masih efektif dalam banyak kasus pemrosesan teks.
4. **Perhitungan Probabilitas:** Multinomial Naive Bayes menggunakan probabilitas untuk membuat keputusan klasifikasi. Ini menghitung probabilitas bahwa suatu dokumen atau data masuk ke dalam setiap kategori yang mungkin berdasarkan frekuensi kemunculan kata-kata atau token. Kategori dengan probabilitas tertinggi akan dipilih sebagai hasil klasifikasi.
5. **Aplikasi Utama:** Multinomial Naive Bayes sangat umum digunakan dalam aplikasi NLP, termasuk klasifikasi teks, analisis sentimen, pengelompokan dokumen, pengenalan bahasa, dan banyak lagi. Hal ini terutama karena kenyataan bahwa kata-kata atau token sering digunakan sebagai fitur dalam analisis teks.

Meskipun algoritma ini sangat cocok untuk pemrosesan dokumen namun algoritma ini juga cocok untuk perhitungan data kategor (Rinaldi & Goejantoro, 2021) Sehingga pada penelitian ini peneliti mencoba menggunakan algoritma Multinomial Naive Bayes untuk pemodelan data.

2.2.7.2 Perbedaan Algoritma Naïve Bayes biasa dengan Gaussian Naïve Bayes dan Multinomial Naïve Bayes

Perbedaan utama antara algoritma Gaussian Naïve Bayes, Multinomial Naïve Bayes , dan Algoritma Naive Bayes biasa adalah:

- Gaussian Naive Bayes:

- Mengasumsikan fitur-fitur mengikuti distribusi Gaussian (normal). Hal ini berguna ketika fitur-fiturnya bersifat kontinu.
- Menghitung rata-rata dan varians setiap fitur untuk setiap kelas untuk memodelkan distribusi Gaussian.
- Cocok digunakan ketika nilai fitur berupa angka dan mengikuti distribusi normal.

- Multinomial Naive Bayes:

- Mengasumsikan nilai fitur berupa hitungan, seperti frekuensi kata dalam dokumen.
- Memodelkan distribusi multinomial dari nilai fitur untuk setiap kelas.
- Sesuai untuk tugas klasifikasi teks di mana fitur-fiturnya adalah jumlah kata.

- Naive Bayes biasa:

- Membuat asumsi penyederhanaan bahwa fitur-fiturnya independen dari kelasnya.
- Tidak membuat asumsi khusus tentang distribusi fitur yang mendasarinya.
- Dapat bekerja dengan fitur kontinu dan diskrit, tetapi mungkin tidak bekerja sebaik Gaussian atau Multinomial Naive Bayes ketika distribusi fitur tidak sesuai dengan asumsi.

Pilihan di antara varian-varian ini tergantung pada karakteristik data dan masalah yang dihadapi. Gaussian baik untuk fitur kontinu, Multinomial untuk fitur hitungan, dan Naive Bayes biasa ketika distribusi fitur tidak diketahui.

2.2.8 Pengertian Prediksi

Prediksi adalah suatu proses untuk memperkirakan secara

sistematis tentang sesuatu yang mungkin akan terjadi di masa mendatang berdasarkan data atau informasi di masa lalu dan sekarang. Prediksi tidak harus memberikan jawaban secara pasti kejadian yang akan terjadi, melainkan berusaha untuk mencari jawaban sedekat mungkin yang akan terjadi (Rianto, 2022)

Pengertian prediksi sama dengan peramalan atau *forecasting*. Menurut Kamus Besar Bahasa Indonesia, prediksi merupakan hasil dari kegiatan meramal, memperkirakan nilai pada masa yang akan datang dengan menggunakan data masa lalu. Peramalan atau *forecasting* adalah suatu prosedur untuk membuat informasi factual tentang situasi sosial masa depan atas dasar informasi yang telah ada tentang masalah kebijakan. Ramalan mempunyai tiga bentuk utama yaitu proyeksi, prediksi dan perkiraan. (Rianto, 2022).

Pengertian lain prediksi adalah suatu proses memperkirakan secara sistematis tentang sesuatu yang paling mungkin terjadi di masa depan berdasarkan informasi masa lalu dan sekarang yang dimiliki, agar kesalahannya (selisih antara sesuatu yang terjadi dengan hasil perkiraan) dapat diperkecil (Setyowanto., 2018) Dapat disimpulkan bahwa prediksi adalah suatu proses memperkirakan secara sistematis tentang sesuatu yang paling mungkin terjadi di masa depan berdasarkan informasi masa lalu dan sekarang yang dimiliki dengan tujuan agar kesalahan, selisih antara sesuatu yang terjadi dengan hasil perkiraan, dapat diperkecil. Istilah prediksi sama dengan ramalan atau perkiraan, Prediksi mempunyai tiga bentuk utama: proyeksi, prediksi, dan perkiraan.

1. Suatu proyeksi adalah ramalan yang didasarkan pada ekstrapolasi atas kecenderungan masa lalu maupun masa kini ke masa depan.Proyeksi membuat pertanyaan yang tegas berdasarkan argument yang diperoleh dari motode tertentu dan kasus yang paralel.
2. Sebuah prediksi adalah ramalan yang didasarkan pada asumsi teoritik yang tegas.Asumsi ini dapat berbentuk hokum teoretis

(misalnya hokum berkurangnya nilai uang), proposisi teoritis (misalnya proposisi bahwa pecahnya masyarakat sipil diakibatkan oleh kesenjangan antara harapan dan kemampuan), atau analogi (misalnya analogi antara pertumbuhan organisasi pemerintah dengan pertumbuhan organisme biologis).

3. Suatu perkiraan (conjecture) adalah ramalan yang didasarkan pada penilaian yang informative atau penilaian pakar tentang situasi masyarakat masa depan. Tujuan dari pada diadakannya peramalan kebijakan adalah untuk memperoleh informasi mengenai perubahan dimasa yang akan dating yang akan mempengaruhi terhadap implementasi kebijakan serta konsekuensinya.

2.2.9 Confusion Matrix

Kinerja sistem klasifikasi menggambarkan seberapa baik sistem dalam mengklasifikasikan data. *Confusion matrix* merupakan salah satu metode yang dapat digunakan untuk mengukur kinerja suatu metode klasifikasi. Pada dasarnya *Confusion Matrix* mengandung informasi yang membandingkan hasil klasifikasi yang dilakukan oleh sistem (prediksi) dengan hasil klasifikasi yang seharusnya. *Confusion matrix* merupakan sebuah tabel yang berisi jumlah banyaknya test record yang diklasifikasi secara benar atau salah (Gorunescu, 2011). Perhitungan kinerja model dalam *Confusion Matrix* berdasarkan pada nilai True Positif (TP), True Negatif (TN), False Positif (FP), False Negatif (FN). Bentuk dari *Confusion Matrix* dilampirkan pada tabel 2.3 dibawah ini :

Tabel 2.3 *Confusion Matrix*

Aktual	Prediksi	
	True	False
True	TP	FN
False	FP	TN

1. Menghitung nilai Akurasi

Nilai akurasi menggambarkan seberapa akurat sistem dapat mengklasifikasikan data secara benar. Nilai akurasi merupakan perbandingan antara data yang diprediksi benar dengan keseluruhan data.

Perhitungan nilai akurasi dilakukan dengan persamaan berikut ini :

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (2.8)$$

2. Menghitung nilai Presisi

Presisi adalah nilai dari ketepatan dari metode yang digunakan dalam klasifikasi. Nilai tersebut menunjukkan banyaknya data yang dapat terklasifikasi di kelas yang benar dalam beberapa pengujian. (Kiding, 2021)

Perhitungan nilai presisi dilakukan dengan persamaan berikut ini :

$$\boldsymbol{Presisi = \frac{TP}{TP+FP} \times 100\%} \quad (2.9)$$

3. Menghitung nilai *Recall*

Recall adalah nilai yang dapat mengukur hasil berapa presentase data yang terklasifikasikan dengan benar. (Kiding, 2021)

Perhitungan nilai *recall* dilakukan dengan persamaan berikut ini :

$$\boldsymbol{Recall = \frac{TP}{TP+FN} \times 100\%} \quad (2.10)$$

4. Menghitung *F-1 Score*

F-1 Score adalah kombinasi rata-rata nilai *presisi* dan *recall* yang berbanding lurus dengan nilai keduanya. (Ramdhani, S. L., Andreswari, R., & Hasibuan, 2018).

Perhitungan nilai *F1-Score* dilakukan dengan persamaan berikut ini :

$$\boldsymbol{F1\ score = \frac{2 \times presisi \times recall}{presisi + recall}} \quad (2.11)$$

Dimana :

TP : Jumlah data positif yang terklasifikasi dengan benar oleh sistem

TN : Jumlah data negatif yang terklasifikasi dengan benar oleh sistem

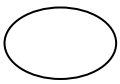
FN : Jumlah data negatif yang terklasifikasi dengan salah oleh sistem

FP : Jumlah data positif yang terklasifikasi dengan salah oleh sistem

2.2.10 Diagram Alir Data (Data Flow Diagram/DFD)

Diagram alir data atau disebut juga Data Flow Diagram (DFD) merupakan salah satu alat yang dapat digunakan untuk memodelkan sistem. Diagram alir yang paling utama disebut dengan diagram konteks yang berada pada level tertinggi (top level) dan dapat dikembangkan sehingga memiliki diagram alir dT level berikutnya seperti DFD level 1, 2, dan seterusnya dengan maksud menampilkan lebih rinci proses-proses yang terjadi dalam sistem. Diagram alir data digambarkan dengan simbol-simbol yaitu simbol ensitas/terminator aliran data (data flow). Proses dan penyimpanan data seperti ditunjukkan pada tabel dibawah ini :

Tabel 2.4 Daftar simbol DFD

Simbol	Keterangan
	Entitas (terminator)
	Aliran data (Data Flow)
Atau  	Proses
	Penyimpanan data

Entitas / terminator dapat berupa orang ,organisasi atau bagian yang ditandai dengan pemberian label pada simbolnya. Entitas dapat memberi dan menerima data dari sistem. Aliran data merupakan penghubung dari satu titik ke titik lainnya. Proses menggambar halaman yang terjadi didalam sistemdengan cara menerima input, melakukan pemprosesan input, lalu menghasilkan output. Penanaman proses dapat berupa nama sistem, subsistem, dan menggunakan kata kerja. Penyimpanan data (data store) dapat berupa penyimpanan manual atau basis data yang terkomputerisasi. (Soulfitri, 2019).