

BAB II LANDASAN TEORI

2.1 Tinjauan Pustaka

Penelitian ini mengacu pada penelitian lain sebagai referensi, salah satu penelitian yang sejenis dilakukan oleh Nazarudin Ahmad, Saifan Hafizh & Rana Sulthanah Ahmad et al., (2024) dengan judul “Prediksi Kelulusan Mata Kuliah Mahasiswa Teknologi Informasi Menggunakan Algoritma K-Nearest Neighbor” . Penelitian ini membahas prediksi kelulusan mata kuliah mahasiswa program studi teknik informasi . Penelitian ini menunjukkan metode K-NN memiliki akurasi keseluruhan sebesar 90%, dengan menggunakan $k = 5$ dengan data latih sebanyak 40 data dan data uji sebanyak 20 data.

Penelitian selanjutnya yang dilakukan oleh Luqman Abdur R A at al ., (2023) dengan judul “Faktor Faktor yang Mempengaruhi Minat Calon Mahasiswa Baru Mendaftar pada FTII UHAMKA Menggunakan Algoritma *K-Nearest Neighbors* (K-NN). Penelitian ini membahas faktor apa saja yang dapat mempengaruhi pendaftar di FTII (Fakultas Teknik Industri dan Informatika) dengan menggunakan 10 atribut yaitu nama lengkap, jurusan, biaya kuliah, Lokasi FTII dengan domisili, terdapat teman/keluarga, akreditasi, fasilitas, pelayanan pmb, informasi pmb, dan keterangan minat. Metode yang digunakan adalah K-NN kemudian dataset ditransformasikan menggunakan split data dengan perbandingan 80% data training sebanyak 95 data dan 20% data testing sebanyak 24 data. Dari hasil pengujian penelitian menggunakan confusion matrix memperoleh hasil rata rata akurasi sebesar 72,5%.

Penelitian selanjutnya yang dilakukan oleh Farhan Hidayat Zulfallah (2022) dengan judul “Implementasi Algoritma KNN Dalam Mengukur Ketetapan Kelulusan Mahasiswa UIN Syarif Hidayatullah Jakarta” Zulfallah, (2022). Pada penelitian ini menunjukkan bahwa pengukuran ketepatan dengan metode K-NN, Stratified K-Fold Cross Validation dan perhitungan confusion matrix didapatkan hasil rata rata dari akurasi sebesar 98,38%.

Penelitian selanjutnya yang dilakukan oleh Mulyati et al., (2020) dengan judul “Rancang Bangun Aplikasi Data Mining Prediksi Kelulusan Ujian Nasional Menggunakan Algoritma (KNN) *K-Nearest Neighbors* Dengan Metode Enclidean Distance pada SMPN 2 Pagedangan”. Penelitian ini membahas tentang penggunaan metode penentuan jarak dengan metode Enclidean Distance menggunakan algoritma KNN sangat membantu masalah efisien waktu dan efektif dalam memprediksi kelulusan. Hasil penelitian ini dengan menggunakan nilai try out siswa 3 tahun terakhir yang berjumlah 744 data menghasilkan nilai rata-rata akurasi dengan nilai k sebesar 7 sebesar 96.26%, presisi sebesar 96.17%, recall sebesar 97.32%.

Berdasarkan penjelasan sebelumnya tentang perbedaan dari beberapa penelitian yang telah dilaksanakan sebelumnya. Perbedaan yang dimiliki dari penelitian ini adalah memprediksi kelulusan dengan penerapan algoritma K-NN. Aplikasi yang digunakan adalah *Jupiter Notebook* di *Visual Studio Code*. Menurut penelitian dari Esti Purwaningsih dan Ela Nurelasari yang berjudul “Penerapan K-Nearest Neighbor Untuk Klasifikasi Tingkat Kelulusan Pada Siswa” tahun 2021, Algoritma K-NN mempunyai kinerja yang baik dalam klasifikasi untuk menangani dataset dengan jumlah kecil. Hasil yang diharapkan dari penelitian ini yaitu memprediksi kelulusan mahasiswa di Universitas Sahid Surakarta.

1. Indeks Prestasi Kumulatif (IPK)

Menurut Sutja & Harahap, (2024) mendefinisikan bahwa IPK didapat oleh mahasiswa melalui berbagai faktor baik secara internal seperti motivasi belajar maupun eksternal yang berkaitan dengan lingkungan individu.

2. Indeks Prestasi Semester (IPS)

Menurut Syamsul Huda & Devy Habibi Muhammad, (2022) mendefinisikan bahwa IPS indeks prestasi yang diperoleh mahasiswa setiap semester berdasarkan nilai mata kuliah yang diambil

Menurut Yunita & Mudlikah, (2020) mendefinisikan bahwa IPS merupakan ukuran keberhasilan mahasiswa dalam menempuh mata kuliah pada satu semester.

3. Satuan Kredit Semester (SKS)

Menurut Primajaya et al., (2019) mendefinisikan bahwa SKS sebagai bentuk penyelenggaraan pendidikan di mana peserta didik menyepakati jumlah beban belajar yang diikuti dan/atau strategi belajar setiap semester pada satuan semester sesuai dengan bakat, minat, dan kemampuan belajarnya.

Menurut Eko Wahyudi & Riayatul Husnan, (2022) mendefinisikan bahwa SKS adalah bentuk penyelenggaraan Pendidikan yang peserta didiknya memilih jumlah beban dan Pelajaran yang diikuti setiap semester.

Menurut Nisa & Pramono, (2023) mendefinikan bahwa SKS adalah salah satu bentuk penyelenggaraan pada bidang Pendidikan yang memberikan peluang kepada peserta didik untuk menentukan jumlah beban belajar serta mata Pelajaran yang diikuti setiap semester.

4. KNN (K-Nearest Neighbors)

Menurut Andi Maulida Argina (2020) mendefiniksikan bahwa sebuah metode klasifikasi terhadap sekumpulan data berdasarkan pembelajaran data yang sudah terklasifikasikan sebelumnya. Termasuk dalam supervised learning, dimana hasil query instance yang baru diklasifikasikan berdasarkan mayoritas kedekatan jarak dari kategori yang ada dalam K-NN

Pada algoritma K-NN terdapat 5 cara untuk mencari tetangga terdekat yaitu :

- Jarak Euclidean
- Jarak Manhattan
- Jarak Cosine
- Jarak Corelation
- Jarak Hamming

Pada penelitian ini menggunakan jarak *Euclidean*, maka rumus perhitungan jarak dengan *Euclidean* seperti dibawah ini :

$$euc = \sqrt{\sum_{i=1} (p_i - q_i)^2} \dots\dots\dots(Persamaan 1)$$

dimana :

p_i = sample data / data training

q_i = data uji / data testing

i = variabel data

Σ = dimensi data

Menurut Yudhistira Adhitya & Charles Chaywin, (2023) perhitungan untuk algoritma *K-Nearest Neighbors* :

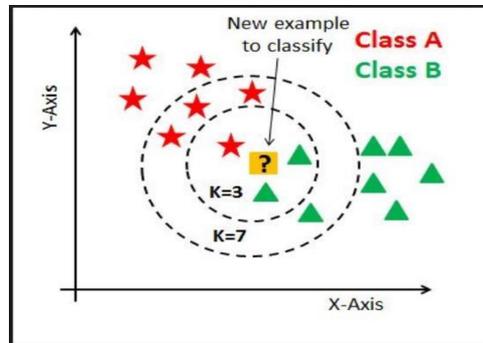
- a. Tentukan nilai K
- b. Dengan data training , tentukan jarak Enclidean kuadrat setiap objek.
- c. Kelompokkan item-item ini ke dalam grup berdasarkan jarak terdekat
- d. Dengan memanfaatkan kategori Nearest Neighbors dengan nilai instance query terhitung paling akurat.

5. Konsep KNN (*K-Nearest Neighbors*)

Menurut Esty Purwaningsih & Ela Nurelasari (2021) menguraikan bahwa konsep KNN tergolong sebagai konsep klasifikasi yang sederhana dan menjadi populer dalam penelitian terkait klasifikasi. Menurut Imandoust dan Bolandraftar terdapat keunggulan yang dimiliki diantaranya yaitu adanya kemudahan, serta efektivitas.

Menurut Faisal et al ., (2023) menguraikan bahwa K-NN merupakan salah satu dari algoritma klasifikasi data mining, algoritma ini digunakan untuk mencari atau mendapatkan prediksi dimasa yang akan datang.

Menurut Fajra Riski & Hurriyah Fitriyah, (2022) menguraikan bahwa K-NN adalah algoritma sederhana yang bekerja berdasarkan atas jarak terkecil dari query instance ke sampel latih untuk menentukan tetangga terdekatnya.



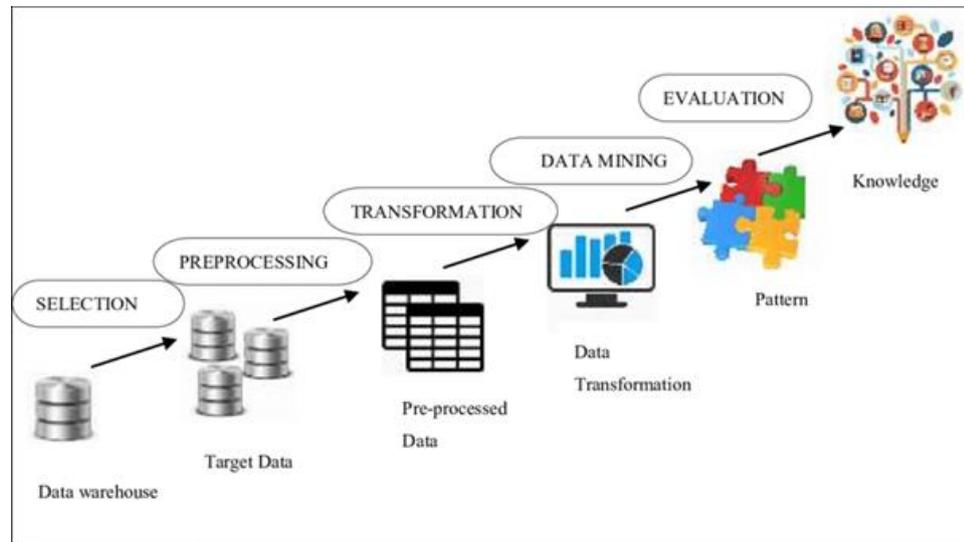
Gambar 2.1 Contoh Algoritma K-NN

6. Kelulusan

Menurut Pratama, (2022) mendefinisikan bahwa kelulusan merupakan keadaan saat berhasil memenuhi persyaratan dan kriteria yang ada sehingga mencapai tujuan dan merubah status. Dalam memberikan penilaian, dosen akan mempertimbangkan faktor-faktor nilai IPK dan batas SKS minimal kelulusan yang mempengaruhi kelulusan. Sebagai contoh dalam wawancara narasumber kepala BAAK Universitas Sahid Surakarta yaitu Firdhaus Hari Saputro Al Haris, ST. M.Eng diketahui batas nilai IPK atau tidak ada mata kuliah yang mengulang bagi mahasiswa Universitas Sahid Sarakarta yaitu 3.00 dan untuk SKS minimalnya yaitu di 144 SKS.

7. Knowledge Discovery in Database (KDD)

Menurut Ghani Muttaqin et al ., (2020) KDD adalah sekumpulan proses untuk menentukan manfaat atau pengetahuan dari sebuah data. Tahapan KDD :



Gambar 2.2 Gambar KDD

- a. Data selection. Proses ini melakukan pemilihan data, menciptakan himpunan data atau memfokuskan pada subset variable (samepl data)
- b. Pre-Processing dan Data cleaning. Proses ini dilakukan membuang data yang tidak konsisen dan noise, duplikasi data, memperbaiki kesalahan kata.
- c. Transformation. Proses ini mentransformasikan atau menggabungkan data ke dalam yang lebih tepat untuk melakukan proses mining dengan cara melakukan peringkasan (agregasi)
- d. Data Mining. Proses ini membuat modeling untuk mencari informasi dalam data terpilih dengan menggunakan teknik, metode, algoritma tertentu sesuai dengan proses KDD..
- e. Evaluation. Proses ini untuk menerjemahkan pola pola hasil dari data mining. Mengevaluasi atau menguji apakah pola atau informasi yang ditemukan bersesuaian atau bertentangan dengan fakta hipotesa sebelumnya. Pengetahuan yang diperoleh berbentuk visualisasi

8. Prediksi Kelulusan

Menurut Al Faruq et al., (2023) mendefinisikan bahwa prediksi kelulusan mahasiswa adalah proses untuk memperkirakan kemungkinan seorang mahasiswa menyelesaikan studinya lulus tepat waktu atau tidak lulus tepat waktu.

9. Data Mining

Menurut Surbakti (2021) mendefinisikan bahwa data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat serta pengetahuan

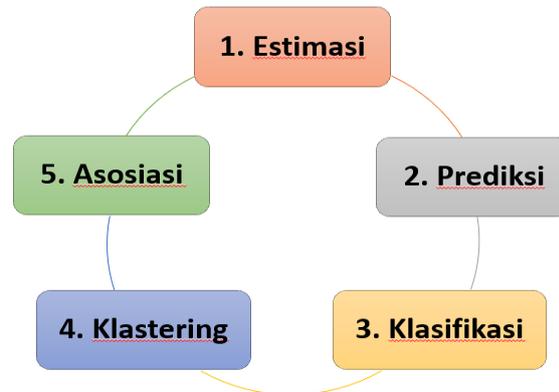
Menurut Ucha Putri et al., (2021) menguraikan bahwa data mining adalah serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data.

Menurut Kristania & Listanto, (2022) menguraikan bahwa data mining adalah analisis data yang menggunakan cara untuk menemukan pola dan aturan dalam sekumpulan data.

Menurut Zulfallah F (2022) menguraikan bahwa data mining adalah menemukan informasi baru, berharga dan penting dalam jumlah data yang besar. Dalam praktiknya, terdapat dua tujuan utama data mining, yaitu prediksi dan deskripsi.

- a. Data mining prediktif adalah untuk menghasilkan model sistem yang dijelaskan oleh sekumpulan data yang diberikan.
- b. Data mining deskriptif adalah menghasilkan informasi baru dan penting berdasarkan kumpulan data yang tersedia.

Data mining dapat diartikan juga sebagai proses menemukan pola dan pengetahuan yang menarik dari jumlah data yang besar maupun kecil. Sumber data yang didapat dari database, gudang data, web dan penyimpanan informasi lainnya. Adapun peran dari data mining :



Gambar 2.3 Peran data mining

1) Estimasi

Estimasi adalah proses dari satu set variable numerik dan/atau kategori digunakan untuk mendekati variable target. Contoh estimasi adalah mengestimasi biaya pengeluaran Universitas untuk periode 1 semester. Algoritma yang digunakan adalah Neural Network

2) Prediksi

Prediksi hampir sama dengan estimasi maupun klasifikasi, perbedaannya pada hasilnya yang ada dimasa depan, misalnya memprediksi harga crypto, saham dan emas. Contoh algoritma yang digunakan adalah K- Nearest Neighbors, Decision tree dan Neural Network

3) Klasifikasi

Klasifikasi hampir sama dengan estimasi dan prediksi, dengan perbedaan pada variable objek bersifat kategoris dari pada numerik. Ada variable tujuan dalam klasifikasi, contohnya pada nilai mahasiswa, yang dapat dibagi menjadi 3 kelompok atau kategori : nilai A (tinggi), nilai B (sedang), dan nilai C (rendah). Contoh algoritmanya adalah K- Nearest Neighbors, Decision Tree.

4) Klustering

Klustering adalah pengelompokan catatan, pengamatan, kejadian dalam kelas yang terkait. Pengelompokan record yang dapat dibandingkan satu sama lain tetapi tidak dengan record di kluster lain. Perbedaan dengan klasifikasi, klustering tidak memiliki variable target. Contoh dari klustering yaitu target pemasaran produk tertentu untuk Perusahaan menengah dengan anggaran pemasaran yang terbatas. Contoh algoritma yang digunakan yaitu K-Means.

5) Asosiasi

Asosiasi adalah proses menentukan kualitas mana yang "berjalan bersama". Proses menemukan kriteria untuk mengukur hubungan antara 2 atau lebih kualitas. Contoh asosiasi adalah jika 100 pelanggan membeli roti, jika 80 dari 100 pelanggan membeli roti dengan selai maka aturan asosisanya adalah "jika membeli roti, maka belilah selai". Algoritma yang dapat digunakan algoritma apriori.

10. Confusion Matrix

Confusion matrix adalah matriks yang digunakan untuk menyatakan klasifikasi jumlah data uji yang benar dan jumlah data uji yang salah Normawati & Prayogi, (2021). *Confusion matrix* juga digunakan untuk menghitung kinerja terkait dengan kebenaran di dalam suatu proses klasifikasi Ahmad et al., (2024) *Confusion matrix* untuk klasifikasi biner ditunjukkan pada tabel 2.1

Tabel 2.1 *Confusion Matrix*

Nilai Aktual	Nilai Prediksi	
	<i>Lulus Tepat Waktu</i>	<i>Tidak Lulus Tepat Waktu</i>
<i>Lulus Tepat Waktu</i>	<i>True Positive</i>	<i>False Positive</i>
<i>Tidak Lulus Tepat Waktu</i>	<i>False Negatif</i>	<i>True Negatif</i>

Keterangan :

1. *TP (True Positif)* = jumlah data nilai aktual kelas lulus tepat waktu dan nilai prediksi kelas lulus tepat waktu
2. *TN (True Negative)* = jumlah data nilai aktual kelas tidak lulus tepat waktu dan nilai prediksi tidak lulus tepat waktu
3. *FP (False Positive)* = jumlah data nilai aktual lulus tepat waktu dan nilai prediksi tidak lulus tepat waktu
4. *FN (False Negative)* = jumlah data nilai aktual tidak lulus tepat waktu dan nilai prediksi lulus tepat waktu

Rumus *Confusion matrix* untuk menghitung *Accuracy*, *Precision* dan *recall* sebagai berikut :

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{Total}} \dots \dots \dots (\text{Persamaan 2})$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \dots \dots \dots (\text{Persamaan 3})$$

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \dots \dots \dots (\text{Persamaan 4})$$