

BAB III

METODOLOGI PENELITIAN

3.1 Metode Pengumpulan Data

Terdapat dua cara yang digunakan untuk mengumpulkan data yakni observasi dan studi pustaka.

3.1.1 Wawancara

Wawancara merupakan salah satu usaha secara sistematis, untuk mengumpulkan informasi yang kita butuhkan. Wawancara dilakukan dengan cara bertanya atau melakukan tanya jawab secara langsung dengan pihak bagian akademik Universitas Nahdlatul Ulama Lampung. Tujuan dari wawancara adalah untuk memperoleh data yang lebih akurat dan lengkap sebagai kebutuhan penelitian.

3.1.2 Studi Pustaka

Metode ini dilakukan dengan cara mempelajari dan mencari referensi pada jurnal maupun literature lain yang berhubungan dengan penelitian yang dilakukan yaitu Prediksi Kelulusan Mahasiswa menggunakan Algoritma *K-Nearest Neighbor*(K-NN) dan *Naive Bayes*. Metode ini digunakan untuk mencari sumber – sumber data yang diperlukan dalam penelitian yang biasa diperoleh dari membaca dan mengutip buku – buku referensi serta jurnal penelitian yang pernah dilakukan sebelumnya yang mendukung dan berkaitan dengan penelitian.

3.2 Penentuan Metode

Dalam penelitian ini penentuan metode dilakukan dengan cara mempelajari studi literatur yang digunakan pada Bab II. Dimana pada studi literatur tersebut terdapat banyak metode yang digunakan , di antaranya Naive Bayes, K-Nearest Neighbors, Decision Tree, dan Artificial Neural Network. Metode Naive Bayes dan K-Nearest Neighbors

dipilih untuk melakukan penelitian ini karena termasuk metode klasifikasi yang memiliki tingkat akurasi cukup tinggi.

3.3 Penentuan Variabel

Dalam penelitian ini variabel yang digunakan diantaranya adalah NIM, jenis kelamin, usia, jurusan, IP Semester 1-4, IPK, lama masa studi dan nilai ijazah SMA. Adapun variabel-variabel dari data kelulusan mahasiswa tersebut dapat dilihat pada tabel 3.1.

Tabel 3.1 Variabel yang Digunakan

Variabel	Keterangan
NIM	Digunakan sebagai identitas mahasiswa
Jenis Kelamin	Digunakan untuk mengetahui jenis kelamin mahasiswa Perempuan/Laki-laki
Usia	Digunakan untuk mengetahui usia mahasiswa saat lulus
Jurusan	Digunakan untuk mengetahui jurusan mahasiswa
IP Semester 1-4	Digunakan untuk mengetahui perkembangan Indeks Prestasi mahasiswa
IPK	Digunakan untuk mengetahui Indeks Prestasi Kumulatif mahasiswa saat lulus
Masa Studi	Digunakan untuk mengetahui lama waktu yang ditempuh mahasiswa untuk menyelesaikan studinya
Ijazah SMA	Digunakan untuk mengetahui nilai ijazah SMA yang digunakan mahasiswa saat mendaftar

3.4 Tools yang Digunakan

3.4.1 Rapidminer

Rapidminer merupakan software untuk analisis data dan sebagai mesin data mining yang menyediakan lingkungan terintegrasi untuk persiapan data, pembelajaran mesin (machine learning), pembelajaran dalam (deep learning), penambangan teks (text mining), dan analisis prediktif (predictive analytics). RapidMiner merupakan salah satu tools yang dipakai dalam data mining. RapidMiner memiliki kurang lebih 500 operator data mining, termasuk operator untuk input, output, data preprocessing dan visualisasi. RapidMiner merupakan software yang berdiri sendiri untuk analisis data dan sebagai mesin data mining yang dapat diintegrasikan pada produknya sendiri.

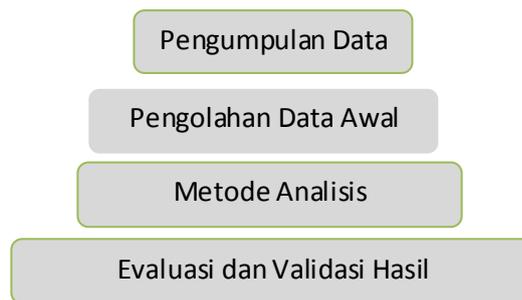
3.4.1.1 Fitur-Fitur yang Tersedia dalam RapidMiner

Dalam menangani pekerjaan yang berhubungan dengan data mining, RapidMiner menyediakan berbagai fitur sebagai berikut:

- Banyak algoritma penambangan data seperti pohon keputusan dan peta yang mengatur sendiri.
- Visualisasi yang canggih seperti diagram histogram yang tumpang tindih, diagram pohon, dan diagram sebar 3D.
- Tersedia banyak pilihan plugin termasuk juga plugin teks yang bisa dimanfaatkan untuk melakukan analisis text mining.
- Menyediakan teknologi data mining dan machine learning termasuk ETL (Extract, Transform, Load), pra-pemrosesan data, visualisasi data, pemodelan dan evaluasi.
- Proses penambangan data ditulis dalam XML dan terdiri dari operator nestable, dideskripsikan dengan XML, dan dibuat dalam GUI.

3.5 Tahapan Penelitian

Dalam penelitian ini, data yang digunakan adalah data kelulusan mahasiswa Universitas Nahdatul Ulama. Data kelulusan mahasiswa tersebut akan diolah menggunakan metode data mining sehingga hasilnya dapat digunakan sebagai rules dalam memprediksi kelulusan mahasiswa. Dalam penelitian ini akan dilakukan beberapa langkah-langkah atau tahapan penelitian seperti yang digambarkan pada gambar 3.1.



Gambar 3.1 Tahapan Penelitian

3.5.1 Pengumpulan Data

Dalam pengumpulan data, sumber data dihimpun langsung oleh peneliti melalui Kepala Bagian BAK Universitas Nahdatul Ulama Lampung. Peneliti mengumpulkan data mahasiswa yang lulus pada tahun 2022 sebagai kasus lama dan data mahasiswa yang lulus pada 2023 sebagai kasus baru.

3.5.2 Pengolahan Data Awal

Untuk mendapatkan data yang berkualitas, beberapa teknik yang dilakukan oleh peneliti adalah sebagai berikut

1. Data validation, untuk mengidentifikasi dan menghapus data yang ganjil (outlier/noise), data yang tidak konsisten, dan data yang tidak lengkap (missing value).
2. Data integration and transformation, untuk meningkatkan akurasi dan efisiensi algoritma. Data yang digunakan dalam penulisan ini bernilai kategorikal.

3. Data size reduction and dicrtization, untuk memperoleh data set dengan jumlah atribut dan record yang lebih sedikit tetapi bersifat informatif.

3.5.3 Metode Analisis

Dalam penelitian ini akan dilakukan eksperimen dengan menggunakan metode klasifikasi data mining *K-Nearest Neighbor* dan *Naive Bayes* terhadap data mahasiswa yang terkait dengan kelulusan mahasiswa.

3.5.3.1 Tahapan Kerja Metode K-Nearest Neighbor

1. Menentukan jumlah tetangga (K) yang akan digunakan untuk pertimbangan penentuan kelas.
2. Menghitung jarak dari data baru ke masing-masing data point di dataset.
3. Ambil sejumlah K data dengan jarak terdekat, kemudian tentukan kelas.

3.5.3.2 Tahapan Kerja Metode Naive Bayes

1. Mengidentifikasi kelas atau label yang ingin diprediksi.
2. Menghitung berapa banyak kasus yang terdapat dalam setiap kelas
3. Menghitung probabilitas untuk setiap variabel dalam kelas. Probabilitas ini akan memberi tahu seberapa sering suatu variabel muncul dalam suatu kelas tertentu.
4. Setelah memiliki probabilitas variabel, peneliti menggunakan Teorema Naive Bayes untuk melakukan prediksi. Teorema Naive Bayes membantu menghitung peluang terjadinya suatu kelas berdasarkan bukti yang dimiliki (probabilitas variabel kelas).

3.5.4 Evaluasi dan Validasi Hasil

Evaluasi adalah langkah pemeriksaan untuk memastikan bahwa data tersebut telah sesuai kriteria yang ditetapkan dengan tujuan untuk memastikan bahwa data yang akan dimasukkan ke dalam basis data telah diketahui dan dapat dijelaskan sumber dan kebenaran datanya. Evaluasi data mining ini bisa digunakan untuk komparasi atau perbandingan antara algoritma yang digunakan.

Validasi adalah proses menilai seberapa baik performa model penambangan Anda terhadap data nyata. Penting bagi Anda untuk memvalidasi model penambangan dengan memahami kualitas dan karakteristiknya sebelum Anda menyebarkannya ke lingkungan produksi.