

## BAB II LANDASAN TEORI

### 2.1 Tinjauan Pustaka

Penelitian sebelumnya menjadi acuan penting dalam pengembangan penelitian ini, khususnya yang berkaitan dengan klasifikasi data akademik menggunakan metode *K-Nearest Neighbor (KNN)*, antara lain :

1. Pada penelitian (5) dengan judul “Klasifikasi Prestasi Akademik Siswa Berdasarkan Nilai Rapor dan Kedisiplinan dengan Metode *K-Nearest Neighbor*”. Penelitian ini melakukan klasifikasi prestasi akademik siswa berdasarkan nilai rapor dan kedisiplinan dengan menggunakan data sebanyak 348 siswa SMA Negeri 2 Batu. Hasil pengujian menunjukkan nilai rata-rata akurasi sebesar 87,73%, precision sebesar 81,78%, recall 67,00%, dan F1-score 73,38%. Pengujian terbaik diperoleh pada data split 70:30 dengan akurasi mencapai 91,39%. Hasil ini membuktikan bahwa algoritma KNN efektif dalam memprediksi prestasi akademik siswa berdasarkan data akademik dan kedisiplinan.
2. Pada penelitian (6) dengan judul “Pengembangan Website dengan Algoritma *K-Nearest Neighbor (KNN)* dalam Klasifikasi Perkembangan Prestasi Siswa terhadap Hasil Belajar (Studi Kasus SD Negeri 107396 Paluh Merbau)”. Penelitian ini mengembangkan sistem berbasis web untuk mengklasifikasikan perkembangan prestasi siswa menggunakan KNN dengan parameter  $k = 3$ . Hasil klasifikasi menunjukkan akurasi sebesar 78%, dengan variabel signifikan meliputi nilai ujian semester, ujian harian, tugas, sikap, dan absensi. Sistem berbasis KNN ini meningkatkan efisiensi kerja guru dalam mengelola dan memvisualisasikan data siswa secara interaktif.
3. Pada penelitian (7) dengan judul “Analisis Klasifikasi Kelulusan Mahasiswa Menggunakan Algoritma *K-Nearest Neighbor* di Fakultas Teknik Universitas Negeri Manado”. Penelitian ini mengidentifikasi pola kelulusan mahasiswa menggunakan KNN. Hasil menunjukkan bahwa sebagian besar mahasiswa tidak lulus tepat waktu meskipun

memiliki IPK tinggi. Atribut seperti jenis kelamin, jenis pembiayaan, IPK, dan IPS berpengaruh terhadap klasifikasi. KNN mampu mengklasifikasikan status kelulusan dengan tingkat akurasi yang cukup baik, menunjukkan potensinya untuk analisis akademik mahasiswa.

4. Pada penelitian (8) dengan judul “Algoritma *K-Nearest Neighbor Classification* sebagai Sistem Pengelompokan Kemampuan Akademik Siswa Berbasis Web”. Penelitian ini berhasil mengembangkan sistem klasifikasi kemampuan akademik siswa berbasis web yang mengelompokkan siswa ke dalam lima kategori berdasarkan nilai rata-rata. Model klasifikasi KNN menunjukkan performa tinggi dengan akurasi sebesar 90,19%, dan mampu memberikan rekomendasi treatment pembelajaran yang sesuai dengan kemampuan siswa.
5. Pada penelitian (9) dengan judul “Implementasi Algoritma Klasifikasi *K-Nearest Neighbor (KNN)* untuk Klasifikasi Seleksi Penerima Beasiswa”. Penelitian ini menghasilkan aplikasi seleksi penerima beasiswa berbasis algoritma KNN. Evaluasi menggunakan confusion matrix menunjukkan rata-rata akurasi sebesar 90,5% dan precision sebesar 89%. Hasil ini menunjukkan bahwa KNN memiliki kinerja stabil dalam menentukan keputusan berbasis data akademik dan administratif.
6. Pada penelitian (10) “Prediksi Kelulusan Mata Kuliah Mahasiswa Teknologi Informasi Menggunakan Algoritma *K-Nearest Neighbor*”. Penelitian ini membangun sistem prediksi kelulusan berbasis PHP dengan konsep data mining KNN. Dengan 20 data uji, diperoleh akurasi sebesar 80% pada  $k = 3$  dan 90% pada  $k = 5$ . Hasil menunjukkan bahwa semakin besar nilai  $k$  dan jumlah data yang digunakan, semakin optimal hasil akurasi prediksi yang dicapai.
7. Pada penelitian (11) “Prediksi Kelulusan Mahasiswa dengan Metode *K-Nearest Neighbor* pada Jurusan Sistem Informasi Institut XYZ”. Penelitian ini mengembangkan sistem berbasis web untuk memprediksi kelulusan mahasiswa dengan metode KNN di Institut Teknologi Adhi

Tama Surabaya. Hasil penelitian menunjukkan nilai akurasi tertinggi sebesar 95,15% dengan 106 data mahasiswa. KNN terbukti efektif dalam memprediksi kelulusan tepat waktu mahasiswa.

8. Pada penelitian (12) “Klasifikasi Kemampuan Berpikir Kreatif Peserta Didik Madrasah Ibtidaiyah Menggunakan Algoritma *K-Nearest Neighbor*”. Penelitian ini menggunakan algoritma KNN dengan parameter  $k = 5$  dan menghasilkan akurasi sebesar 86%. Atribut yang paling berpengaruh adalah indikator orisinalitas dan kelancaran berpikir. Penelitian ini menegaskan bahwa KNN merupakan metode sederhana namun efektif untuk menganalisis kemampuan berpikir kreatif siswa secara objektif dan berbasis data empiris.

Berdasarkan berbagai penelitian di atas, dapat disimpulkan bahwa algoritma *K-Nearest Neighbor (KNN)* memiliki tingkat akurasi tinggi dalam klasifikasi data akademik, baik untuk memprediksi prestasi, kemampuan, maupun kelulusan siswa dan mahasiswa. Nilai akurasi yang dihasilkan dari berbagai penelitian berada pada kisaran 78% hingga 95,15%, dengan rata-rata akurasi keseluruhan sekitar 80,92%. Hal ini menunjukkan bahwa KNN merupakan metode yang andal, sederhana, dan fleksibel dalam mendukung proses analisis dan pengambilan keputusan di bidang pendidikan.

## **2.2 Landasan Teori**

### **2.2.1 Pendidikan dan Prestasi Akademik**

Pendidikan merupakan proses yang berlangsung secara sadar, terencana, dan sistematis untuk menciptakan suasana belajar yang memungkinkan peserta didik mengembangkan potensi dirinya secara optimal. Hal ini sejalan dengan Undang-Undang Republik Indonesia Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional, yang menegaskan bahwa pendidikan berfungsi mengembangkan kemampuan serta membentuk watak dan peradaban bangsa yang bermartabat dalam rangka mencerdaskan kehidupan bangsa (1). Dengan demikian, pendidikan tidak hanya menekankan pada penguasaan pengetahuan, tetapi juga pada pembentukan karakter dan kualitas individu secara menyeluruh.

Prestasi akademik merupakan salah satu indikator penting untuk mengukur keberhasilan proses pendidikan. Menurut Slameto, prestasi belajar dipengaruhi oleh dua kelompok faktor utama, yaitu faktor internal dan faktor eksternal. Faktor internal mencakup aspek psikologis, motivasi, minat, bakat, serta kondisi fisiologis peserta didik, sedangkan faktor eksternal berkaitan dengan lingkungan keluarga, sekolah, masyarakat, serta sarana dan prasarana yang menunjang proses pembelajaran. Slameto juga menyatakan bahwa kebiasaan belajar merupakan bagian dari faktor internal yang turut menentukan keberhasilan belajar, sementara lingkungan belajar terdiri atas tiga indikator utama, yaitu lingkungan keluarga, sekolah, dan masyarakat (13).

Dalam konteks pendidikan modern, prestasi akademik umumnya diukur melalui nilai yang tercantum dalam rapor atau hasil evaluasi pembelajaran lainnya. Nilai rapor dianggap sebagai indikator kuantitatif yang dapat menggambarkan tingkat pencapaian peserta didik pada mata pelajaran tertentu. Selain itu, nilai rapor juga dapat diolah menggunakan metode analisis lanjutan, termasuk teknik data mining, untuk menemukan pola, kecenderungan, atau karakteristik tertentu yang berkaitan dengan hasil belajar siswa (5). Temuan dari analisis tersebut dapat digunakan sebagai dasar untuk meningkatkan efektivitas pembelajaran maupun kebijakan pendidikan.

Dengan demikian, pendidikan sebagai proses pengembangan potensi peserta didik serta prestasi akademik sebagai indikator keberhasilan proses tersebut memiliki keterkaitan yang erat. Dukungan lingkungan belajar, kebiasaan belajar, serta pemanfaatan teknologi analitik modern turut berperan dalam memahami dan meningkatkan pencapaian akademik peserta didik.

### 2.2.2 Sistem Informasi

Sistem informasi adalah kombinasi antara teknologi, manusia, dan prosedur yang bekerja bersama untuk mengumpulkan, mengolah, menyimpan, dan mendistribusikan informasi guna mendukung pengambilan keputusan dalam organisasi (14). Sistem informasi digunakan untuk

meningkatkan efektivitas dan efisiensi proses kerja, termasuk dalam bidang pendidikan.

Dalam konteks akademik, sistem informasi dapat dimanfaatkan untuk mengelola data siswa, data nilai, hasil evaluasi, maupun proses pengolahan data akademik lainnya sehingga informasi yang dihasilkan lebih akurat dan mudah diakses (15).

### 2.2.3 Data Mining

Data mining adalah proses untuk menemukan pola, pengetahuan, atau informasi bermakna dari kumpulan data dalam jumlah besar melalui teknik statistik, kecerdasan buatan, dan *machine learning* (16). Data mining bertujuan mengubah data mentah menjadi informasi yang bernilai untuk mendukung pengambilan keputusan. Proses data mining meliputi beberapa tahapan, yaitu seleksi data, transformasi, data mining (penggalian pola), dan evaluasi hasil (17).

Dalam konteks pendidikan, data mining dapat digunakan untuk menganalisis data akademik seperti nilai rapor, absensi, maupun data perilaku siswa untuk menemukan pola prestasi dan memprediksi perkembangan akademik (5). Pemanfaatan data mining memberikan dukungan bagi pendidik dan pihak sekolah dalam merumuskan kebijakan maupun treatment pembelajaran yang berbasis data. Dengan demikian, data mining merupakan teknik analitik yang efektif untuk mengolah data akademik dalam skala besar menjadi informasi yang akurat dan relevan bagi proses evaluasi maupun prediksi.

### 2.2.4 Klasifikasi

Klasifikasi merupakan salah satu teknik dalam data mining yang bertujuan untuk mengelompokkan objek atau data ke dalam kelas tertentu berdasarkan atribut yang dimilikinya. Proses klasifikasi bekerja dengan membangun suatu model dari data pelatihan (*training data*) yang telah memiliki label kelas, kemudian model tersebut digunakan untuk memprediksi kelas dari data baru yang belum diketahui labelnya (16). Dengan demikian, klasifikasi berfungsi sebagai proses pemetaan dari sekumpulan data input

menuju kategori atau kelas yang sudah ditentukan sebelumnya. (18) menjelaskan bahwa klasifikasi umumnya ditujukan untuk mengidentifikasi pola atau hubungan dalam data sehingga dapat mendukung keputusan prediktif secara objektif.

Metode klasifikasi memiliki peran penting dalam berbagai bidang, termasuk pendidikan, bisnis, kesehatan, dan keuangan. Dalam konteks pendidikan, klasifikasi dapat dimanfaatkan untuk memprediksi prestasi akademik siswa, tingkat kemampuan belajar, hingga potensi kelulusan berdasarkan data historis seperti nilai rapor, absensi, maupun data perilaku siswa (5). Beberapa algoritma klasifikasi yang sering digunakan antara lain *Decision Tree*, *Naïve Bayes*, *Support Vector Machine (SVM)*, dan *K-Nearest Neighbor (KNN)*, di mana masing-masing algoritma memiliki karakteristik dan performa yang berbeda sesuai kebutuhan penelitian (17). Oleh karena itu, pemilihan algoritma klasifikasi yang tepat perlu mempertimbangkan jenis data, jumlah data, kompleksitas variabel, serta tingkat akurasi yang diharapkan pada penelitian.

#### 2.2.5 Algoritma *K-Nearest Neighbor (KNN)*

Algoritma *K-Nearest Neighbor (KNN)* merupakan salah satu metode klasifikasi dalam data mining yang bekerja berdasarkan prinsip kedekatan jarak antara data baru dan data yang telah ada sebelumnya. KNN termasuk ke dalam kategori *supervised learning* dan *lazy learning*, karena proses pembelajaran tidak dilakukan pada tahap pelatihan, melainkan pada saat proses klasifikasi berlangsung. Dengan kata lain, KNN tidak membangun model secara eksplisit, tetapi menyimpan seluruh data latih dan melakukan pencarian kelas pada saat data uji diberikan (19).

Proses klasifikasi pada KNN dilakukan dengan mengukur jarak antara data uji dengan seluruh data latih menggunakan perhitungan jarak, seperti *Euclidean Distance*, kemudian memilih sejumlah tetangga terdekat berdasarkan nilai  $k$ , dan hasil klasifikasi ditentukan berdasarkan mayoritas kelas dari tetangga yang terpilih (18). Agar perhitungan jarak berlangsung secara proporsional, seluruh atribut perlu terlebih dahulu dinormalisasi

sehingga berada pada skala yang sama. Setelah proses normalisasi dilakukan, KNN menghitung tingkat kemiripan antara data uji dan data latih menggunakan *Euclidean Distance*, yang merupakan metode pengukuran jarak paling umum digunakan dalam penelitian pendidikan berbasis data numerik. Rumus *Euclidean Distance* adalah sebagai berikut:

$$d(p, q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (2.1)$$

Dimana :

$d(p, q)$  = jarak antara data uji (p) dan data latih (q)

$p_1 \dots p_n$  = nilai atribut pada data uji

$q_1 \dots q_n$  = nilai atribut pada data latih

$n$  = jumlah atribut

Tahap berikutnya adalah menentukan nilai k, yaitu jumlah tetangga terdekat yang dipertimbangkan dalam klasifikasi. Setelah diperoleh k data dengan jarak paling kecil, penentuan kelas dilakukan melalui mekanisme voting mayoritas, yaitu memilih kelas yang paling banyak muncul di antara tetangga tersebut. Voting mayoritas dirumuskan sebagai:

$$\hat{y} = \text{mode}(y_1, y_2, \dots, y_k) \quad (2.2)$$

Dimana :

$\hat{y}$  = kelas hasil prediksi

$y_1 \dots y_k$  = kelas dari k tetangga terdekat

KNN banyak digunakan karena karakteristiknya yang sederhana, fleksibel, dan mampu memberikan akurasi yang tinggi, terutama pada data numerik yang memiliki distribusi tidak kompleks. Selain itu, KNN dapat bekerja efektif tanpa memerlukan banyak parameter untuk pengaturan model, sehingga sering menjadi pilihan pada penelitian klasifikasi data akademik

(16). Beberapa penelitian pada bidang pendidikan juga menunjukkan bahwa KNN mampu melakukan prediksi prestasi peserta didik, menentukan tingkat kemampuan akademik, hingga memetakan kelulusan dengan hasil akurasi yang tinggi, yaitu berada pada kisaran 78%–95% (5,8). Temuan tersebut menunjukkan bahwa algoritma KNN merupakan metode klasifikasi yang andal dalam mengolah data akademik dan mendukung proses pengambilan keputusan berbasis data.

