BAB 2
LANDASAN TEORI

Bab ini memaparkan tentang penelitian-penelitian terdahulu yang berhubungan dengan topik skripsi. Bab 2 ini juga menjelaskan teori-teori yang digunakan oleh penulis dalam pembuatan tugas akhir/skrpsi.

2.1 Penelitian Terdahulu

Penelitian yang diusulkan oleh penulis dengan judul "Implementasi Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Beasiswa Kip-K Di Unipdu Berbasis Web" merupakan sebuah studi lanjutan yang didasarkan pada beberapa penelitian sebelumnya yang telah dilakukan oleh para peneliti terdahulu dengan topik yang sejenis.

Beberapa penelitian terkait yang menggunakan metode Naïve Bayes ialah penelitian pada jurnal Fakultas Teknik UNISLA oleh Alief Ulfa Kurnia, yang berjudul "Sistem Pendukung Keputusan Penerimaan Beasiswa Menggunakan Metode Naive Bayes" menunjukan bahwa sistem pendukung keputusan penerimaan beasiswa menggunakan metode Naive Bayes efektif dalam menyeleksi mahasiswa sesuai kategori yang ditetapkan. Proses pengambilan keputusan sangat bergantung pada dataset yang digunakan, dan sistem ini mencapai tingkat akurasi sebesar 92,7% dengan error 7,3% berdasarkan perhitungan dari 110 data (Kurnia dkk., 2020).

Penelitian oleh Ahmad Misbachudin Riyadi pada *Journal of Engineering and Informatic* yang berjudul "Klasifikasi Penerima Beasiswa Menggunakan Metode Naïve Bayes (Studi Kasus SMP Negeri3 Selomerto)" dari 186 data siswa yang terdiri atas 150 data untuk pelatihan dan 36 data untuk pengujian, diperoleh tingkat akurasi yang cukup tinggi, yaitu sebesar 91,67%, baik melalui perhitungan manual maupun dengan menggunakan aplikasi RapidMiner. Dengan demikian,

dapat disimpulkan bahwa algoritma Naïve Bayes sangat tepat diterapkan untuk mendukung pengklasifikasian penerima beasiswa, sehingga proses pemberian beasiswa dapat dilaksanakan dengan lebih cepat dan akurat (Misbachudin Riyadi dkk., 2023).

Penelitian oleh (Iskandar dkk., 2021) pada jurnal Karismatika yang berjudul “Metode Naive Bayes Classifer Dalam Penentuan Penerima Beasiswa Bidikmisi Di Universitas Negeri Medan” Pengujian dengan perbandingan data training dan data testing sebesar 80:20 menghasilkan akurasi tertinggi sebesar 79%. Dari total 318 mahasiswa, sebanyak 250 mahasiswa dikategorikan layak menerima beasiswa, sementara 68 mahasiswa dikategorikan tidak layak.

Penelitian oleh(H. Muslim dkk., 2023) pada jurnal Ilmiah Teknik dan Ilmu Komputer yang berjudul “Studi Komparasi Algoritma Naïve Bayes Dan K-Nn Untuk Klasifikasi Penerimaan Beasiswa Di Mi Al-Islamiyah Karangsawah” Dari 186 data siswa yang terdiri atas 150 data training dan 36 data testing, hasil klasifikasi menggunakan algoritma Naïve Bayes dan K-Nearest Neighbor masing-masing menghasilkan tingkat akurasi sebesar 91,67% dan 75,00%. Berdasarkan nilai akurasi yang diperoleh dari kedua algoritma tersebut, algoritma Naïve Bayes terbukti lebih unggul dalam klasifikasi penerimaan beasiswa dibandingkan dengan algoritma K-Nearest Neighbor.

Penelitian oleh (Sumiah & Mirantika, 2020) pada jurnal Buffer Informatika yang berjudul “Perbandingan Metode K-Nearest Neighbor Dan Naive Bayes Untuk Rekomendasi Penentuan Mahasiswa Penerima Beasiswa Pada Universitas Kuningan” menunjukan bahwa hasil pengukuran tingkat akurasi algoritma K-Nearest Neighbor menunjukkan akurasi sebesar 100%, sedangkan algoritma Naive Bayes memperoleh akurasi sebesar 99,89%. Hal ini mengindikasikan bahwa pada data penerimaan beasiswa, tingkat akurasi K-Nearest Neighbor lebih tinggi dibandingkan dengan Naive Bayes.

2.2 Perbandingan Penelitian Terdahulu

Penelitian yang diusulkan oleh penulis dengan judul "Implementasi Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Beasiswa Kip-K Di Unipdu Berbasis Web" memiliki beberapa kesamaan serta perbedaan dengan penelitian-penelitian terdahulu yang telah dilakukan. Berdasarkan keterangan 2.1, beberapa kesamaan antara penelitian terdahulu dan penelitian yang dilakukan oleh penulis dapat diidentifikasi. Berikut ini adalah tabel yang memperlihatkan perbandingan antara penelitian terdahulu dan penelitian yang sedang dilakukan oleh penulis berdasarkan judul penelitian.

Tabel 2. 1 Penelitian Terdahulu

NO	Judul dan Penulis	PERSAMAAN	PERBEDAAN	HASIL
1	Sist Pendukung Kep usan Per maan Beasiswa Me unakan Metode Nai Bayes (Kurnia dkk 020)	• Metode Naïve Bayes • SPK	• Atribut : IPK, Surat Keterangan Dekan, Surat keterangan Tidak Mampu, KTP, Pekerjaan Orang Tua, Jumlah Tanggungan Orang Tua, Surat keterangan NU, Rekening Listrik, Foto Rumah • Data Klasifikasi	Nilai akurasi 92,7% dengan eror 7,3%
2	Kla kasi Penerima Bea wa Me unakan Metode Nai Bayes (Mis chudin Riyadi dkk 023)	• Metode Naïve Bayes	• Atribut : Nama, Jenis Kelamin, Jarak rumah ke sekolah, Penghasilan orang tua, Tanggungan orang tua, kelengkapan keluarga (Yatim/ Piatu), Pekerjaan orang tua dan Kelayakan.	Nilai akurasi 91,67%

NO	Judul dan Penulis	PERSAMAAN	PERBEDAAN	HASIL
			• Data Beasiswa SMP Negeri 3 Selomerto • *Software Rapidminer*	
3	Metode Naive Bayes Classifier Dalam Pertimbangan Penerima Beasiswa Bidikmisi Di Universitas Negeri Medan (Iskandar dkk., 2021)	• Metode Naïve Bayes	• Atribut : Nilai Rata-rata UN, Jumlah Anak, Pekerjaan Orangtua, Penghasilan Orangtua dan IPK • Data Mahasiswa UNM • SPK	Nilai akurasi sebesar 79%
4	Studi Komparasi Algoritma Naïve Bayes Dan K-Nn Untuk Klasifikasi Penerimaan Beasiswa Di Mi Al-Islamiyah	• Metode Naïve Bayes • SPK	• Atribut : Nama, Jenis Kelamin, Jarak rumah ke sekolah, Penghasilan orang tua, Tanggungan orang tua, kelengkapan keluarga (yatim/piatu),	Tingkat akurasi Naïve Bayes 91,67%, dan KNN 75,00%

NO	Judul dan Penulis	PERSAMAAN	PERBEDAAN	HASIL
	Kan gsawah (H. Mus dkk., 2023)		pekerjaan orang tua dan kelayakan. • Data Beasiswa Mi Al – Islamiyah • Metode Algoritma KNN • SPK	
5	Per ndingan Metode K-N rest Neighbor Dan Naive Bayes Unt Rekomendasi Per tuan Mahasiswa Per ma Beasiswa Pad Universitas Kur gan (Sumiah & Mir ka, 2020)	• Metode Naïve Bayes • SPK	• Atribut : Nim, Nama mahasiswa, Program studi, IPK, Pekerjaan orang tua, Jumlah tanggungan Penghasilan orang tua, Status_beasisw. • Data beasiswa Uniku • Metode KNN	Nilai akurasi KNN sebesar 100%, dan Naïve Bayes 99,89%

2.3 Kajian Pustaka

Bagian ini memuat rangkuman teori-teori yang diambil dari buku, jurnal ilmiah atau artikel ilmiah yang mendukung penelitian, serta memuat penjelasan tentang konsep dan prinsip dasar yang diperlukan untuk pemecahan permasalahan.

2.3.1.Sistem Pendukung Keputusan

a. Pengertian Sistem Pendukung Keputusan

Sistem Pendukung Keputusan (SPK) atau *Decision Support System* (DSS) adalah sistem berbasis komputer yang membantu dalam proses pengambilan keputusan dengan memanfaatkan data, model, dan teknologi untuk memecahkan masalah. SPK pertama kali diperkenalkan pada awal tahun 1970 oleh Scott Morton dengan istilah "*Management Decision System*.". Sesuai dengan namanya, tujuan penggunaan sistem ini adalah sebagai "*second opinion*" atau "*information sources*" yang dapat digunakan sebagai bahan pertimbangan sebelum menetapkan suatu kebijakan tertentu (Wahyuningsih & Patima, 2018).

Sistem Pendukung Keputusan (SPK) dapat diterapkan di berbagai bidang, seperti bisnis, pemerintahan, kesehatan, pendidikan, dan lain-lain. Beberapa contoh aplikasi SPK yang umum meliputi penentuan kelayakan kredit, evaluasi kinerja karyawan, pengelolaan persediaan, serta perencanaan strategis (Jeperson Hutahaean, Fifto Nugroho, Dahlan Abdullah Kraugusteeliana, 2023). SPK dapat dikembangkan dengan memanfaatkan berbagai teknologi, seperti pemrosesan bahasa alami, data mining, *artificial intelligence, machine learning*, dan lain sebagainya (Jeperson Hutahaean, Fifto Nugroho, Dahlan Abdullah Kraugusteeliana, 2023).

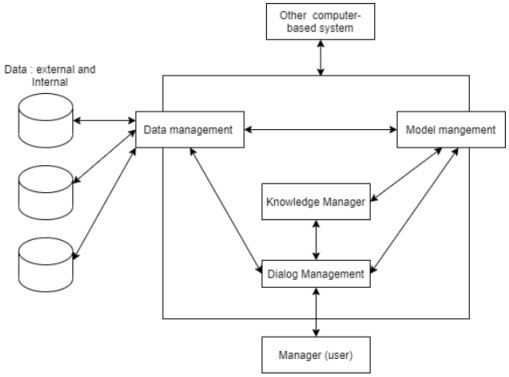
b. Tujuan Sistem Pendukung Keputusan

Dalam proses pengolahannya, *Decision Support System* (DSS) didukung oleh berbagai *system* lain, seperti Kecerdasan Buatan (AI), Sistem Pakar (*Expert System*), *Fuzzy Logic*, dan lain-lain. Dengan demikian, tujuan dari penerapan Sistem Pendukung Keputusan (SPK) ini adalah sebagai berikut:

1) Membantu dalam menyelesaikan permasalahan yang bersifat semi-struktural.
2) Mampu mendukung aktivitas manajer dalam mengambil keputusan terkait suatu masalah.
3) Mampu meningkatkan keefektifan pengambilan keputusan, bukan sekadar tingkat efisiensi.

c. Komponen Sistem Pendukung Keputusan

Dalam Sistem Pendukung Keputusan (SPK) terdapat setidaknya 4 (empat) sub sistem komponen utama, seperti yang ditunjukkan pada gambar berikut (Argistiawan, 2021):



Gambar 2. 1 Komponen SPK

1) Subsistem Manajemen Data

Mencakup basis data yang memiliki data relevan untuk situasi tertentu, yang dikelola oleh perangkat lunak yang dikenal sebagai *Database Management System* (DBMS). Subsistem manajemen data ini dapat terhubung

dengan data warehouse perusahaan. Umumnya, data disimpan atau diakses melalui database pada server web.

2) Subsistem Manajemen Model

Merupakan salah satu keunggulan dalam sistem pendukung keputusan, yang memungkinkan integrasi antara akses data dan model-model keputusan. Subsistem ini terdiri dari paket perangkat lunak yang berisi model-model finansial, statistik, ilmu manajemen, atau model kuantitatif lainnya yang menyediakan kemampuan analisis dan manajemen perangkat lunak yang sesuai.

3) Subsistem Manajemen Dialog

Memungkinkan pengguna untuk berinteraksi dan memberikan perintah kepada Sistem Pendukung Keputusan melalui sistem tersebut. Dengan kata lain, sistem ini menyediakan antarmuka (*interface*). Istilah *User Interface* mencakup seluruh aspek komunikasi antara pengguna dan DSS, yang tidak hanya melibatkan perangkat keras dan perangkat lunak, tetapi juga faktor-faktor yang berkaitan dengan kemudahan penggunaan, aksesibilitas, serta interaksi antara manusia dan mesin.

4) Knowledge Management System

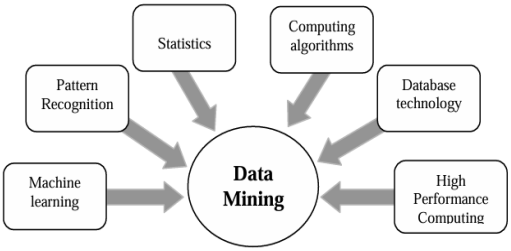
Subsistem ini dapat mendukung subsistem lain atau berfungsi secara independen. Masalah yang kompleks dan tidak terstruktur sering kali memerlukan keahlian, yang dapat diberikan oleh sistem pakar atau sistem cerdas lainnya. Oleh karena itu, DSS yang lebih canggih dilengkapi dengan komponen manajemen berbasis pengetahuan, yang menyediakan keahlian dan pengetahuan untuk meningkatkan kualitas komponen DSS lainnya.

2.3.2.Data Mining

a. Pengertian Data Mining

Data mining mulai dikenal sejak tahun 1990-an, ketika pekerjaan yang memanfaatkan data menjadi semakin penting dalam berbagai bidang, seperti pemasaran dan bisnis, ilmu pengetahuan dan teknik, serta seni dan hiburan(A. Muslim, 2019). Data mining, atau *Knowledge Discovery in Database* (KDD), adalah proses pengambilan informasi tersembunyi yang sebelumnya tidak diketahui. Secara umum, data mining bertujuan untuk mengekstrak kumpulan data guna menemukan informasi yang tidak terduga. Proses ini mampu menangani data berskala besar dan memanfaatkan data dari pengalaman masa lalu untuk meningkatkan model pembelajaran, seperti dalam penerapan klasifikasi (Ginantra dkk., 2021).

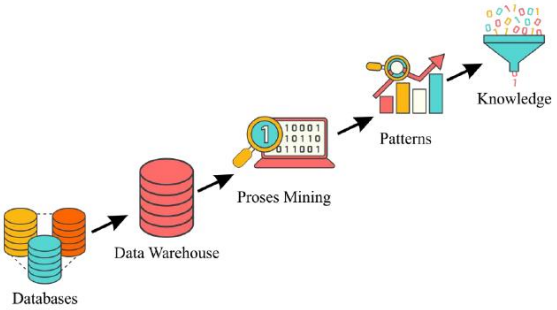
Data mining memiliki hubungan dengan berbagai bidang, antara lain statistik, *machine learning* (pembelajaran mesin), *pattern recognition*, algoritma komputasi, teknologi basis data, dan komputasi berkinerja tinggi. Diagram hubungan antara data mining dan bidang-bidang tersebut dapat dilihat pada Gambar berikut ini (A. Muslim, 2019) :



Gambar 2. 2 Diagram hubungan data mining

b. Tahapan proses data mining

Tahapan proses data mining terdiri dari beberapa langkah yang sesuai dengan proses *Knowledge Discovery in Database* (KDD), sebagaimana dijelaskan pada Gambar 3 dibawah ini (A. Muslim, 2019):



Gambar 2. 3 Proses KDD

1) *Databases*: Langkah pertama dalam alur ini dimulai dari kumpulan data yang tersimpan di berbagai database (basis data). Ini adalah tempat penyimpanan data mentah dari berbagai sumber.
2) Data *Warehouse*: Data dari berbagai database dikumpulkan dan diorganisir dalam sebuah data *warehouse* (gudang data), yang merupakan tempat penyimpanan terpusat untuk analisis data. Gudang data ini membantu menyatukan data dalam format yang lebih konsisten dan terstruktur.
3) Proses Mining: Setelah data tersimpan dalam data warehouse, data tersebut dieksplorasi menggunakan teknik process mining (penambangan proses). Proses mining bertujuan untuk menggali pola-pola dari data yang dapat diolah lebih lanjut.
4) *Patterns*: Hasil dari proses mining berupa pola-pola (*patterns*) yang ditemukan dalam data. Pola ini menggambarkan hubungan atau tren penting yang tersembunyi dalam data mentah.
5) *Knowledge*: Pada tahap akhir, pola-pola ini diinterpretasikan menjadi *knowledge* (pengetahuan) yang dapat digunakan untuk pengambilan keputusan, pengembangan strategi, atau peningkatan proses bisnis.

c. Data mining berdasarkan fungsionalitasnya dikelompokkan menjadi enam bagian yaitu (Ginantra dkk., 2021) :

1) Klasifikasi (*classification*)

Klasifikasi termasuk dalam model *supervised*, di mana kita memiliki sampel data yang digunakan untuk memprediksi beberapa kelas berdasarkan data yang ada. Dalam proses ini, hanya ada satu atribut yang disebut *atribut target*, sedangkan atribut lainnya disebut *atribut predator* atau atribut pendukung dalam menentukan klasifikasi.

2) Klastering (*clustering*)

Bagian dari model *unsupervised*. Dalam klastering, data yang tidak memiliki label sebelumnya dikelompokkan berdasarkan kesamaan tertentu. Klastering yang diatur dalam struktur hierarkikal akan mendefinisikan taksonomi atau pengelompokan dari data tersebut.

3) Regresi (*regression*)

Suatu fungsi yang digunakan untuk memodelkan data dengan tujuan meminimalkan kesalahan prediksi. Regresi umumnya diterapkan pada data yang bersifat *time series.*

4) Association Rule

Suatu pemodelan yang mengidentifikasi hubungan atau kebergantungan antar item dalam dataset. Fungsi ini sering disebut sebagai "*market basket analysis*," yang bertujuan menemukan relasi atau korelasi antara himpunan item, seperti produk yang sering dibeli bersama di toko.

5) Anomaly detection

Mengidentifikasi data yang tidak umum, seperti outlier atau perubahan deviasi/bias, merupakan langkah penting dalam analisis data. Data tersebut sering kali menunjukkan anomali atau pola yang berbeda dari mayoritas data lainnya, dan oleh karena itu memerlukan investigasi lebih lanjut.

6) *Summarization*

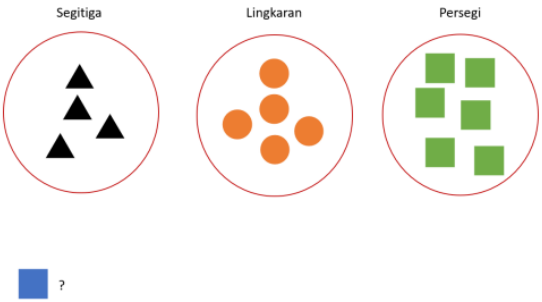
Menyediakan representasi data yang lebih sederhana meliputi pelaporan dan visualisasi data yang bertujuan untuk mendukung pengambilan informasi dan memperkuat Keputusan.

2.3.3.Klasifikasi

Klasifikasi merupakan teknik pengolahan data yang membagi objek ke dalam beberapa kelas berdasarkan jumlah kelas yang telah ditentukan (Rambe, 2022). Klasifikasi dapat diartikan sebagai suatu proses pembelajaran terhadap fungsi target ***f*** yang memetakan setiap vektor (set fitur) ***x*** ke dalam salah satu dari sejumlah label kelas ***y*** yang tersedia. Dalam klasifikasi, disediakan sejumlah data yang disebut training set, yang terdiri dari beberapa atribut, baik kontinyu maupun kategoris, di mana salah satu atribut tersebut menunjukkan kelas dari setiap record. Contoh Blok Model Klasifikasi yang di tujukan pada gambar berikut ini, (A. Muslim, 2019) :

Gambar 2. 4 Blok Model Klasifikasi

Teknik ini mirip dengan teknik *clustering* yang membagi data ke dalam kelompok. Namun, pada teknik klasifikasi, kelompok tersebut disebut kelas data. Perbedaan utama dengan teknik *clustering* adalah adanya *training* set yang telah dilengkapi dengan label. Sebuah model dibangun menggunakan *training* set berlabel ini agar mampu melakukan klasifikasi terhadap data baru dan memprediksi kelas dari objek data yang labelnya belum diketahui. Ilustrasi berikut menggambarkan proses tersebut (Ginantra dkk., 2021) :



Gambar 2. 5 Ilustrasi Teknik Classification

Pada gambar 2, terdapat tiga kelas utama, yaitu segitiga, lingkaran, dan persegi. Setiap kelas ini telah didefinisikan dengan jelas, dan objek data yang ada di dalam training set telah diberi label yang sesuai berdasarkan karakteristik visualnya, seperti bentuk. Proses *classification* memanfaatkan data yang telah diberi label ini untuk membangun sebuah model. Model tersebut mampu mengenali pola-pola yang ada pada data, sehingga ketika model dihadapkan pada objek baru yang belum diberi label, ia dapat memprediksi ke kelas mana objek tersebut termasuk (Ginantra dkk., 2021).

Dalam contoh gambar ini, terdapat sebuah objek berwarna biru yang belum memiliki label. Berdasarkan ciri visualnya yakni bentuknya yang persegi model yang telah dibangun akan mengklasifikasikan objek tersebut ke dalam kelas persegi. Proses ini menggambarkan bagaimana teknik *classification* bekerja dengan menggunakan data latih (*training set*) yang sudah dilabeli untuk membuat prediksi pada data yang belum teridentifikasi(Ginantra dkk., 2021). Berikut beberapa metode algoritma yang dapat digunakan dalam proses klasifikasi (LBS, 2024):

1) *K-Nearest Neighbor Classifier*

   Algoritma ini mengklasifikasikan objek berdasarkan kedekatannya dengan sejumlah objek tetangga terdekat, di mana kelas objek ditentukan oleh mayoritas kelas dari tetangga tersebut.

2) Naive Bayes

   Metode ini menggunakan teorema Bayes dengan asumsi independensi antar fitur, yang sangat efektif untuk dataset besar dan digunakan dalam klasifikasi berbasis teks.

3) *Backpropagation Classification*

   Algoritma ini diterapkan pada jaringan saraf tiruan dengan memperbarui bobot berdasarkan kesalahan prediksi hingga model mencapai hasil yang diinginkan.

4) Pendekatan Sistem Fuzzy

   Klasifikasi dilakukan dengan menentukan derajat keanggotaan suatu objek pada kelas tertentu, memberikan fleksibilitas dalam menghadapi ketidakpastian data.

5) *Decision Tree*

   Algoritma ini menggunakan struktur pohon keputusan untuk memetakan pilihan klasifikasi berdasarkan fitur-fitur dalam data.

2.3.4.Naïve bayes

Naïve Bayes merupakan sebuah metode klasifikasi yang didasarkan pada Teorema Bayes. Metode ini menggunakan pendekatan probabilitas dan statistik yang pertama kali dikemukakan oleh ilmuwan asal Inggris, Thomas Bayes, yang bertujuan untuk memprediksi peluang kejadian di masa depan berdasarkan pengalaman di masa lalu, sehingga dikenal sebagai Teorema Bayes. Algoritma ini mengasumsikan bahwa setiap atribut objek bersifat independen. Probabilitas yang terlibat dalam menghasilkan perkiraan akhir dihitung

berdasarkan jumlah frekuensi dari tabel keputusan "master" (Ginantra dkk., 2021).

Tahapan Algoritma Naïve Bayes, adalah sebagai berikut (Hansen, 2022) :

1. Menghitung jumlah kelas per label.
2. Menghitung jumlah kasus per kelas.
3. Kalikan semua variabel kelas.
4. Bandingkan hasil per kelas.

$$P(C|X) = \frac{P(X|C)P\,(C)}{P(X)} \qquad (1)$$

Berikut keterangan dari rumus Naïve Bayes diatas :

1. **X =** Data dengan class yang belum diektahui.
2. **C =** Hipotesis data merupakan suatu class spesifik.
3. **P (C | X) =** Probabilitas bersyarat C yang diberikan oleh X (*posteriori probability*).
4. **P (X | C) =** Probabilitas bersyarat X yang diberikan oleh C (*likelihood*)
5. **P (C) =** Probabilitas kejadian C (*prior probability*)
6. **P (X) =** Probabilitas kejadian X (*evidence*)

Persamaan 1 mengilustrasikan bahwa probabilitas terjadinya sampel dengan karakteristik tertentu dalam kelas C (*posterior*) merupakan hasil perkalian antara probabilitas kemunculan kelas C sebelum sampel tersebut diperhitungkan (dikenal sebagai prior) dengan probabilitas kemunculan karakteristik sampel pada kelas C (sering disebut *likelihood*), kemudian dibagi dengan probabilitas kemunculan karakteristik sampel secara keseluruhan (dikenal sebagai *evidence*). Oleh karena itu, pada persamaan 1 dapat dituliskan sebagai berikut (Alber, 2021):

$$Posterior = \frac{Prior\ x\ Likelihood}{Evidence} \qquad (2)$$

Nilai *evidence* bersifat konstan untuk setiap kelas pada satu sampel. Nilai posterior tersebut kemudian akan

dibandingkan dengan nilai posterior dari kelas-kelas lainnya guna menentukan kelas mana suatu sampel akan diklasifikasikan.

2.3.5.Gaussian Naïve Bayes

Metode Gaussian Naïve Bayes adalah teknik klasifikasi yang mengasumsikan bahwa data kontinu mengikuti distribusi normal (Gaussian). Selama proses pelatihan, data dibagi menjadi beberapa kelas, dan untuk setiap kelas, dihitung rata-rata dan deviasi standar dari fitur kontinu.

Ketika model Gaussian Naive Bayes digunakan untuk memprediksi kelas data baru, algoritma memanfaatkan distribusi Gaussian yang sudah dihitung sebelumnya untuk memperkirakan probabilitas data kontinu berasal dari masing-masing kelas. Berdasarkan rata-rata dan deviasi standar yang telah dihitung, algoritma menghitung probabilitas data baru untuk setiap kelas. Kelas dengan probabilitas tertinggi dipilih sebagai prediksi kelas yang paling mungkin untuk data tersebut (Ghozali, 2024).

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} . e - \frac{[x-\pi]^2}{2\sigma^2} \qquad (3)$$

Keterangan rumus Gaussian:

1) **P(x)** adalah probabilitas atau fungsi densitas untuk nilai **x**
2) **μ** adalah mean atau rata-rata dari atribut
3) **σ** adalah standar deviasi dari atribut
4) ***e*** adalah bilangan Euler
5) **π**\pi adalah konstanta pi (sekitar 3.14),
6) **x** adalah nilai yang akan dihitung probabilitasnya

Sebelum mengncari nilai gaussian dari atribut, yang pertama adalah membaca data training. Jika data bersifat numerik, maka perlu dilakukan perhitungan nilai rata-rata (mean) dan standar deviasi untuk setiap parameter numerik

tersebut. Sementara itu, apabila data berupa kategori, cari nilai rata-rata (mean) dari data tersebut.

1) Berikut rumus mencari nilai rata-rata (mean)

$$\mu = \frac{\sum_{i=1}^{n} Xi}{n} \qquad (4)$$

Keterangan :
μ : rata-rata hitung (mean)
xi : nilai sampel ke i
n : jumlah sampel

2) Untuk Menghitung Standar Daviasi, dapat dilihat pada rumus dibawah ini :

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu)^2} \qquad (5)$$

Keterangan :
X : nilai fitur asli
μ : nilai rata-rata dari semua data
n : jumlah data
σ : Deviasi standar dari dataset

2.3.6.Confusion Matrix

Confusion matrix merupakan salah satu metode yang digunakan untuk menghitung tingkat akurasi dan tingkat kesalahan (*error rate*). Akurasi didefinisikan sebagai perbandingan antara jumlah kasus yang teridentifikasi dengan benar terhadap total jumlah kasus. Sedangkan, *error rate* adalah perbandingan antara jumlah kasus yang teridentifikasi salah dengan total jumlah kasus(Alber, 2021). Contoh *confusion matrix* untuk klasifikasi dapat dilihat pada Tabel dibawah ini (P. Normawati Dwi, 2021).

Tabel 2. 2 Confusion Matrix

		Kelas Prediksi	
		1	0
Kelas Sebelumnya	1	TP	FN
	0	FP	TN

Keterangan:

1) **TP (*True Positive*) =** jumlah dokumen dari kelas 1 yang benar diklasifikasikan sebagai kelas 1
2) **TN (*True Negative*) =** jumlah dokumen dari kelas 0 yang benar diklasifikasikan sebagai kelas 0
3) **FP (*False Positive*) =** jumlah dokumen dari kelas 0 yang salah diklasifikasikan sebagai kelas 1
4) **FN (*False Negative*) =** jumlah dokumen dari kelas 1yang salah diklasifikasikan sebagai kelas 0.

Confusion Matrix dapat dihitung menggunakan rumus yang melibatkan akurasi, presisi, *recall*, dan *error*, sebagaimana dinyatakan dalam persamaan berikut (Pratiwi dkk., 2021) :

1) Persamaan Akurasi

$$\boldsymbol{Akurasi} = \frac{\boldsymbol{TP+TN}}{\boldsymbol{TP+TN+FP+FN}} \boldsymbol{X}\ \mathbf{100\%} \qquad (6)$$

Nilai akurasi merupakan perbandingan antara jumlah data yang terklasifikasi dengan benar dan total keseluruhan data.

2) Persamaan Presisi

$$\boldsymbol{Presisi} = \frac{\boldsymbol{TP}}{\boldsymbol{TP+FN}} \boldsymbol{X}\ \mathbf{100\%} \qquad (7)$$

Nilai presisi menggambarkan jumlah data kategori positif yang terklasifikasi dengan benar, dibagi dengan total data yang diklasifikasikan sebagai positif.

3) Persamaan *Recall*

$$Recall = \frac{TP}{TP+FN} X\ 100\% \quad (8)$$

Recall menunjukkan persentase data kategori positif yang terklasifikasi dengan benar oleh sistem.

4) Persamaan *F1 Score*

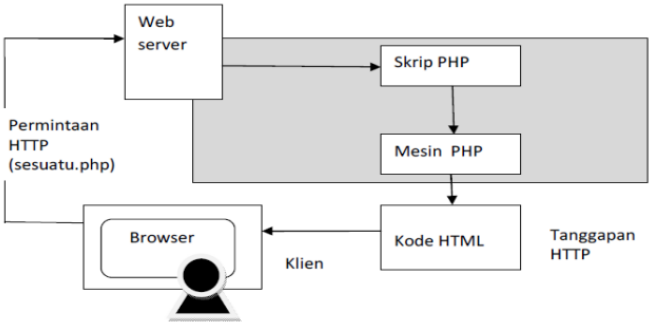
$$F1 = 2 . \frac{Precision\ x\ Recall}{Precision+Recall} \quad (9)$$

F1 Score adalah metrik evaluasi dalam pembelajaran mesin yang mengukur keseimbangan antara *precision* (ketepatan prediksi positif) dan *recall* (kemampuan model mendeteksi seluruh data positif). *F1 Score* dihitung sebagai rata-rata harmonik dari *precision* dan *recall*, sehingga memberikan penilaian yang lebih adil ketika terdapat ketidakseimbangan antara keduanya.

2.3.7.PHP

PHP merupakan bahasa pemrograman server-side scripting yang terintegrasi dengan HTML untuk membangun halaman web dinamis. Sebagai *server-side scripting, sintaks* dan perintah PHP dieksekusi pada server, kemudian hasilnya dikirimkan ke browser dalam format HTML (Muammar, 2019).

Salah satu keunggulan PHP adalah kemampuannya untuk berinteraksi dengan berbagai jenis database. Hal ini memungkinkan implementasi yang mudah dalam menampilkan data dinamis yang diambil dari database, seperti yang ditunjukkan pada gambar dibawah ini (Theodorus Yagoyamu, 2020).



Gambar 2. 6 Skema PHP

2.3.8.UML

UML (*Unified Modeling Language*) dikembangkan oleh Object Management Group pada Januari 1997. Fungsi utama UML adalah menyediakan model visual bagi penggunanya. UML memiliki berbagai jenis diagram, seperti *Use Case Diagram, Activity Diagram, Sequence Diagram*, dan lainnya, yang digunakan untuk memvisualisasikan sistem atau perangkat lunak secara terstruktur dan mudah dipahami (Hansen, 2022).

Menurut (Hansen, 2022) UML merupakan metode untuk menggambarkan perangkat lunak dalam bentuk diagram dan simbol yang mewakili masing-masing fungsinya. UML juga mempermudah programmer dalam memahami dan mengikuti alur kerja perangkat lunak yang sedang dibuat, sehingga dapat memberikan gambaran yang jelas tentang struktur dan interaksi dalam sistem tersebut.

1) Diagram Kelas (*Class Diagram*)

Bersifat statis dan menunjukkan sekumpulan kelas, antarmuka, kolaborasi, serta relasi di antara mereka. Diagram ini sering digunakan dalam pemodelan sistem berorientasi objek untuk memberikan gambaran yang jelas tentang struktur dan hubungan antar elemen dalam sistem.

2) Diagram *Use-Case (Use Case Diagram)*

Bersifat statis dan menggambarkan sekumpulan use-case serta aktor-aktor, yang merupakan jenis khusus dari kelas. Diagram ini sangat penting untuk mengorganisasi dan memodelkan perilaku sistem yang dibutuhkan dan diharapkan oleh pengguna.

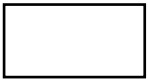
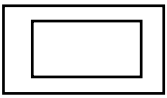
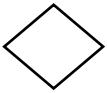
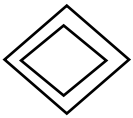
3) Diagram Aktivitas (*Activity Diagram*)

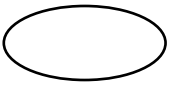
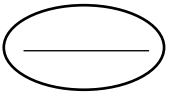
Bersifat dinamis. Diagram ini adalah tipe khusus dari diagram status yang menunjukkan aliran dari satu aktivitas ke aktivitas lainnya dalam suatu *system* (Lysander, 2019).

2.3.9. Entity Relation Diagram (ERD)

ERD, singkatan dari *Entity Relation Diagram*, adalah alat yang digunakan untuk memodelkan struktur data dengan menggambarkan entitas dan hubungan antar entitas. ERD memiliki tiga fungsi, yaitu sebagai alat untuk memodelkan hasil analisis data, memodelkan data konseptual, dan memodelkan objek-objek dalam suatu sistem. ERD terdiri dari beberapa notasi, yaitu Entity, Relationship, dan Attribute (Hansen, 2022).

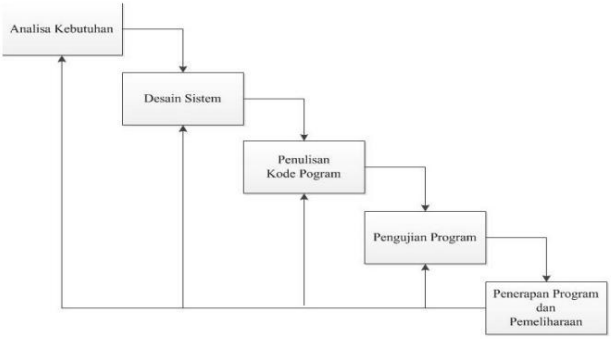
Tabel 2. 3 Simbol *Entity Relationship Diagram (ERD)*

No	Simbol	Nama	Deskripsi
1		*Entity*	*Entity* yang mendeskripsikan tabel.
2		*Weak Entity*	*Entity* yang keberadaannya bergantung pada *entity* lain
3		*Relationship*	*Relationship* yang terjadi di antara satu atau lebih entitas.
4		*Identifying Relationship*	Hubungan yang terjadi antara satu atau lebih weak *entity*

5		*Atribut/Field*	Atribut yang bernilai tunggal atau atribut atomic
6		*Atribut Primary Key*	Satu atau gabungan dari beberapa atribut

2.3.10. Metode Waterfall

Metode air terjun, atau yang lebih dikenal sebagai metode *waterfall*, juga dikenal sebagai siklus hidup klasik atau "*Linear Sequential Model*." Model ini menggambarkan pendekatan yang sistematis dan berurutan dalam pengembangan perangkat lunak. Metode waterfall pertama kali diperkenalkan oleh Winston Royce sekitar tahun 1970. Dalam pendekatan ini, setiap tahap dilakukan secara sistematis dan berurutan, di mana setiap tahap harus menunggu tahap sebelumnya selesai, dan tidak ada kemungkinan untuk kembali ke tahap sebelumnya. Model pengembangan ini bersifat linear, dimulai dari tahap perencanaan hingga tahap pemeliharaan. Gambar 6 menunjukkan siklus pengembangan perangkat lunak dengan metode *Waterfall* (Theodorus Yagoyamu, 2020).



Gambar 2. 7 Alur Metode *Waterfall*

1) Analisis Kebutuhan
   Langkah ini menganalisis kebutuhan sistem melalui penelitian, wawancara, atau studi literatur. Hasilnya adalah

dokumen user requirement yang berisi kebutuhan pengguna, yang nantinya akan menjadi acuan sistem analis untuk mengembangkannya ke dalam bahasa pemrograman sesuai kebutuhan.

2) Desain Sistem
   Tahapan ini merumuskan solusi dengan merancang sistem menggunakan perangkat pemodelan seperti *data flow diagram* (DFD) untuk aliran informasi, *entity relationship diagram* (ERD) untuk hubungan data, serta struktur data yang akan digunakan.
3) Penulisan Kode Program
   Penulisan kode program atau coding merupakan tahap di mana desain sistem diterjemahkan ke dalam bahasa pemrograman yang dapat dikenali oleh komputer. Setelah pengkodean selesai, dilakukan testing untuk menemukan dan memperbaiki kesalahan atau bug sebelum sistem diimplementasikan.
4) Pengujian Program
   Tahapan akhir adalah evaluasi di mana sistem yang baru diuji untuk menilai kemampuan dan keefektifannya. Pada tahap ini, kekurangan dan kelemahan sistem akan teridentifikasi, yang kemudian menjadi dasar untuk melakukan pengkajian ulang dan perbaikan.
5) Penerapan Program dan Pemeliharaan
   Perangkat lunak yang telah diserahkan kepada pelanggan pasti akan mengalami perubahan. Perubahan ini dapat terjadi karena adanya kesalahan yang ditemukan atau karena perangkat lunak perlu menyesuaikan dengan lingkungan baru, seperti perangkat keras (peripheral) atau sistem operasi yang baru. Selain itu, perubahan juga bisa disebabkan oleh kebutuhan pelanggan untuk menambah atau mengembangkan fungsionalitas perangkat lunak.

2.3.11. Pengujian Black Box

Metode *Blackbox Testing* merupakan suatu pendekatan pengujian yang berfokus pada spesifikasi

fungsional perangkat lunak, di mana penguji dapat menentukan serangkaian kondisi input dan melakukan pengujian berdasarkan spesifikasi fungsional program tersebut. Proses *Blackbox Testing* dilakukan dengan menguji program yang telah dikembangkan, yakni dengan mencoba memasukkan data pada setiap formya. Pengujian ini penting untuk memastikan bahwa program berfungsi sesuai dengan kebutuhan perusahaan(Shadiq dkk., 2021).

Terdapat berbagai teknik pengujian dalam Black Box Testing, antara lain (Uminingsih dkk., 2022) :

1) Teknik *Equivalence Partitioning*, metode dengan melakukan partisi atau pembagian input data ke dalam beberapa kategori.
2) Teknik *Boundary Value Analysis*, metode yang bertujuan untuk mendeteksi adanya kesalahan baik dari sisi luar maupun dalam perangkat lunak, dengan menekankan pada nilai minimum dan maksimum dari kesalahan yang ditemukan.
3) Teknik *Fuzzing*, untuk menemukan bug atau gangguan pada perangkat lunak melalui injeksi data yang cacat.
4) Teknik *Cause-Effect Graph*, yaitu metode pengujian yang menggunakan grafik sebagai acuan, di mana grafik tersebut menggambarkan relasi antara efek dan penyebabnya.
5) Teknik *Orthogonal Array Testing*, yaitu metode yang digunakan ketika input domain relatif kecil, namun cukup kompleks untuk diujikan pada skala besar.
6) Teknik *All Pair Testing*, di mana semua pasangan dari test case dirancang agar mencakup setiap kemungkinan kombinasi diskrit dari semua pasangan berdasarkan parameter inputnya, dengan tujuan untuk menguji seluruh pasangan tersebut.
7) Teknik *State Transition*, yang berguna untuk menguji kondisi dari mesin dan navigasi dalam bentuk grafis.

2.3.12. Beasiswa

Beasiswa dalam Kamus Besar Bahasa Indonesia, merupakan tunjangan yang diberikan kepada pelajar atau mahasiswa sebagai bentuk bantuan biaya untuk mendukung proses belajar mereka. Menurut Murniasih (2009), beasiswa diartikan sebagai sebuah bentuk penghargaan yang diberikan kepada individu untuk mendukung mereka melanjutkan pendidikan ke jenjang yang lebih tinggi. Penghargaan tersebut dapat berupa akses ke institusi tertentu atau berupa bantuan keuangan (Adrianto, 2022). Dari berbagai definisi tersebut, dapat disimpulkan bahwa beasiswa adalah bantuan finansial atau penghargaan yang diberikan kepada individu untuk mendukung mereka dalam melanjutkan pendidikan, baik dalam bentuk tunjangan biaya belajar maupun akses ke institusi pendidikan tertentu.

Beasiswa dapat diberikan oleh berbagai pihak, termasuk pemerintah, perusahaan, atau yayasan. Beasiswa ini dapat bersifat cuma-cuma atau diberikan dengan syarat adanya ikatan kerja (sering disebut ikatan dinas) setelah penerima menyelesaikan pendidikannya. Durasi ikatan dinas bervariasi, tergantung pada kebijakan lembaga yang memberikan beasiswa tersebut (Juliawan, 2021). Beasiswa yang diberikan oleh lembaga pemerintah, perusahaan, atau yayasan umumnya terbagi menjadi dua golongan, yaitu beasiswa prestasi dan beasiswa tidak mampu. Pihak pemberi beasiswa akan memberikan beasiswa setelah penerima memenuhi persyaratan dan kriteria yang telah ditetapkan (Juliawan, 2021).

Beberapa tujuan pemberian beasiswa antara lain (Juliawan,2021):

1) Membantu pelajar atau mahasiswa untuk mendapatkan ilmu sesuai bidang yang diminati, terutama bagi mereka yang mengalami kesilitan pembayaran.
2) Mendorong pemerataan pengetahuan atau pendidikan kepada semua yang membutuhkan.
3) Membentuk generasi baru yang lebih cerdas, dengan memberikan kesempatan kepada kaum muda untuk

melanjutkan pendidikan ke jenjang lebih tinggi, sehingga tercipta sumber daya manusia yang mampu menghadapi tantangan zaman.

2.3.13. Kartu Indonesia Pintar (KIP-K)

KIP Kuliah merupakan salah satu program beasiswa dari pemerintah yang ditujukan kepada peserta didik dan mahasiswa dari keluarga kurang mampu untuk mendukung biaya pendidikan. Beasiswa ini merupakan program lanjutan dari Beasiswa Bidikmisi, yang sejak tahun 2020 mengalami perubahan nama menjadi KIP-Kuliah (Safii & Amanda, 2023). KIP Kuliah bertujuan untuk meningkatkan potensi ekonomi serta mobilitas sosial bagi mahasiswa dari keluarga miskin atau rentan miskin agar dapat melanjutkan pendidikan di perguruan tinggi. Beasiswa tersebut diberikan kepada yang berhak menerima berdasarkan klasifikasi, kualitas dan kompetensi penerimanya.

Manfaat utama KIP Kuliah Merdeka 2024 adalah jaminan biaya pendidikan yang dibayarkan langsung kepada perguruan tinggi berdasarkan akreditasi Program Studi (Prodi). Semua perguruan tinggi yang menerima mahasiswa KIP Kuliah Merdeka harus terakreditasi secara resmi dan tercatat dalam sistem akreditasi nasional. Selain itu, mahasiswa penerima KIP Kuliah Merdeka juga mendapatkan bantuan biaya hidup bulanan, yang jumlahnya disesuaikan dengan 5 klaster wilayah, yakni Rp800.000, Rp950.000, Rp1.100.000, Rp1.250.000, dan Rp1.400.000, berdasarkan survei dari Badan Pusat Statistik (Kemdikbud, 2024).

Adapun syarat bagi calon penerima kip-k, dan syaran ekonimi calon penerima kip-k, antara lain sebagai berikut (Kemdikbud, 2024) :

a. Syarat umum bagi calon penerima KIP-K :
   1) Penerima KIP Kuliah Merdeka adalah lulusan Sekolah Menengah Atas (SMA), Sekolah Menengah Kejuruan (SMK), atau sederajat yang lulus pada

tahun berjalan atau maksimal lulus dua tahun sebelumnya.

2) Telah lulus seleksi penerimaan mahasiswa baru melalui semua jalur masuk Perguruan Tinggi Akademik atau Vokasi, baik di Perguruan Tinggi Negeri (PTN) maupun Perguruan Tinggi Swasta (PTS), yang telah terakreditasi beserta Program Studi yang diakui secara resmi dalam sistem akreditasi nasional.
3) Memiliki potensi akademik yang baik namun mengalami keterbatasan ekonomi, atau berasal dari keluarga miskin/rentan miskin, yang dibuktikan dengan dokumen resmi.

b. Syarat Ekonomi Penerima KIP-K :
   1) Mahasiswa pemegang Kartu Indonesia Pintar (KIP) Pendidikan Menengah.
   2) Mahasiswa dari keluarga yang terdaftar dalam Data Terpadu Kesejahteraan Sosial (DTKS) atau menerima program bantuan sosial, seperti: Peserta Program Keluarga Harapan (PKH), Pemegang Kartu Keluarga Sejahtera (KKS).
   3) Termasuk dalam kelompok masyarakat miskin/rentan miskin hingga desil 3 dalam Data Percepatan Penghapusan Kemiskinan Ekstrem (PPKE) yang ditetapkan oleh Kementerian Koordinator Bidang Pembangunan Manusia dan Kebudayaan.
   4) Mahasiswa dari panti sosial atau panti asuhan.

Jika calon penerima tidak memenuhi salah satu dari kriteria di atas, mereka tetap dapat mendaftar KIP Kuliah Merdeka dengan syarat:

1) Bukti pendapatan kotor gabungan orang tua/wali paling banyak Rp4.000.000 per bulan, atau

pendapatan kotor gabungan orang tua/wali dibagi jumlah anggota keluarga paling banyak Rp750.000.

2) Surat Keterangan Tidak Mampu (SKTM) yang dikeluarkan dan dilegalisasi oleh pemerintah desa/kelurahan sebagai bukti bahwa keluarga calon penerima termasuk golongan miskin atau tidak mampu.