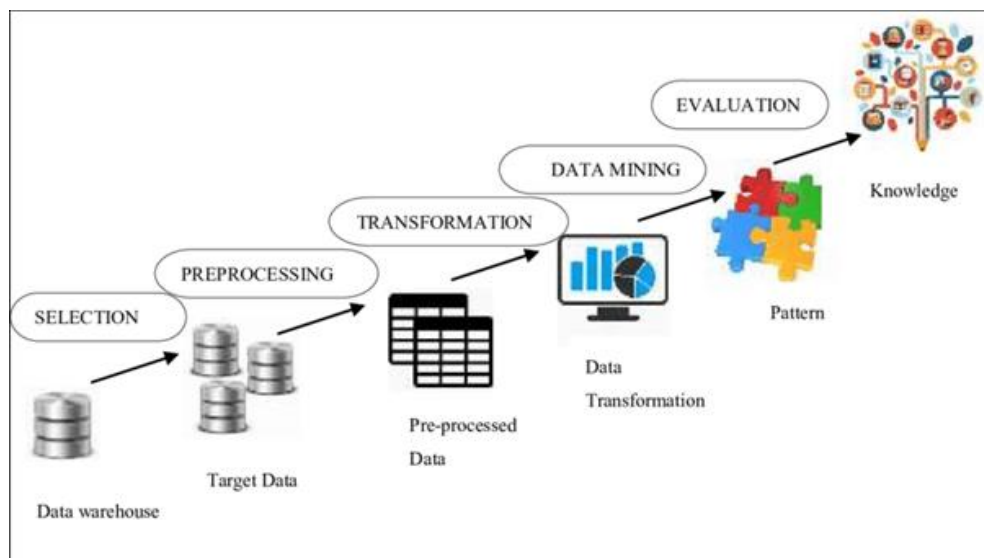


BAB II LANDASAN TEORI

2.1. Knowledge Discovery in Database (KDD)

Knowledge Discovery in Database (KDD) merupakan metode teknis untuk menemukan pola bermakna dalam data yang dapat dimanfaatkan dalam pengambilan keputusan. Proses KDD terdiri atas lima tahapan utama, yaitu seleksi data, *pre-processing*, transformasi, *data mining*, dan evaluasi. Tahap seleksi bertujuan menentukan data yang relevan, sedangkan *pre-processing* mencakup integrasi dan pembersihan data untuk menghilangkan *noise*, redundansi, dan inkonsistensi. Tahap transformasi menyesuaikan format data agar siap diproses pada tahap *data mining* menggunakan algoritma yang sesuai. Tahap evaluasi digunakan untuk menilai kualitas pola atau model sebelum dijadikan pengetahuan (Alfandi & Fatah, 2024).



Gambar 2.1. Tahapan Knowledge Discovery in Database (KDD)

Sumber: (Alfandi & Fatah, 2024)

Gambar 2.1 menampilkan alur visual tahapan Knowledge Discovery in Database (KDD) yang menggambarkan hubungan antar proses dalam pengolahan data. Diagram memperlihatkan perpindahan data dari *data warehouse* menuju data target hingga terbentuknya pengetahuan. Setiap tahapan ditunjukkan secara berurutan mulai dari seleksi data, *preprocessing*, transformasi, *data mining*, hingga evaluasi. Ilustrasi ini menegaskan bahwa proses KDD bersifat berkesinambungan, di mana keluaran dari setiap tahap menjadi masukan bagi tahap berikutnya hingga menghasilkan pola dan pengetahuan yang dapat dimanfaatkan.

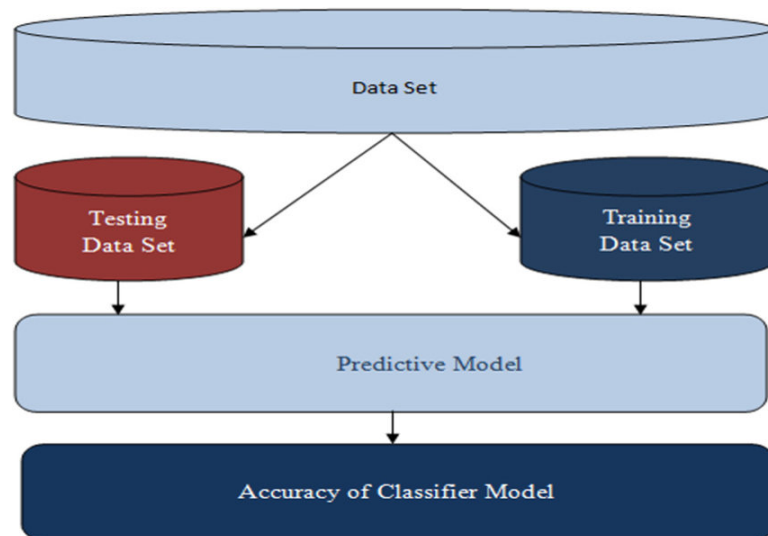
2.2. Data Mining

Data mining merupakan proses untuk menggali dan mengekstraksi informasi atau pola penting yang sebelumnya tidak diketahui dari sekumpulan data dalam jumlah besar. Proses ini bertujuan menghasilkan informasi yang bernilai guna mendukung pengambilan keputusan secara lebih efektif. Dalam kerangka *Knowledge Discovery in Database* (KDD), data mining merupakan bagian dari rangkaian proses yang mencakup pembersihan data, integrasi data, transformasi data, penerapan teknik data mining, evaluasi pola, serta penyajian pengetahuan yang dihasilkan kepada pengguna (Harist Nst et al., 2021).

Penerapan data mining berkembang di berbagai bidang, termasuk pendidikan, seiring meningkatnya ketersediaan data digital (*digital footprint*) dari aktivitas pembelajaran. Data akademik seperti nilai tugas, nilai ujian, kehadiran, dan interaksi dalam sistem pembelajaran daring dapat dianalisis untuk mengidentifikasi pola belajar, memahami perilaku peserta didik, serta memprediksi kinerja akademik (Kharis & Zili, 2022).

2.2.1. Peran Data Mining dalam Klasifikasi Prediktif

Klasifikasi merupakan teknik utama dalam *data mining* yang termasuk dalam pendekatan *supervised learning* dan digunakan untuk menentukan kategori data berdasarkan pola pada data pelatihan. Teknik ini dimanfaatkan untuk memprediksi perilaku, mengidentifikasi risiko, serta mengelompokkan individu berdasarkan karakteristik tertentu. Bidang pendidikan memanfaatkan metode klasifikasi untuk memetakan potensi akademik dan memprediksi kelulusan mahasiswa secara lebih terukur. Penerapan algoritma *K-Nearest Neighbors* (KNN) menunjukkan bahwa klasifikasi mampu mengidentifikasi pola akademik berdasarkan tingkat kemiripan karakteristik antar data (Atmaja et al., 2023). Penerapan algoritma *Naive Bayes* juga membuktikan bahwa metode klasifikasi berbasis probabilitas dapat menentukan kategori data secara efisien melalui distribusi fitur, sehingga mendukung pengambilan keputusan berbasis data (Julianto & Andayani, 2024).



Gambar 2.2. Alur Klasifikasi Prediktif Berbasis Data Mining

Sumber: (Sharma et al., 2017)

Gambar 2.2 memperlihatkan alur kerja klasifikasi prediktif dalam proses data mining, yang dimulai dari satu data set utama yang dibagi menjadi training data set dan testing data set. Training data set digunakan untuk membentuk model prediktif melalui pembelajaran pola dan karakteristik data. Model yang terbentuk kemudian diuji menggunakan testing data set untuk menilai kemampuan klasifikasi terhadap data baru. Proses ini menghasilkan nilai akurasi yang menunjukkan tingkat ketepatan model dalam menentukan kategori data, sehingga menegaskan pentingnya evaluasi kinerja dalam klasifikasi prediktif.

2.3. Analisis

Analisis merupakan proses pengolahan dan penafsiran data secara sistematis untuk memperoleh pemahaman yang jelas terhadap fenomena yang diteliti. Proses ini dilakukan dengan mengorganisasi data, menyusun pola, serta menarik kesimpulan yang dapat dipertanggung jawabkan secara ilmiah. Dalam penelitian kuantitatif, analisis berperan penting dalam menentukan kualitas temuan dan menjadi dasar dalam menjawab rumusan masalah, sehingga diperlukan ketelitian agar hasilnya mencerminkan kondisi objek penelitian.

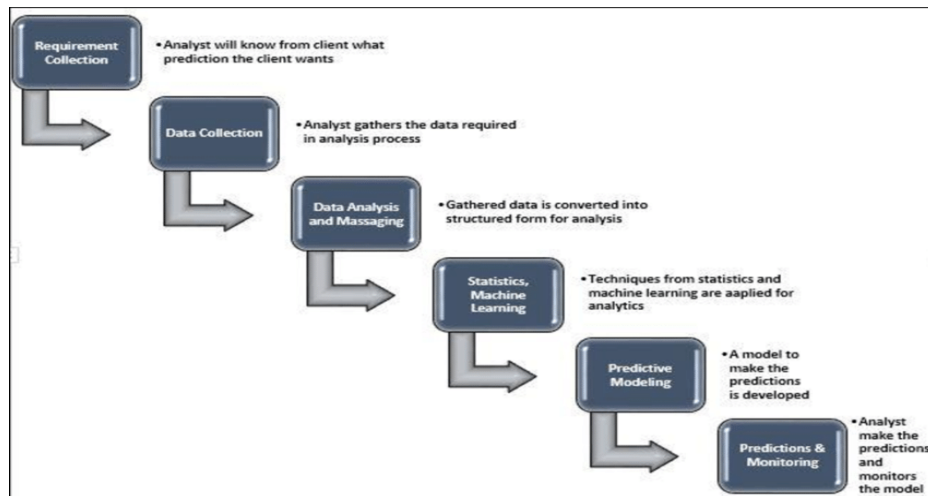
Penelitian mengenai teknik analisis data kuantitatif menjelaskan bahwa proses analisis dilaksanakan melalui beberapa tahapan, meliputi penyuntingan data, pengkodean, tabulasi, dan pengujian statistik untuk menghasilkan temuan yang terstruktur. Setiap tahapan memiliki peran penting dalam menjaga keakuratan interpretasi data, sehingga kesimpulan yang dihasilkan tetap relevan dengan permasalahan penelitian dan dapat dimanfaatkan sebagai dasar penentuan strategi maupun kebijakan terkait objek penelitian (Sofwatillah et al., 2024).

2.4. Analisis Prediktif

Analisis prediktif merupakan pendekatan yang memanfaatkan data historis, teknik statistik, dan algoritma *machine learning* untuk memproyeksikan kondisi pada masa mendatang. Pendekatan ini digunakan untuk mengenali pola berulang, memetakan kecenderungan, serta menghasilkan estimasi berdasarkan struktur data yang dipelajari oleh model. Bidang pendidikan memanfaatkan analisis prediktif untuk memahami faktor yang memengaruhi performa siswa dan mengantisipasi risiko akademik secara lebih dini. Penerapan analisis prediktif dengan algoritma seperti *Random Forest* dan *K-Nearest Neighbors* menunjukkan kemampuan dalam memprediksi kinerja siswa secara akurat, sehingga mendukung perancangan intervensi pendidikan yang lebih tepat dan berbasis data. (Amin et al., 2025).

2.4.1. Konsep Dasar Analisis Prediktif

Analisis prediktif membangun model berdasarkan pola pada data historis untuk memperkirakan nilai atau kategori pada periode berikutnya. Proses ini bertujuan mengidentifikasi hubungan antar variabel serta kecenderungan tersembunyi dalam *data set* sebagai dasar peramalan dan pengambilan keputusan. Penerapan analisis prediktif memanfaatkan metode seperti regresi, klasifikasi (*classification*), dan *clustering* yang dikombinasikan dengan teknik statistik serta algoritma *machine learning*. Model yang dihasilkan memungkinkan peneliti dan lembaga pendidikan mengevaluasi kondisi aktual, memprediksi perkembangan siswa, serta meningkatkan efektivitas pemantauan kinerja akademik dan perencanaan pembelajaran berbasis data.



Gambar 2.3. Alur Proses Analisis Prediktif

Sumber: (Kumar & Garg, 2018)

Gambar 2.3 menggambarkan alur konseptual analisis prediktif yang disusun secara bertahap untuk menghasilkan prediksi berbasis data historis. Proses dimulai dari *requirement collection* dan *data collection*, kemudian dilanjutkan dengan *data analysis and messaging* untuk menyiapkan data agar layak dianalisis. Tahap *predictive modeling* dilakukan untuk membangun model prediksi, yang selanjutnya digunakan pada tahap *predictions and monitoring* guna mendukung pengambilan keputusan serta memantau kinerja model secara berkelanjutan.

2.4.2. Peran Analisis Prediktif dalam Evaluasi Program Publik

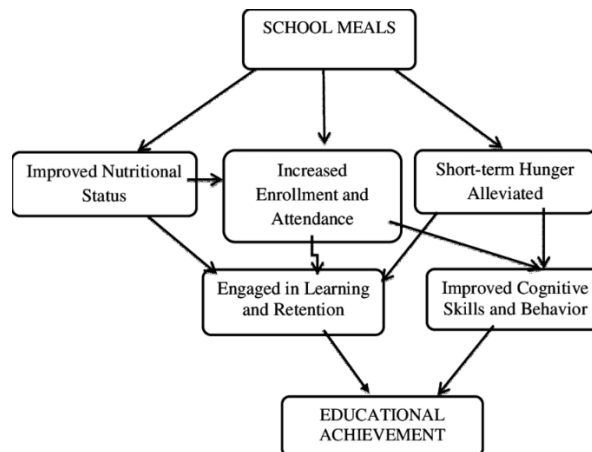
Evaluasi program publik membutuhkan pendekatan yang mampu menggambarkan efektivitas kebijakan serta memproyeksikan dampaknya terhadap kelompok sasaran. Analisis prediktif memenuhi kebutuhan tersebut dengan memanfaatkan data historis untuk menilai capaian program, memperkirakan hambatan, dan mengidentifikasi peluang perbaikan sehingga membantu lembaga pemerintah membuat keputusan yang lebih berbasis data.

Penelitian mengenai penerapan *artificial intelligence* dalam administrasi publik menunjukkan bahwa analisis prediktif mampu meningkatkan efektivitas evaluasi program dan kebijakan publik melalui penggunaan model analitis berbasis data historis. Pendekatan ini memungkinkan pemerintah memahami dinamika pelaksanaan program, memproyeksikan dampak kebijakan, serta menghasilkan rekomendasi berbasis data yang mendukung pengambilan keputusan secara lebih akurat dan terukur (Huby & Suriadi, 2025).

2.5. Program Makan Bergizi Gratis

Program Makan Bergizi Gratis merupakan upaya penyediaan makanan bergizi secara teratur kepada siswa sebagai dukungan terhadap kesehatan dan proses pembelajaran. Program ini dirancang sesuai dengan kebutuhan gizi anak usia sekolah dan dilaksanakan melalui mekanisme yang melibatkan sekolah, penyedia makanan, serta pengawasan mutu. Penyediaan makanan bergizi diharapkan mampu memenuhi sebagian kebutuhan energi dan zat gizi harian siswa sehingga meningkatkan kesiapan fisik dan mental dalam mengikuti pembelajaran.

Pelaksanaan Program Makan Bergizi Gratis diharapkan berdampak pada peningkatan kehadiran, konsentrasi, dan hasil belajar siswa, serta menurunkan risiko kekurangan gizi yang dapat menghambat perkembangan akademik. Temuan penelitian menunjukkan bahwa penyediaan makanan sehat di sekolah tidak hanya berperan dalam pemenuhan kebutuhan gizi, tetapi juga memberikan dampak positif terhadap motivasi, kesiapan belajar, kenyamanan, dan antusiasme siswa dalam mengikuti proses pembelajaran (Dian. et al., 2025).



Gambar 2.4. Alur Konseptual Program Makan Bergizi Gratis Sekolah

Sumber: (Yendaw & Dayour, 2015)

Gambar 2.4 menunjukkan alur konseptual Program Makan Bergizi Gratis sebagai intervensi pendukung capaian pendidikan siswa. Penyediaan *school meals* berkontribusi terhadap perbaikan status gizi dan pengurangan rasa lapar jangka pendek, sehingga mendorong peningkatan pendaftaran, kehadiran, serta keterlibatan siswa dalam proses pembelajaran dan retensi materi. Peningkatan asupan gizi juga berpengaruh terhadap kemampuan kognitif dan perilaku belajar siswa. Rangkaian proses tersebut bermuara pada peningkatan capaian pendidikan (*educational achievement*) sebagai hasil akhir pelaksanaan program.

2.5.1. Tantangan dan Isu dalam Pelaksanaan Program Makan Bergizi Gratis

Pelaksanaan Program Makan Bergizi Gratis menghadapi berbagai tantangan, antara lain keterbatasan pendanaan, kualitas dan variasi menu, serta koordinasi antara sekolah dan penyedia makanan. Tantangan tersebut diperkuat oleh kebutuhan standar operasional yang konsisten, keterlibatan tenaga pendukung, dan mekanisme pengawasan yang memadai, yang secara langsung memengaruhi efektivitas

program. Penelitian menunjukkan bahwa keberhasilan program makan bergizi sangat ditentukan oleh kualitas implementasi, terutama konsistensi penyediaan makanan, ketepatan distribusi, dan kesesuaian menu dengan kebutuhan siswa. Temuan tersebut menegaskan pentingnya pengawasan berkelanjutan sebagai faktor kunci dalam meningkatkan efektivitas Program Makan Bergizi Gratis (Rozak et al., 2025).

2.6. Algoritma K-Nearest Neighbors (KNN)

Algoritma *K-Nearest Neighbors* (KNN) merupakan metode klasifikasi yang bekerja dengan mengidentifikasi tetangga terdekat dari data baru berdasarkan ukuran jarak. Algoritma ini mengelompokkan data dengan karakteristik serupa ke dalam kelas yang sama sehingga proses klasifikasi bergantung pada kedekatan nilai antar objek. KNN dikenal sederhana, mudah diterapkan, dan efektif pada berbagai jenis data, serta banyak digunakan dalam bidang pendidikan untuk memprediksi kelulusan, memetakan performa akademik, dan mengidentifikasi risiko belajar.

2.6.1. Konsep Dasar Algoritma KNN

Algoritma *K-Nearest Neighbors* (KNN) bekerja dengan menentukan sejumlah tetangga terdekat berdasarkan ukuran jarak untuk mengelompokkan data baru ke dalam kelas tertentu. Perhitungan jarak antara data uji dan data latih umumnya menggunakan *Euclidean distance*, kemudian penentuan kelas dilakukan berdasarkan mayoritas kelas dari tetangga terdekat. Kinerja algoritma KNN dipengaruhi oleh kualitas data dan pemilihan nilai k yang tepat.

Penelitian mengenai prediksi kelulusan siswa menunjukkan bahwa penggunaan jarak *Euclidean* dan pemilihan nilai k yang optimal berpengaruh signifikan terhadap akurasi model, sehingga KNN mampu mengidentifikasi pola akademik siswa secara lebih tepat (Rizal & Khotimah, 2022). Perhitungan jarak *Euclidean* pada algoritma KNN dinyatakan dengan rumus sebagai berikut:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

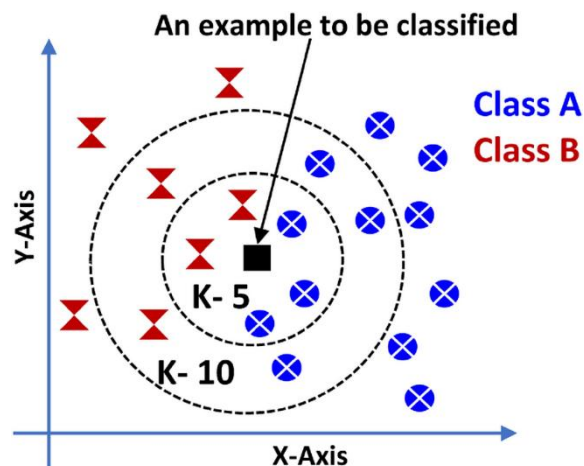
Keterangan:

$d(x, y)$: jarak antara data baru dan data latih

x_i : nilai fitur ke- i pada data baru

y_i : nilai fitur ke- i pada data latih

Semakin kecil nilai jarak, semakin mirip kedua data tersebut



Gambar 2.5. Ilustrasi Mekanisme Klasifikasi Algoritma K-Nearest Neighbors

Sumber: (Taha, 2025)

Gambar 2.5 mengilustrasikan mekanisme kerja algoritma *K-Nearest Neighbors* (KNN) dalam melakukan klasifikasi berdasarkan kedekatan jarak antar

data. Visual tersebut menampilkan sebaran data latih yang telah memiliki kelas, yaitu *Class A* dan *Class B*, serta satu data baru yang belum memiliki kelas. Proses klasifikasi dilakukan dengan menentukan sejumlah tetangga terdekat berdasarkan nilai k yang direpresentasikan melalui lingkaran jarak tertentu. Penentuan kelas data baru dilakukan berdasarkan kelas yang paling dominan di antara tetangga terdekat tersebut. Ilustrasi ini menegaskan bahwa KNN mengandalkan kemiripan karakteristik data tanpa membangun model secara eksplisit, sehingga sesuai digunakan dalam analisis prediktif berbasis kedekatan data.

2.6.2. Kelebihan Algoritma KNN

Beberapa kelebihan algoritma *K-Nearest Neighbors* (KNN) yang mendukung penerapannya dalam penelitian pendidikan dan kesehatan antara lain:

1. Tidak memerlukan proses pelatihan model, karena KNN termasuk algoritma *lazy learning* yang tidak membangun model klasifikasi secara eksplisit, sehingga tidak memerlukan estimasi parameter yang kompleks.
2. Mudah diimplementasikan dan dipahami, karena konsep kerja KNN sederhana dan berbasis perhitungan jarak, sehingga cocok digunakan pada penelitian awal maupun pengembangan sistem klasifikasi.
3. Efektif pada *dataset* kecil hingga menengah, KNN memberikan performa baik ketika ukuran data tidak terlalu besar dan telah dinormalisasi.

2.6.3. Kelemahan Algoritma KNN

Meskipun memiliki kelebihan, algoritma *K-Nearest Neighbors* (KNN) juga memiliki beberapa kelemahan, yaitu:

1. Sensitif terhadap skala dan *noise* data, tanpa proses normalisasi, jarak antar data dapat menghasilkan prediksi yang bias. Kondisi ini dapat mengurangi ketepatan hasil klasifikasi.
2. Kurang efisien untuk *dataset* berukuran besar, KNN harus menghitung jarak setiap *instance* sehingga waktu komputasi meningkat pada *dataset* yang sangat besar. Situasi ini dapat menghambat proses analisis.
3. Rentan terhadap fitur tidak relevan, keberadaan atribut yang tidak signifikan dapat menurunkan akurasi model karena fitur semacam ini dapat mengaburkan pola data utama.

2.7. Algoritma Naive Bayes

Algoritma *Naive Bayes* merupakan metode klasifikasi berbasis probabilitas yang menghitung peluang suatu data masuk ke dalam kategori tertentu berdasarkan distribusi fitur pada data pelatihan. Algoritma ini menggunakan Teorema Bayes dengan asumsi independensi antarfitur, sehingga proses perhitungan menjadi lebih sederhana dan efisien. Karakteristik tersebut menjadikan *Naive Bayes* sebagai algoritma yang ringan dan mampu memberikan performa yang baik, khususnya pada *dataset* berukuran besar.

Bidang pendidikan memanfaatkan *Naive Bayes* untuk memprediksi kategori siswa dan menganalisis performa belajar. Penelitian mengenai prediksi penerima bantuan pendidikan menunjukkan bahwa algoritma ini mampu mengolah data berlabel secara efisien dan menghasilkan klasifikasi yang relevan untuk mendukung pengambilan keputusan berbasis data (Permadi et al., 2025).

2.7.1. Konsep Dasar Algoritma Naive Bayes

Algoritma *Naive Bayes* membangun model klasifikasi menggunakan perhitungan probabilitas dari data pelatihan. Algoritma ini menilai kemungkinan suatu data masuk ke kelas tertentu berdasarkan nilai fitur, kemudian mengalikan peluang tiap fitur terhadap kelas hingga menghasilkan probabilitas akhir sebagai dasar klasifikasi. Pendekatan probabilistik tersebut membuat *Naive Bayes* efektif untuk *dataset* terstruktur dengan distribusi fitur yang stabil. Perhitungan probabilitas pada algoritma *Naive Bayes* didasarkan pada Teorema Bayes berikut:

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

Untuk data dengan banyak fitur, perhitungan probabilitas dilakukan sebagai berikut:

$$P(C | X_1, X_2, \dots, X_n) = P(C) \prod_{i=1}^n P(X_i | C)$$

Keterangan:

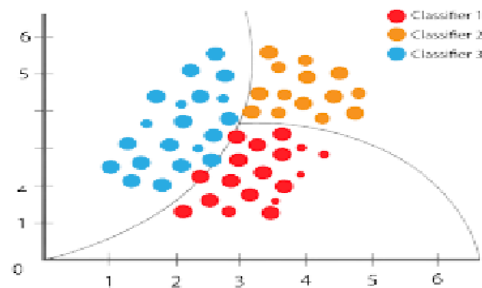
$P(C | X)$: probabilitas data termasuk kelas C berdasarkan fitur X

$P(X | C)$: probabilitas fitur X muncul pada kelas C

$P(C)$: probabilitas awal (prior) kelas C

\prod : simbol perkalian antar fitur

Asumsi utama *Naive Bayes* adalah setiap fitur dianggap saling independen



Gambar 2.6. Ilustrasi Klasifikasi Probabilistik Algoritma Naive Bayes

Sumber: (Widyawati et al., 2023)

Gambar 2.6 menunjukkan ilustrasi mekanisme klasifikasi algoritma *Naive Bayes* yang membedakan data ke dalam beberapa kelas berdasarkan pendekatan probabilistik. Setiap titik data direpresentasikan dalam ruang fitur dan dikelompokkan sesuai kelas dengan probabilitas *posterior* tertinggi. Garis batas keputusan pada gambar menggambarkan pemisahan kelas yang dihasilkan dari perbandingan nilai probabilitas, menegaskan bahwa proses klasifikasi *Naive Bayes* tidak bergantung pada jarak antar data, melainkan pada peluang statistik.

2.7.2. Kelebihan Algoritma Naive Bayes

Beberapa kelebihan algoritma *Naive Bayes* yang menjadikannya banyak digunakan dalam bidang kesehatan dan pendidikan antara lain:

1. Cepat dan efisien, karena proses pelatihan dan klasifikasi dilakukan menggunakan perhitungan probabilitas sederhana, sehingga sesuai untuk pengolahan data dalam jumlah besar maupun analisis *real-time*.
2. Performa stabil pada data kategorikal, terutama pada *dataset* dengan fitur diskrit seperti data survei, angket, dan data administrasi pendidikan.

3. Relatif lebih tahan terhadap *overfitting*, karena struktur model yang sederhana dan jumlah parameter yang terbatas, sehingga cocok untuk *dataset* dengan dimensi fitur yang cukup besar.

2.7.3. Kelemahan Algoritma Naive Bayes

Meskipun memiliki sejumlah kelebihan, algoritma *Naive Bayes* juga memiliki beberapa kelemahan yang perlu diperhatikan, antara lain:

1. Asumsi independensi fitur yang sering tidak terpenuhi, karena dalam kondisi nyata banyak fitur saling berkorelasi, sehingga pelanggaran asumsi ini dapat menurunkan akurasi klasifikasi.
2. Kurang optimal pada data dengan distribusi kompleks, terutama ketika distribusi probabilitas fitur tidak mengikuti pola yang sederhana atau bersifat nonlinier.
3. Masalah *zero probability*, yaitu kondisi ketika suatu fitur tidak pernah muncul pada data pelatihan sehingga menghasilkan probabilitas nol, yang dapat memengaruhi hasil prediksi jika tidak ditangani dengan teknik tertentu seperti *smoothing*.

2.8. Evaluasi Model Klasifikasi

Evaluasi model klasifikasi bertujuan untuk menilai sejauh mana algoritma mampu menghasilkan prediksi yang benar berdasarkan pola data yang dianalisis. Proses ini berperan penting dalam memahami tingkat kesalahan model serta menentukan kelayakan algoritma yang digunakan dalam konteks penelitian. Evaluasi dilakukan secara terukur menggunakan *confusion matrix* dan metrik evaluasi agar performa model dapat dinilai secara lebih objektif.

		Predicated Class		
		Positive	Negative	
Actual Class	Positive	True Positive	False Negative Type II Error	Sensitivity TP $\frac{TP}{TP+FN}$
	Negative	False Positive Type I Error	True Negative	Specificity TN $\frac{TN}{TN+FP}$
		Precision TP $\frac{TP}{TP+FP}$	Negative Predictive Value TN $\frac{TN}{TN+FN}$	Accuracy TP $\frac{TP}{TP+TN+FP+FN}$

Gambar 2.7. Confusion Matrix sebagai Dasar Evaluasi Model Klasifikasi

Sumber: (Ratnakar, 2024)

Gambar 2.7 menunjukkan *confusion matrix* sebagai alat utama dalam evaluasi model klasifikasi yang membandingkan kelas aktual dengan kelas hasil prediksi. Matriks ini terdiri atas empat komponen utama, yaitu *true positive* (TP), *true negative* (TN), *false positive* (FP), dan *false negative* (FN). Keempat komponen tersebut digunakan untuk mengidentifikasi jumlah prediksi yang benar maupun kesalahan klasifikasi yang dihasilkan oleh model. Melalui *confusion matrix*, berbagai metrik evaluasi dapat dihitung untuk menilai performa model secara objektif.

Penggunaan *confusion matrix* dan metrik evaluasi memberikan gambaran menyeluruh terhadap kualitas prediksi model klasifikasi serta membantu menilai efektivitas algoritma secara akurat dalam mendukung pengambilan keputusan berbasis data (Naseer et al., 2024). Metrik evaluasi yang digunakan dalam penelitian meliputi *accuracy*, *precision*, *recall*, dan *F1-score*. Rumus masing-masing metrik adalah sebagai berikut:

1. *Accuracy* mengukur tingkat ketepatan model dalam melakukan klasifikasi dengan membandingkan jumlah prediksi yang benar terhadap keseluruhan data.

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

2. *Precision* menunjukkan tingkat ketepatan model dalam memprediksi kelas positif, yaitu perbandingan antara data positif yang diprediksi benar dengan seluruh data yang diprediksi sebagai positif.

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

3. *Recall* adalah menggambarkan kemampuan model dalam mengenali seluruh data positif secara benar dengan membandingkan jumlah data positif yang diprediksi benar terhadap total data positif.

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

4. *F1-score* merupakan rata-rata harmonis antara *precision* dan *recall* yang digunakan untuk menilai keseimbangan performa model, terutama pada *dataset* dengan distribusi kelas yang tidak seimbang.

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

2.9. Penelitian Terdahulu dan Perbandingan Dengan Penelitian Ini

Berikut merupakan rangkuman penelitian terdahulu yang relevan dengan topik penelitian ini. Tabel ini disajikan untuk menunjukkan posisi penelitian ini dibandingkan studi sebelumnya.

Tabel 2.1 Penelitian Terdahulu

Penelitian	Judul	Tahun	Data dan Metode yang Digunakan	Hasil
Dian. Kurnianto, Dian Apriani, Feny Dinoto, Muad	Makna Program Makan Bergizi Gratis terhadap Peningkatan Hasil Belajar Siswa SMK Menuju Generasi Emas 2045	2025	Penelitian kualitatif dengan pendekatan fenomenologis dilakukan melalui wawancara, observasi, dan dokumentasi pada dua SMK untuk memahami dampak Program Makan Bergizi terhadap siswa.	Program makan bergizi meningkatkan fokus, energi, motivasi, dan prestasi siswa, sementara sekolah tanpa program mengalami kelelahan, pasif belajar, dan prestasi rendah.
Rozak, Abdur Izzah, Armiya Nur Lailatul Chusna, Cahya Amalia	Dampak Program Makan Bergizi Gratis terhadap Konsentrasi Siswa Kelas III Sekolah Dasar	2025	Penelitian kualitatif dengan desain studi kasus dilakukan melalui wawancara, observasi partisipatif, dan dokumentasi dengan melibatkan informan siswa, guru, serta orang tua di SD Islam Al Hikmah.	Program makan bergizi meningkatkan fokus visual, keterlibatan aktif, stabilitas perilaku, daya tahan konsentrasi, kesiapan fisik dan mental serta motivasi belajar siswa secara signifikan.
Permadi, M Rizal Mahendra, Safira Putri Qondias, Dimas	Perbedaan Tingkat Konsumsi Zat Gizi Makro, Status Gizi dan Indeks Prestasi antara <i>School Feeding</i> dan	2025	Penelitian menggunakan <i>Systematic Literature Review</i> dengan pencarian <i>Google Scholar</i> dan <i>Science Direct</i> , melalui identifikasi	<i>School feeding</i> meningkatkan konsumsi zat gizi makro dan status gizi anak, tetapi tidak menunjukkan

	<i>Non School Feeding</i> Anak Usia Sekolah Dasar		masalah, screening, penilaian kualitas, dan ekstraksi delapan literatur.	perbedaan indeks prestasi dibanding <i>non school feeding</i> .
Atmaja, I Putu Yogista Putra Setiawan, Gde Iwam Dika, I Wayan Imawati, Ida Ayu Putu Febri	<i>Data Mining</i> Memprediksi Kelulusan Mahasiswa Menggunakan Metode <i>K-Nearest Neighbors</i> (KNN)	2023	Penelitian menggunakan metode CRISP-DM melalui observasi, studi literatur, persiapan data, pemodelan KNN, evaluasi akurasi dan presisi, serta <i>deployment</i> untuk prediksi kelulusan mahasiswa.	KNN mampu memprediksi kelulusan tepat atau terlambat berdasarkan data akademik; hasil stabil meski nilai <i>K</i> berubah mengikuti tetangga terdekat dominan.
Julianto, Alvin Andayani, Sri	Penerapan <i>Data Mining</i> untuk Klasifikasi Produk Terlaris Menggunakan Algoritma <i>Naive Bayes</i> pada Bengkel Motor	2024	Penelitian memakai tahapan KDD dengan seleksi, <i>preprocessing</i> , integrasi, transformasi data, lalu klasifikasi <i>Naive Bayes</i> menggunakan <i>Rapid Miner</i> pada data penjualan bengkel.	<i>Naive Bayes</i> mencapai akurasi 86,77%, presisi 85,39%, <i>recall</i> 76%, sehingga efektif menentukan produk laris dan tidak laris pada bengkel.

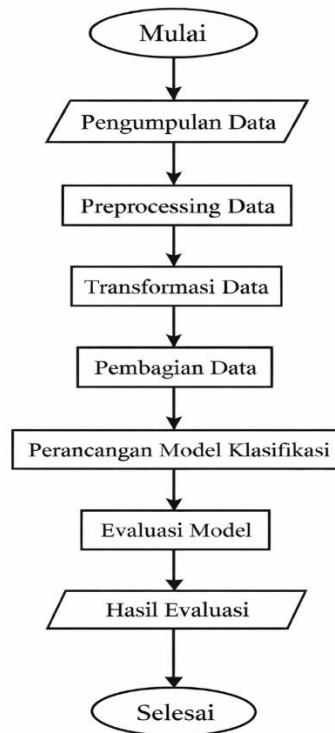
Penelitian ini memiliki kebaruan (*novelty*) pada penerapan analisis prediktif dengan pendekatan klasifikasi berbasis *data mining* untuk mengevaluasi efektivitas Program Makan Bergizi Gratis di tingkat sekolah menengah atas. Pendekatan tersebut belum pernah diterapkan dalam penelitian sejenis. Kajian terdahulu mengenai program makan bergizi gratis atau program gizi sekolah umumnya menggunakan pendekatan kualitatif dan deskriptif untuk menilai dampaknya

terhadap motivasi belajar, konsentrasi, status gizi, dan hasil belajar siswa. Oleh karena itu, penelitian ini menghadirkan pendekatan kuantitatif berbasis *machine learning* untuk mengklasifikasikan tingkat efektivitas program secara lebih terukur pada populasi siswa SMA.

Kebaruan penelitian ini juga terletak pada penggunaan dua algoritma klasifikasi, yaitu *K-Nearest Neighbors* dan *Naive Bayes*, yang dianalisis secara komparatif untuk menentukan model terbaik dalam mengklasifikasikan tingkat efektivitas program makan bergizi gratis. Penelitian terdahulu pada bidang *data mining* umumnya hanya menggunakan satu algoritma dalam satu konteks tertentu, seperti prediksi kelulusan atau klasifikasi data penjualan. Sementara itu, penelitian ini mengintegrasikan kedua algoritma tersebut dalam konteks evaluasi program intervensi gizi, sehingga memberikan kontribusi baru dalam penerapan metode klasifikasi pada bidang pendidikan dan kesehatan.

2.10. Kerangka Penelitian

Kerangka penelitian disusun untuk menggambarkan alur umum penelitian dalam menganalisis efektivitas Program Makan Bergizi Gratis menggunakan pendekatan analisis prediktif dengan metode klasifikasi. Kerangka ini memberikan gambaran konseptual mengenai tahapan penelitian yang dilaksanakan secara sistematis, mulai dari pengumpulan data hingga evaluasi hasil klasifikasi sebagai dasar penarikan kesimpulan. Penyusunan kerangka penelitian memastikan proses penelitian berjalan terarah, terstruktur, dan selaras dengan tujuan penelitian yang telah ditetapkan.



Gambar 2.8. Flowchart Kerangka Penelitian

Gambar 2.8 menunjukkan alur kerangka penelitian yang diawali dengan tahap pengumpulan data sebagai proses menghimpun informasi yang relevan dengan pelaksanaan Program Makan Bergizi Gratis dan kondisi siswa. Data yang diperoleh kemudian diproses melalui tahap *preprocessing* untuk memastikan kualitas data sebelum dilakukan analisis. Tahap transformasi data dilakukan untuk menyesuaikan format data agar siap digunakan dalam proses analisis lebih lanjut.

Pembagian data dilakukan sebagai bagian dari analisis prediktif dengan pendekatan klasifikasi sebelum perancangan model. Perancangan model bertujuan membangun model klasifikasi yang digunakan untuk mengelompokkan tingkat efektivitas program, kemudian dilanjutkan dengan tahap evaluasi model untuk menilai kinerja model klasifikasi yang dihasilkan. Hasil evaluasi menjadi dasar penarikan

kesimpulan mengenai tingkat efektivitas Program Makan Bergizi Gratis, sedangkan tahap akhir menandai selesainya seluruh rangkaian proses penelitian.

2.11. Tools Pendukung Penelitian

Penelitian ini memanfaatkan beberapa perangkat lunak untuk mendukung proses klasifikasi data sosial menggunakan algoritma *K-Nearest Neighbors* (KNN) dan *Naive Bayes*. Perangkat lunak tersebut digunakan dalam pengolahan data, pembangunan model, serta evaluasi hasil analisis dengan tujuan meningkatkan akurasi perhitungan, mempercepat proses pemodelan, dan memudahkan visualisasi data. *Tools* yang digunakan dalam penelitian ini adalah sebagai berikut:

1. Microsoft Excel

Microsoft Excel digunakan pada tahap awal penelitian untuk melakukan input dan pembersihan data (*data cleaning*), pengolahan statistik dasar, serta penyusunan distribusi frekuensi dan perhitungan probabilitas sederhana. Aplikasi ini juga dimanfaatkan dalam pembuatan *confusion matrix* secara manual guna menyiapkan *dataset* secara sistematis sebelum proses pemodelan dilakukan menggunakan perangkat lunak analisis lanjutan.



Gambar 2.9. Aplikasi Microsoft Excel

Gambar 2.9 menunjukkan tampilan antar muka Microsoft Excel yang digunakan dalam proses pengolahan dan persiapan data penelitian.

2. RapidMiner

RapidMiner digunakan sebagai platform utama untuk membangun dan mengevaluasi model *K-Nearest Neighbors* (KNN) dan *Naive Bayes*, karena mendukung *preprocessing* data serta perhitungan metrik evaluasi seperti akurasi, presisi, *recall*, dan *confusion matrix* secara efisien tanpa penulisan kode.



Gambar 2.10. Aplikasi RapidMiner

Gambar 2.10 menunjukkan tampilan antar muka RapidMiner yang digunakan dalam proses pemodelan dan evaluasi algoritma klasifikasi pada penelitian.