

## **BAB II**

### **METODE PENELITIAN**

#### **2.1 Objek Penelitian**

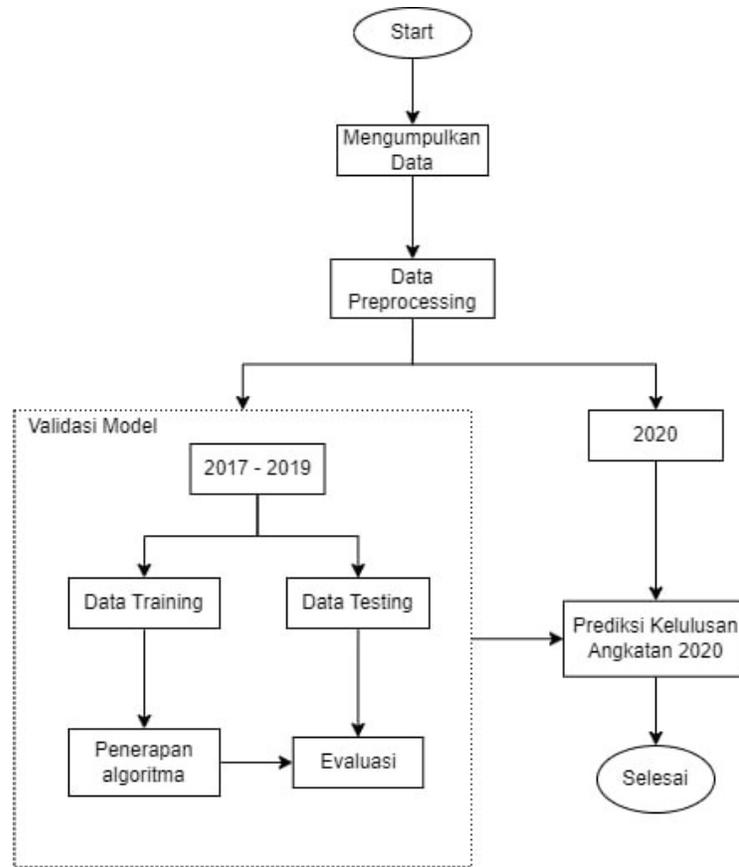
Objek pada penelitian ini yaitu memprediksi kelulusan mahasiswa Program Studi Teknik Informatika Universitas Muhammadiyah Kalimantan Timur.

#### **2.2 Alat dan Bahan**

- 1) Peralatan yang digunakan untuk penelitian ini adalah:
  - a. Satu laptop dengan spesifikasi AMD Ryzen 5 4500U dengan Radeon Graphics 2.38 GHz, dan RAM 8 GB.
  - b. Tools untuk menjalankan algoritma K-NN menggunakan bahasa pemrograman *python*.
  - c. Menggunakan *RapidMiner studio* untuk menjalankan bahasa *python*.
  
- 2) Bahan yang digunakan untuk penelitian ini adalah
  - a. Data kelulusan mahasiswa angkatan 2017 – 2019
  - b. Data penerimaan mahasiswa baru angkatan 2017 – 2020
  - c. Data masa studi berupa nilai matakuliah, Indeks Prestasi Semester (IPS), dan Indeks Prestasi Kumulatif (IPK).

#### **2.3 Prosedur penelitian**

Tujuan utama penelitian ini ialah melakukan prediksi terhadap kelulusan mahasiswa dengan menggunakan algoritma K-NN dan mengetahui akurasi dari algoritma ini pada data mahasiswa Teknik Informatika Universitas Muhammadiyah Kalimantan Timur. Untuk mencapai tujuan tersebut, maka penelitian ini membuat tahapan penelitian seperti ditunjukkan pada **Gambar 2.1**.



**Gambar 2.1** Desain Penelitian

Penelitian dimulai dengan mengumpulkan data akademik mahasiswa, kemudian melakukan *data preprocessing* sehingga data siap untuk dimasukkan sebagai inputan algoritma K-NN, selanjutnya model dievaluasi berdasarkan akurasi. Setelah itu model diterapkan untuk memprediksi kelulusan mahasiswa angkatan 2020.

### 2.3.1 Pengumpulan Data

Dalam pengumpulan data untuk penelitian ini, penulis melakukan observasi langsung dengan cara mengambil data kelulusan dan penerimaan mahasiswa baru di Universitas Muhammadiyah Kalimantan Timur tepatnya pada bagian Program Studi Teknik Informatika. Data yang dikumpulkan berupa data jumlah kelulusan mahasiswa, jumlah penerimaan mahasiswa, nilai Indeks Prestasi Semester (IPS), nilai Satuan Kredit Semester (SKS). Pengguna IPS dan SKS untuk memprediksi kelulusan mahasiswa telah digunakan pada

penelitian terdahulu seperti yang telah dilakukan oleh Anwar et al (2021) dengan akurasi yang mencapai 90,6%.

### **2.3.2 Data Preprocessing**

Setelah pengumpulan data, maka tahap selanjutnya adalah *data preprocessing*. *Data preprocessing* merupakan suatu proses yang mengolah data mentah dengan menghapus atribut yang tidak terpakai sehingga menjadi lebih berkualitas (Pane & Ramdan, 2022). Tujuan dari *data preprocessing* ini adalah untuk menyempurnakan data sehingga model yang dihasilkan memiliki akurasi yang tinggi (Putra et al., 2024). *Preprocessing* yang dilakukan ialah memisahkan, pembersihan dan menormalisasikan data sesuai dengan metode yang dipakai agar siap digunakan untuk analisis dan permodelan. Berdasarkan penelitian yang dilakukan oleh Amri et al. (2023) yang melakukan penelitian terhadap kelulusan mahasiswa, *data preprocessing* akan dilakukan jika terdapat data yang sama atau data yang bersifat *outlier*. Dan juga melakukan penghapusan terhadap atribut yang tidak digunakan (Aprilio Arifin & Malago, 2023).

### **2.3.3 Pembagian Data**

Dari *preprocessing* data di atas maka di peroleh hasil yaitu penyesuaian dan pemisahan data menjadi data validasi model yaitu angkatan 2017 – 2019 dan data prediksi yaitu pada angkatan 2020. Yang mana data validasi model digunakan sebagai pengujian terhadap algoritma yang dipakai untuk menentukan akurasi. Selanjutnya data prediksi digunakan untuk memprediksi kelulusan yang akan datang.

### **2.3.4 Split Data**

Selanjutnya untuk pembuatan model K-NN dan evaluasinya, maka data mahasiswa angkatan 2017 - 2019 dibagi menjadi *data training* dan *data testing* menggunakan *split data*. *Split data* adalah suatu proses membagi dataset menjadi dua yaitu *data training* dan *data testing* 90% :10%, 85%:15%, 83%:17%, 80%:20%, 75%:25%, 73%:27%, 70%:30%, 60%:40%,

50%:50 tergantung tingginya tingkat akurasi yang dihasilkan (Baiq Nurul Azmi et al., 2023). Yang dimana *data training* merupakan sebagian dari dataset yang dimanfaatkan untuk melatih model dalam pembelajaran mesin pada algoritma yang dipakai. Data ini digunakan untuk mengajarkan model pola dan keterkaitan dalam data dengan tujuan membuat prediksi. Sementara itu, *data testing* merupakan sebagian lain dari dataset yang tidak digunakan selama pelatihan model. Pada penelitian oleh Qoiriyah et al (2021) menggunakan perbandingan 75% : 25%, yang menghasilkan rata – rata akurasi 85%. Untuk itu pembagian data yang dilakukan pada penelitian ini adalah 75% : 25%.

### 2.3.5 Klasifikasi K-NN

Setelah membagi data di atas, maka tahap selanjutnya adalah menjalankan *data training* pada algoritma K-NN. Yang dimana, konsep penghitungan algoritma K-NN adalah untuk mengelompokkan objek yang baru berdasarkan (K) tetangga terdekatnya. K-NN termasuk dalam *supervised learning*, yang berarti klasifikasi *instance* ditentukan berdasarkan mayoritas kategori pada K tetangga terdekat pada data. Kelas yang paling sering muncul di antara tetangga maka akan menentukan kelas hasil klasifikasi (Dodi Guswandi et al., 2021). Langkah-langkah untuk menghitung metode K-Nearest Neighbor antara lain (Mutiara Ayu Banjarsari, H. Irwan Budiman, 2015):

1. Menentukan parameter
2. Menghitung jarak antara data yang akan dievaluasi dengan semua pelatihan
3. Mengurutkan jarak yang terbentuk
4. Menentukan jarak terdekat sampai urutan K
5. Memasangkan kelas yang bersesuaian
6. Mencari jumlah kelas dari tetangga yang terdekat dan tetapkan kelas tersebut sebagai kelas data yang akan dievaluasi

Penghitungan jarak algoritma KNN dilakukan menggunakan rumus *Euclidean Distance* dibawah ini (Informatika & Dahlan, 2020):

$$d = \sqrt{\sum_{i=1}^p (x_{2i} - x_{1i})^2} \quad (2.1)$$

Keterangan :

$d$  = Jarak antar objek

$x_1$  = Sampel data

$x_2$  = Data testing

$p$  = Dimensi data

$i$  = Variabel data

### 2.3.6 Evaluation

Tahap *evaluation* digunakan untuk memberikan informasi tentang seberapa baik model dalam mempelajari pola yang ada dalam *data training* dan *data testing* menggunakan akurasi, presisi, dan recall. Pada penelitian ini, pengukuran kinerja atau evaluasi model dilakukan dengan cara menghitung nilai akurasi, presisi dan recall pada *data testing*. Berikut ini adalah rumus menentukan akurasi, presisi dan recall (Muslim et al., 2019, p. 48)

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (2.2)$$

$$Presisi = \frac{TP}{TP + FP} \quad (2.3)$$

$$Recall = \frac{TP}{TP + FN} \quad (2.4)$$

Keterangan :

1. TP (*True Positif*) = *Observed Class* benar dengan hasil *predicted class* benar.
2. TN (*True Negatif*) = *Observed Class* benar dengan hasil *predicted class* salah.
3. FP (*False Positif*) = *Observed Class* salah dengan hasil *predicted class* benar.
4. FN (*False Negatif*) = *Observed Class* salah dengan hasil *predicted class* salah.

## **2.4 Prediksi Kelulusan**

Setelah berbagai tahapan di atas, maka selanjutnya menerapkan model K-NN untuk memprediksi kelulusan mahasiswa angkatan 2020. Mahasiswa Prodi Teknik Informatika adalah mahasiswa tahun terakhir (semester 8) yang telah menempuh seluruh mata kuliah dan seharusnya akan lulus pada akhir semester 8 ini.