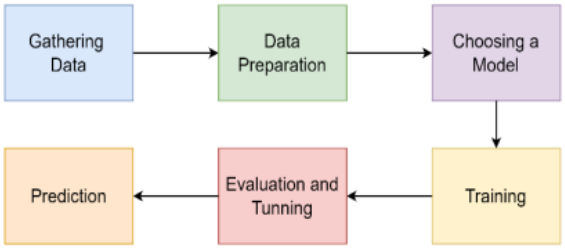
BAB II
LANDASAN TEORI

2.1 Machine Learning

Machine Learning merupakan bagian dari kecerdasan buatan yang secara terstruktur menggunakan algoritma untuk menghasilkan hubungan mendasar antara data dan informasi[3]. Algoritma *machine learning* dibagi menjadi *supervised learning*, *unsupervised learning*, dan *reinforcement learning*. Algoritma *supervised learning* umumnya dibagi menjadi regresi dan klasifikasi. Regresi merupakan metode menggunakan fungsi *continue* untuk menyesuaikan dengan variabel *input* dan variabel *output*. Klasifikasi adalah pencocokan variabel *input* dan kategori diskrit. Algoritma *unsupervised learning* merupakan algoritma dimana *output* yang akan dihasilkan tidak diketahui sebelumnya dimana tidak terdapat label atau hanya terdapat satu label dalam *unsupervised learning*. Dan algoritma *reinforcement learning* adalah model pembelajaran observasi melalui mekanisme *try* an *error*. Hasil yang dianggap baik atau tepat pada algoritma *reinforcement learning* akan disimpan dan menjadi penguat pada *data training* selanjutnya[4].



Gambar 2.1 Proses Machine Learning[5]

Secara luas proses *machine learning* terdiri dari enam proses utama, berikut penjelasannya:

1. *Gathering Data*

   Langkah pengumpulan data menjadi langkah dasar untuk proses *machine learning*. Meskipun merupakan langkah awal, proses ini sangat penting karena kualitas dan kuantitas data dari proses ini akan membantu menentukan model prediksi.

2. *Data Preparation*

   Setelah data terkumpul dari sumber, selanjutnya mempersiapkan data sehingga dapat digunakan untuk proses training *machine learning*. Visualisasi yang relevan pada data dapat dilakukan untuk menemukan hubungan yang relevan antar. Variabel yang berbeda atau menemukan ketidakseimbangan data.

3. *Choosing a Model*

   Proses selanjutnya adalah memilih model yang relevan dengan studi penelitian. Model umumnya dipilih berdasarkan relevansinya terhadap kasus penelitian.

4. *Training*

   Salah satu proses utama dari *machine learning* adalah *training*. Pada proses ini digunakan data dalam perkembangan untuk meningkatkan kemampuan model untuk memprediksi.

5. *Evaluation* dan *Tuning Parameter*

   Setelah proses *training* selesai dilakukan, selanjutnya dilakukan evaluasi untuk memeriksa keakuratannya. Evaluasi akan menggunakan *data test* yang belum pernah digunakan untuk memungkinkan dalam melihat bagaimana kinerja model terhadap data yang belum digunakan. Setelah proses evaluasi, biasanya dilakukan *tuning parameter* untuk melihat apakah berdasarkan parameter yang dimodifikasi akan menyebabkan peningkatan atau penurunan pada hasil evaluasi.

6. *Prediction*

   Proses terakhir dari *machine learning* adalah menggunakan data untuk memberikan jawaban terhadap pertanyaan. Prediksi merupakan proses tujuan untuk menjawab beberapa pertanyaan.

2.1.1 Machine Learning Dalam Pendidikan

Machine learning merupakan mesin yang bisa belajar layaknya manusia, Kerangka kerja *Machine Learning* memerlukan penangkapan dan pemeliharaan serangkaian informasi yang kaya dan mengubahnya menjadi basis pengetahuan terstruktur untuk penggunaan yang berbeda di berbagai bidang, salah satunya di bidang pendidikan[6]. Teknologi ini sangat berperan dalam bidang pendidikan, khususnya dalam analisis data yang digunakan untuk memprediksi hasil belajar, seperti tingkat kelulusan siswa, *Machine Learning* menyediakan metode yang efisien untuk mengolah data besar dalam waktu yang lebih cepat, sehingga memungkinkan pengambilan keputusan berbasis data yang lebih objektif[7]. Penerapan *Machine Learning* dalam mendukung proses evaluasi dan penilaian dalam pendidikan adalah langkah revolusioner yang menjanjikan. *Machine Learning* memungkinkan penggunaan data untuk menghasilkan

wawasan yang lebih dalam tentang kemajuan siswa, mengidentifikasi area-area di mana mereka mungkin mengalami kesulitan, dan merancang kurikulum yang disesuaikan dengan kebutuhan individu[8]. Penerapan *machine learning* dalam pendidikan sangat luas, salah satunya adalah untuk memprediksi tingkat kelulusan siswa. Algoritma *Machine Learning* dapat mengidentifikasi pola-pola yang ada dalam data akademik dan non-akademik siswa, yang pada gilirannya membantu sekolah dalam menentukan kebijakan yang lebih tepat sasaran, seperti program pembinaan atau intervensi untuk siswa yang berisiko tidak lulus.

Dalam konteks pendidikan, *Machine Learning* digunakan untuk menganalisis data yang berhubungan dengan hasil belajar siswa. Proses ini memungkinkan sistem untuk memprediksi tingkat kelulusan siswa berdasarkan data yang telah dianalisis sebelumnya. Penerapan algoritma *machine learning* dapat dilakukan dengan mengidentifikasi pola dalam data akademik (nilai ujian akhir semester, rata-rata rapor, nilai tugas harian) dan data non-akademik (kehadiran, kegiatan ekstrakurikuler, serta motivasi dalam pembelajaran), yang membantu dalam menentukan faktor-faktor yang memengaruhi kelulusan siswa. Melalui pendekatan ini, sistem dapat memberikan wawasan yang lebih mendalam mengenai kesuksesan atau kegagalan siswa[9].

2.1.2 Penerapan Machine Learning untuk Prediksi Kelulusan Siswa

Salah satu aplikasi utama *Machine Learning* dalam pendidikan adalah prediksi kelulusan siswa, di mana algoritma digunakan untuk menganalisis data akademik dan non-akademik siswa. Dengan memanfaatkan teknik klasifikasi dan regresi, *machine learning* mampu mengidentifikasi faktor-faktor yang

mempengaruhi kelulusan siswa secara lebih akurat, sebagai contoh, model *machine learning* dapat memprediksi kemungkinan kelulusan siswa berdasarkan pola yang ditemukan dalam data historis, dengan demikian metode berbasis *machine learning* dapat meningkatkan akurasi prediksi, yang pada gilirannya membantu pengambil keputusan di sekolah dalam merencanakan kebijakan yang lebih efektif dan tepat sasaran[10].

Sejalan dengan hal tersebut, berbagai penelitian menunjukkan bahwa penggunaan *machine learning* dalam analisis data pendidikan tidak hanya terbatas pada prediksi kelulusan, tetapi juga pada pengembangan sistem pembelajaran berbasis *e-learning* yang lebih adaptif dan interaktif[11]. *Machine Learning* memungkinkan pengembangan sistem pembelajaran yang menyesuaikan cara belajar siswa dengan kebutuhan dan gaya belajarnya masing-masing, ini menciptakan lingkungan belajar yang lebih fleksibel, di mana siswa dapat belajar dengan cara yang paling sesuai bagi mereka, meningkatkan efisiensi dan hasil pembelajaran secara keseluruhan[12].

2.1.3 Keunggulan Penerapan Machine Learning dalam Pendidikan

Tabel 2.1 Keunggulan Penerapan Machine Learning dalam Pendidikan

Keunggulan	Penjelasan
Pembelajaran Yang di Sesuaikan	Sistem M*achine Learning* dapat menyesuaikan materi pembelajaran dengan kebutuhan siswa.
Analisis Data	Meningkatkan efisiensi analisis data besar untuk memahami perilaku siswa.

Penilaian	Sistem *Machine Learning* juga digunakan untuk mengurangi jumlah waktu yang dibutuhkan untuk menilai pekerjaan siswa.
Inovasi Dalam Pengajaran	Memfasilitasi pengembangan metode pengjaran yang lebih inovatif dan adaptif.

Penerapan *machine learning* dalam pendidikan menawarkan berbagai keunggulan, antara lain kemampuan untuk mengidentifikasi pola-pola yang tidak terlihat dengan pendekatan tradisional. Dengan demikian, model berbasis *machine learning* dapat memberikan wawasan yang lebih mendalam tentang faktor-faktor yang mempengaruhi hasil belajar siswa, baik itu faktor akademik maupun non-akademik, hal ini memungkinkan pendidik untuk melakukan intervensi yang lebih tepat dan sasaran yang lebih fokus. Selain itu, penerapan *machine learning* juga meningkatkan efisiensi dalam pengolahan dan analisis data besar, yang sebelumnya sulit dilakukan dengan menggunakan metode manual. Keunggulan lainnya adalah kemampuan *machine learning* untuk mengurangi dimensi data yang kompleks, menyederhanakan proses analisis, dan meningkatkan pemahaman terhadap variabel-variabel yang mempengaruhi hasil belajar siswa[13].

2.2 Data Mining

Data mining menurut Suyanto adalah gabungan sejumlah disiplin ilmu komputer yang mendefinisikan sebagai proses penemuan pola-pola baru dari kumpulan-kumpulan data sangat besar, meliputi metode-metode yang merupakan

irisan dari *artificial intelligence*, *machine learning*, *statistics,* dan *database systems.*

Data mining merupakan proses eksplorasi dan analisis data yang bertujuan untuk menemukan pola-pola yang bermanfaat, hubungan tersembunyi, dan informasi yang dapat diambil dari sekumpulan data besarc Adapun proses yang umumnya dilakukan oleh *data Mining* meliputi deskripsi, prediksi, estimasi, klasifikasi, clushtering dan asosiasi. Dalam data mining prediksi dan klasifikasi banyak digunakan untuk menganalisis suatu data yang dapat menggambarkan kelas data atau untuk memprediksi data di masa depan[15]. *Data Mining* merupakan metode untuk menemukan informasi tersembunyi dalam *database* dan bagian dari proses *Knowledge Discovery in Databases (KDD)* untuk menemukan informasi dan pola yang berguna dalam data. Secara umum, Data Mining dibagi menjadi dua kategori utama yaitu[16]:

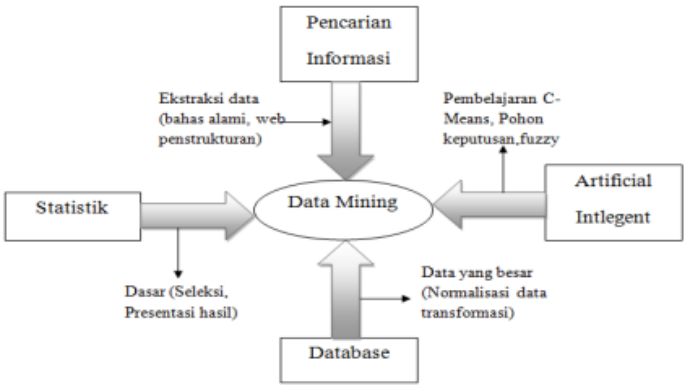
1) Prediktif

   Proses untuk menemukan karakteristik penting dari data dalam satu basis data. Teknik data mining yang termasuk *descriptive mining* adalah c*lustering, asosiation*, dan *sequential mining*.

2) Deskriptif

   Proses untuk menemukan pola dari data dengan menggunakan beberapa variable lain di masa depan. Salah satu teknik yang terdapat dalam *predictive mining* adalah klasifikasi. Secara sederhana data mining biasa dikatakan sebagai proses penyaring atau “menambang” pengtahuan dari sejumlah data yang besar.

Tujuan utama dari *data mining* adalah untuk mengekstraksi pola dari data yang ada, menambah nilai intrinsik dari data serta mengubahnya menjadi pengetahuan. Nama lain *data mining* adalah *Knowledge discovery from Databases (KDD)*[17].



Gambar 2.2 Bidang Ilmu Data Mining[17]

Karakteristik *data mining* adalah sebagai berikut[18]:

a) *Data mining* berhubungan dengan penemuan sesuatu yang tersembunyi dan pola data tertentu yang tidak diketahui sebelumnya.
b) *Data mining* biasa menggunakan data yang sangat besar.
c) Biasanya data yang besar digunakan untuk membuat hasil lebih dipercaya.
d) *Data mining* berguna untuk membuat keputusan yang kritis, terutama dalam strategi.

2.2.1 Data Mining Dalam Pendidikan

Data mining merupakan suatu proses pencarian pola dari data-data dengan jumlah yang sangat banyak yang tersimpan dalam suatu tempat penyimpanan dengan menggunakan teknologi pengenal pola, teknik statistik, dan matematika[19]. Dalam konteks pendidikan, *Data Mining* diterapkan untuk

menganalisis data akademik dan non-akademik siswa, seperti nilai ujian, kehadiran, partisipasi dalam kegiatan ekstrakurikuler, serta faktor-faktor lain yang dapat mempengaruhi hasil belajar siswa[20]. Tujuan utama *Data Mining* adalah mengeksplorasi data dengan tujuan menemukan wawasan yang bermanfaat dan mendukung pengambilan keputusan yang lebih baik. Melalui teknik seperti klastering, klasifikasi, dan regresi, *Data Mining* dapat membantu mengidentifikasi pola tersembunyi, memprediksi perilaku masa depan, dan mengoptimalkan strategi bisnis[21].

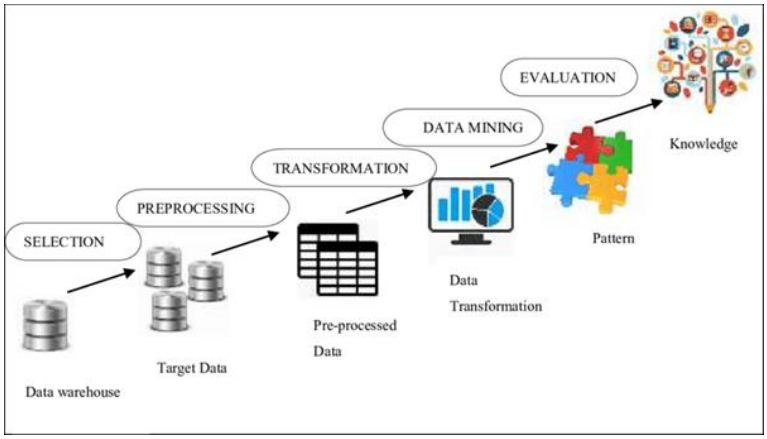
Penerapan *data mining* dalam pendidikan merupakan suatu metode yang diterapkan dalam pemrosesan data berukuran besar, proses pengolahan data ini melibatkan teknik-teknik tertentu untuk menghasilkan pemahaman baru dari informasi yang terdapat dalam data tersebut, yang pada gilirannya dapat digunakan sebagai dasar untuk pengambilan keputusan, dalam ranah *data mining*, terdapat beragam metode yang dapat digunakan, seperti Klasifikasi, Klastering, Estimasi, Prediksi, dan Asosiasi[22]. Teknik klasifikasi digunakan untuk memprediksi kelulusan siswa berdasarkan data akademik dan non-akademik mereka. Beberapa algoritma yang sering digunakan dalam klasifikasi adalah *decision trees*, *support vector machines (SVM)*, dan *naive bayes*. Selain klasifikasi, clustering merupakan teknik lain yang sangat berguna untuk memahami karakteristik siswa berdasarkan atribut tertentu, seperti nilai, kehadiran, atau partisipasi dalam kegiatan ekstrakurikuler. Sementara itu, analisis asosiasi digunakan untuk menemukan hubungan atau korelasi antara berbagai faktor yang mempengaruhi hasil akademik siswa. Teknik ini memungkinkan

identifikasi hubungan antara faktor-faktor akademik dan non-akademik yang dapat mempengaruhi kelulusan siswa[23].

2.2.2 Penerapan Data Mining untuk Memprediksi Kelulusan Siswa

Penerapan *data mining* dalam memprediksi kelulusan siswa menggunakan berbagai teknik, seperti klasifikasi, regresi, dan *clustering*, telah terbukti sangat efektif. Teknik-teknik ini membantu mengidentifikasi pola-pola dalam data yang berkaitan dengan hasil belajar siswa, seperti siswa yang sering absen atau yang memiliki nilai rendah dalam ujian tertentu. Dengan memanfaatkan teknik ini, sistem dapat memberikan prediksi yang lebih tepat mengenai kemungkinan kelulusan siswa[24].

Penerapan *data mining* dalam memprediksi kelulusan siswa menggunakan suatu langkah dalam *Knowledge Discovery in Databases* (KDD). *Knowledge discovery* sebagai suatu proses terdiri atas pembersihan data (data *cleaning*), integrasi data (data *integration*), pemilihan data (data *selection*), transformasi data (data *transformation*), data mining, evaluasi pola (*pattern evaluation*) dan penyajian pengetahuan (*knowledge presentation*). Data mining mengacu pada proses untuk menambang (*mining*) pengetahuan dari sekumpulan data yang sangat besar[25]. Berikut ini aliran informasi dalam *data mining*:



Gambar 2.3 Aliran Informasi dalam Data Mining[26]

1) Data *Selection* (Seleksi Data) : Sebelum di integrasikan ke dalam satu database dan gudang, prosedur ini digunakan untuk menghapusdata yang berisik dan tidak konsisten dari data yang terdapat di banyak database. Basis data ini mungkinmemiliki format atau platform yang berbeda.
2) Data *Transformation* (Transformasi Data) : Data database kemudian diubah menggunakan berbagai metode. Khususnya ketika berhadapan dengan masalah skala besar, prosedur pemilihan sangat penting untuk mendapatkan hasil yang lebih tepat dan mempercepat perhitungan. Agar data yang telah diproses siap untuk ditambang, diperlukan transformasi data sebagai langkah preprocessing.
3) Data *Mining* (Penambangan Data) : Proses penambangan data atau data mining dapat dilakukan setelah data yang telah diseleksi dan ditransformasikan ditambang dengan menggunakan berbagai macam metode. Selama proses penambangan data, fungsi-fungsi khusus harus digunakan untuk mencari pola atau informasi menarik dalam data yang

dipilih. Tujuan dan keseluruhan proses pencarian pengetahuan sangat mempengaruhi pemilihan fungsi atau algoritma yang sesuai.

4) *Evaluation* (Evaluasi) : Tahap ini melibatkan penggambaran visualisasi, penyajian pengetahuan dalam bentuk yang mudah dipahami oleh pengguna, dan menentukan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.
5) *Knowledge* (Pengetahuan) : Memvisualisasikan data yang diproses untuk memudahkan pengguna memahami dan mengambil tindakanberdasarkan analisis adalah langkah terakhir dalam proses KDD. Akibatnya, salah satu langkah penting dalam proses penambangan data adalah menyajikan hasil dengan cara yang dapat dipahami semua orang.

2.2.3 Keunggulan Penerapan Data Mining dalam Pendidikan

Tabel 2.2 Keunggulan Penerapan Data Mining dalam Pendidikan

Keunggulan	Penjelasan
Personalisasi Pembelajaran	Pembelajaran dapat disesuaikan dengan karakteristik dan kebutuhan masing-masing siswa berdasarkan data yang dikumpulkan.
Pengambilan Keputusan Berbasis Data	Data mining memungkinkan pengambilan keputusan yang lebih objektif dan efisien berdasarkan analisis data besar .
Intervensi Tepat Sasaran	Teknik seperti klasifikasi dan *clustering* memungkinkan pendidik mengidentifikasi

	siswa yang berisiko gagal dan memberikan intervensi yang efektif.
Perbaikan Kurikulum	Analisis data memungkinkan pengembangan kurikulum yang lebih responsif terhadap kebutuhan siswa dan perkembangan pendidikan.

Penerapan *data mining* dalam pendidikan menawarkan berbagai keunggulan, antara lain[27]:

1. Pengambilan Keputusan Berbasis Data : *Data mining* memungkinkan pengambilan keputusan yang lebih objektif dan berbasis bukti, dengan menganalisis data besar untuk menemukan pola-pola yang tidak terlihat dengan metode tradisional.
2. Personalisasi Pembelajaran : Teknik *clustering* memungkinkan pendidik untuk mengelompokkan siswa berdasarkan kesamaan karakteristik, sehingga pembelajaran dapat disesuaikan dengan kebutuhan masing-masing kelompok siswa.
3. Intervensi yang Lebih Tepat Sasaran : Dengan menggunakan *data mining*, pendidik dapat mengidentifikasi siswa yang berisiko gagal, sehingga dapat memberikan intervensi yang lebih tepat sasaran dan mendukung siswa dalam mencapai kelulusan.
4. Perbaikan Kurikulum dan Sumber Daya : *Data mining* membantu dalam merancang kurikulum yang lebih responsive dan efisien, serta mengalokasikan sumber daya secara optimal untuk mendukung pembelajaran.

2.3 Algoritma Naive Bayes

Naïve Bayes adalah perhitungan probabilitas dengan metode pengklasifikasian, model ini mudah untuk dibangun dan tidak *complicated*, sehingga dianggap tepat untuk database yang berukuran kecil sampai berukuran besar. Algoritma *Naïve Bayes* menghitung probabilitas kejadian masa dating dari kejadian sebelumnya di mana masing- masing variable dianggap tidak saling tergantung[28]. *Naive bayes* adalah sebuah metode yang paling sering digunakan untuk mengklasifikasi statistik sehinngga banyak digunakan oleh para pengembang sistem berbasis online maupun offline untuk mengetahui probabilitas suatu kelas. Algoritma *Naive Bayes* adalah metode yang dilakukan dalam bentuk klasifikasi data training dan data testing. Data tersebut dihitung dengan cara menghitung peluang dari suatu kelas masing – masing atribut yang ada, dengan menentukan kelas mana yang paling optimal sehingga menghasilkan suatu hipotesa[29].

Kelebihan dari algoritma *Naïve Bayes* adalah sederhana tapi mempunyai akurasi yang tinggi meskipun menggunakan data yang sedikit, Sedangkan kelemahan dalam *Naive Bayes*, yaitu prediksi hasil probabilitas berjalan tidak optimal serta kurangnya pemilihan fitur yang relevan terhadap klasifikasi sehingga akurasi menjadi rendah Hal tersebut dapat diatasi dengan cara pemilihan fitur yang berguna untuk meningkatkan akurasi.

Naive Bayes adalah algoritma pembelajaran mesin berbasis probabilitas yang menggunakan *Teorema Bayes* untuk menghitung probabilitas posterior dari suatu kelas berdasarkan fitur-fitur yang ada. Dalam formula *Teorema Bayes*,

probabilitas suatu kelas P(H|X) dihitung berdasarkan probabilitas awal kelas P(H), probabilitas fitur diberikan kelas P(X|H), dan probabilitas fitur P(X) [30].

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

Rumus teorema Naïve bayes :

X = Data dengan class yang belum diketahui

H = Hipotesis data X merupakan suatu class spesifik

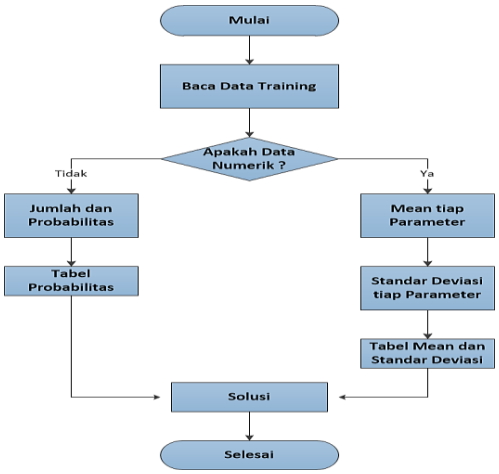
P(H|X) = Probabilitas hipotesis H berdasarkan kondisi x (posteriori prob.)

P(H) = Probabilitas hipotesis H (prior prob.)

P(X|H) = Probabilitas X berdasarkan kondisi tersebut

P(X) = Probabilitas dari X

Asumsi utama dari algoritma ini adalah independensi antar fitur, yang berarti setiap fitur dianggap tidak saling memengaruhi satu sama lain. Asumsi ini menyederhanakan perhitungan probabilitas sehingga proses klasifikasi menjadi cepat dan efisien.



Gambar 2.4 Alur Metode Naive Bayes[31]

Adapun keterangan dari gambar di atas sebagai berikut :

1. Baca data *training*
2. Hitung Jumlah dan probabilitas, namun apabila data numerik maka:
   a) Cari nilai mean dan standar deviasi dari masing-masing parameter yang merupakan .
   b) Cari nilai probabilistik dengan cara menghitung jumlah data yang sesuai dari kategori tersebut.
3. Mendapatkan nilai dalam tabel *mean*, standard deviasi dan probabilitas.
4. Solusi kemudian dihasilkan

2.4 Alat Bantu Pemrograman/Tools Pendukung

2.4.1 Aplikasi RapidMiner

RapidMiner digunakan sebagai Platform utama untuk implementasi algoritma dan analisis data visual. *RapidMiner* merupakan salah satu alat bantu yang populer dalam proses Data Mining dan analisis data. Alat ini memungkinkan pengguna untuk melakukan berbagai tahapan analisis data, mulai dari pembersihan data, transformasi, hingga penerapan algoritma seperti *Naive Bayes* untuk prediksi dan klasifikasi. Selain itu, *RapidMiner* juga menyediakan antarmuka yang user-friendly, memudahkan peneliti dalam melakukan eksplorasi data dan visualisasi hasil analisis.

RapidMiner memiliki keunggulan dalam hal fleksibilitas dan kemampuan untuk mengolah data dalam skala besar, serta mendukung berbagai teknik Analisis Data, seperti klasifikasi, regresi, dan klasterisasi. Platform ini sangat sesuai untuk

penelitian yang melibatkan analisis data sosial-ekonomi, karena dapat mengintegrasikan berbagai atribut seperti pendapatan keluarga, pendidikan orang tua, dan kondisi sosial siswa untuk menghasilkan model prediktif yang lebih akurat [32].

RapidMiner juga mendukung visualisasi data yang memungkinkan peneliti untuk menggambarkan hasil analisis secara grafis, mempermudah pemahaman terhadap pola-pola yang ada dalam dataset. Visualisasi ini sangat berguna untuk menggambarkan hubungan antara atribut sosial-ekonomi siswa dan hasil belajar mereka, yang dapat membantu dalam merancang kebijakan pendidikan yang lebih responsif [33].

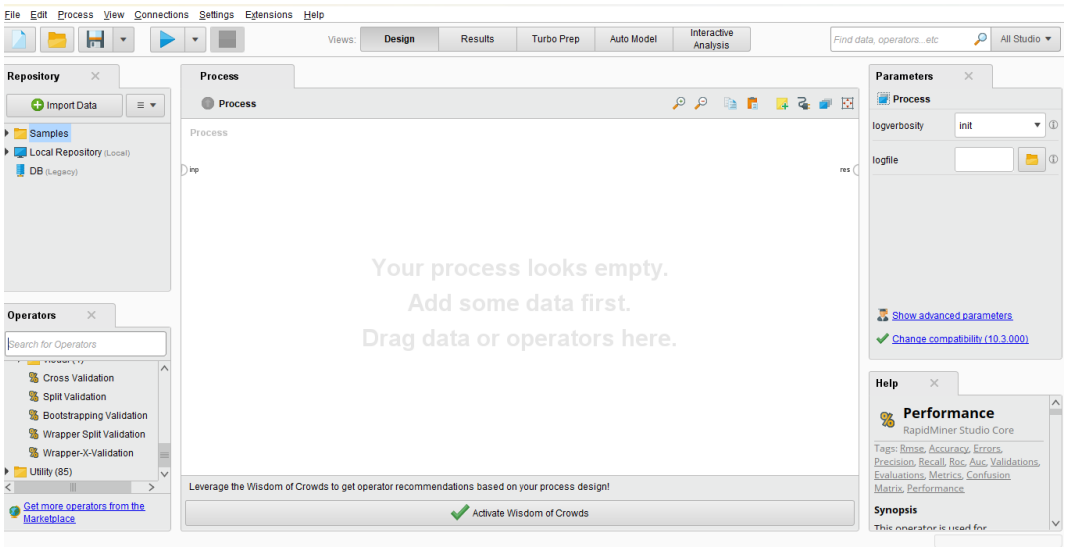
Secara keseluruhan, penggunaan *RapidMiner* dalam penelitian ini membantu dalam pengolahan data, penerapan algoritma, dan visualisasi hasil analisis, yang pada gilirannya meningkatkan akurasi dalam memprediksi hasil belajar siswa berdasarkan atribut sosial-ekonomi mereka. Alat ini berperan penting dalam memastikan bahwa hasil penelitian dapat diterima dengan baik dan memberikan kontribusi yang signifikan terhadap pengembangan kebijakan pendidikan yang berbasis data. [34]. *RapidMiner* menawarkan berbagai keunggulan dalam analisis data yang kompleks, berikut ini keunggulan *RapidMiner* yaitu :

a) Antarmuka Visual adalah memungkinkan pengguna untuk mengoperasikan dan memanipulasi data dengan cara yang intuitif.
b) Fleksibilitas dalam memproses data ini mendukung berbagai jenis algoritma dan teknik analisis, termasuk *Naive Bayes* untuk Klasifikasi, yang digunakan

dalam penelitian ini untuk memprediksi kelayakan penerima bantuan pendidikan berdasarkan atribut sosial-ekonomi.

c) Kemampuan untuk mengelola data besar dengan kemampuan untuk memproses dan menganalisis dataset besar, *RapidMiner* sangat berguna dalam penelitian yang melibatkan banyak variabel, seperti data sosial-ekonomi siswa.

d) Visualisasi data yang kuat menyediakan berbagai jenis visualisasi yang membantu dalam mengidentifikasi pola dan hubungan dalam data, yang memperjelas hasil analisis dan membuatnya lebih mudah dipahami.



Gambar 2.5 Tampilan Awal RapidMiner

2.5 Metodologi Penelitian

2.5.1 Penelitian Terdahulu

Dibawah ini adalah refrensi yang penulis gunakan untuk mendukung dan sebagai landasan pembuatan proposal yaitu :

Tabel 2.3 Penelitian Terdahulu

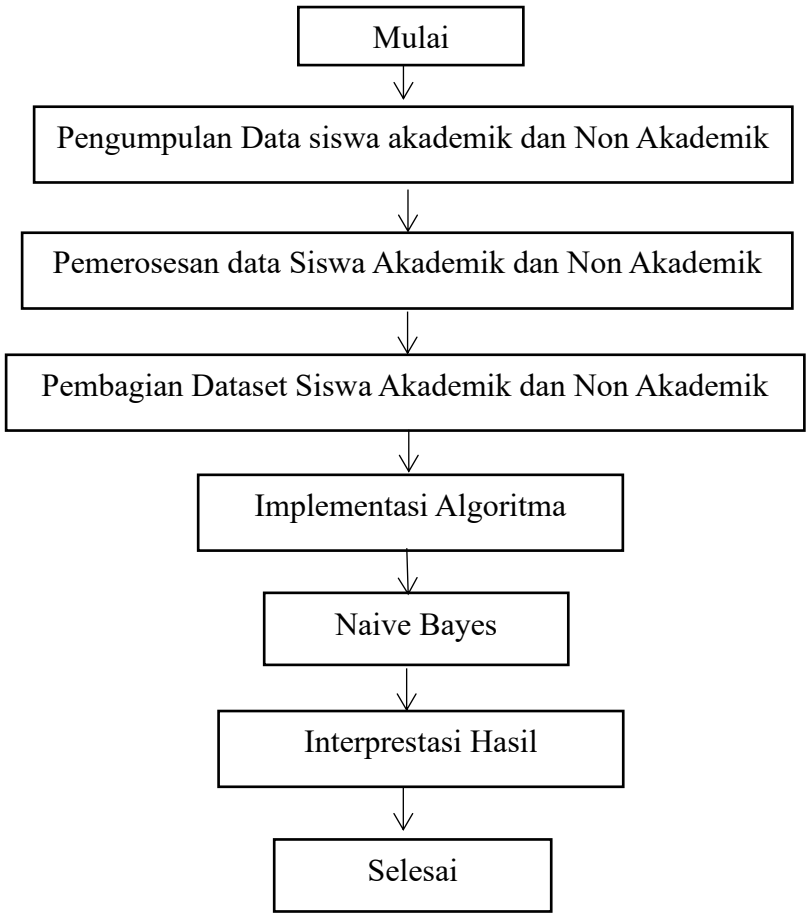
Refrensi Penelitian	1
Judul	Prediksi Kelulusan Tepat Waktu Berdasarkan Riwayat Akademik Menggunakan Metode Naïve Bayes
Nama Penulis	Imam Riadi1, Rusydi Umar, Rio Anggara
Tahun	2024
Hasil	Pada refrensi jurnal ini dijelaskan mengenai Kelulusan tepat waktu. Penelitian data kelulusan bisadilakukan menggunakan tehnik klasifikasi *data mining*. Klasifikasi merupakan salah satu pengolahan dalam data mining dilakukan dengan cara mengelompokkan dengan metode tertentu. Penelitian ini membangun aplikasi dengan implementasi metode *naive bayes* dengan mempertimbangkan parameter menghasilkan klasifikasi mahasiswa lulus tidak tepat waktu dan lulus tepat waktu. Tahapan pada penelitian seperti *load data*, *cleaning data*, *selection data*, *transformation data*, *data training*, *data testing*, dan hasil prediksi.
Refrensi Penelitian	2
Judul	Analisis Manajemen Kesiswaan dalam Meningkatkan Prestasi Akademik dan Non Akademik di SMP Negeri 1 Banyuglugur Situbondo
Nama Penulis	Nurul Alifah

Tahun	2023
Hasil	Pada refrensi jurnal ini dijelaskan mengenai pentingnya untuk memiliki manajemen kesiswaan yang efektif dalam mengatur semua kegiatan yang terkait dengan siswa agar mencapai tujuan pendidikan secara efisien. Berdasarkan hal tersebut, peneliti bermaksud untuk melakukan penelitian dengan tujuan sebagai berikut: (1) Memahami perencanaan pembinaan untuk meningkatkan prestasi akademik dan non-akademik di SMP Negeri 1 Banyuglugur. (2) Memahami pelaksanaan pembinaan kesiswaan dalam upaya meningkatkan prestasi akademik dan non-akademik di SMP Negeri 1 Banyuglugur. (3) Memahami evaluasi pelaksanaan pembinaan kesiswaan dalam rangka meningkatkan prestasi akademik dan non-akademik di SMP Negeri 1 Banyuglugur Situbondo. Penelitian ini menggunakan pendekatan kualitatif, yang memungkinkan peneliti untuk mengamati situasi di lapangan sesuai dengan pertanyaan penelitian. Data dikumpulkan melaluiobservasi, wawancara, dan dokumentasi. Teknik analisis data yang digunakan mencakup reduksi data, penyajian data, dan penarikan

	kesimpulan.
Refrensi Penelitian	3
Judul	Data Mining Model Klasifikasi Untuk Ketepatan Waktu Kelulusan Mahasiswa
Nama Penulis	Chetsi Rafelin Patrisia, Sylvia, Vivian
Tahun	2024
Hasil	Pada refrensi jurnal ini dijelaskan bahwa menggunakan metode klasifikasi algoritma *naïve bayes* untuk mengklasifikasi waktu kelulusan mahasiswa. Metode penelitian yang digunakan adalah jenis penelitian kuantitatif yang mengikuti prosedur data mining berdasarkan prosedur CRISP-DM (*Cross Industry Standard Process Model for Data Mining*). Pengumpulan data menggunakan kuesioner. Teknik pengolahan data yang diterapkan untuk mengklasifikasikan waktu kelulusan mahasiswa menggunakan metode klasifikasi data mining dengan algoritma *Naïve Bayes*, yang dibantu oleh software *RapidMine*r Studio Versi 10.1. Penelitian ini menunjukkan nilaiakurasi sebesar 75%, yang menunjukkan jikadata mining dengan model klasifikasi menggunakan algoritma *Naïve Bayes* efektif untuk mengklasifikasikan ketepatan waktu

	kelulusan mahasiswa.

2.5.2 Kerangka Kerja Penelitian



Gambar 2.6 Kerangka Kerja Penelitian

Kerangka kerja penelitian di atas berbasis *machine learning* dalam pendidikan dirancang secara sistematis untuk memastikan keandalan model prediksi yang digunakan. Langkah-langkah utama dalam penelitian ini meliputi:

a) Pengumpulan Data siswa akademik dan non akademik : Data dikumpulkan dari berbagai sumber, mencakup data akademik seperti rata-rata nilai ujian, nilai tugas harian dan absensi, serta data non-akademik seperti

ekstrakurikuler, motivasi dalam belajar dan latar belakang ekonomi. Pengumpulan data yang menyeluruh sangat penting untuk memastikan bahwa semua aspek yang memengaruhi kelulusan siswa dapat dianalisis secara holistik.

b) Pemrosesan Data siswa akademik dan non akademik: Pemrosesan adalah dimana data siswa akademik dan non akademik di proses yang nantinya pembagian dataset, dipilih berdasarkan signifikansinya terhadap prediksi hasil kelulusan siswa .Teknik seperti analisis korelasi dapat digunakan untuk mengidentifikasi yang memiliki pengaruh besar terhadap hasil prediksi.

c) Pembagian Data Set siswa akademik dan non akademik : Dibagian ini nantinya data akademik dan non akademik akan dipisahkan sesuai data set.

d) Pemilihan Algoritma *Machine Learning* : Algoritma yang dipilih harus sesuai dengan karakteristik data yang digunakan. Dalam konteks penelitian ini, algoritma seperti *Naive Bayes* digunakan untuk memprediksi kelulusan siswa. Pemilihan algoritma didasarkan pada kekuatan masing-masing metode dalam menangani data kategori dan kuantitatif.

e) Pelatihan Model : Dataset yang telah diproses digunakan untuk melatih model prediksi. Pada tahap ini, data dibagi menjadi dua bagian: data pelatihan dan data uji, dengan tujuan untuk mengukur kinerja model pada data baru. Validasi silang (*cross-validation*) sering diterapkan untuk memastikan model tidak overfitting terhadap data pelatihan. Evaluasi Hasil : Model yang telah dilatih dievaluasi menggunakan metrik seperti akurasi, precision, recall, dan

F1-Score. keseluruhan dan memastikan bahwa model dapat digunakan untuk memprediksi kelulusan siswa secara andal.