

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pendidikan merupakan hal yang terpenting dalam kehidupan setiap manusia. Adanya suatu pendidikan sebagai usaha untuk menyiapkan manusia melalui berbagai kegiatan yang berpengaruh terhadap sikap dan tingkah laku untuk meningkatkan kualitas kehidupan. Mahasiswa sering disebut kelompok masyarakat yang memiliki ciri intelektualitas yang lebih luas dibandingkan dengan kelompok seusia mereka yang bukan mahasiswa ataupun kelompok usia lain yang dibawah mereka (Rohman, 2015). Keberhasilan sistem pendidikan khususnya perguruan tinggi dapat dilihat dari meningkatnya kualitas pendidikan dari tahun ketahun. Salah satu yang menjadi indikator kualitas pendidikan adalah tingkat kelulusan mahasiswa dan banyaknya lulusan yang bekerja di suatu lapangan pekerjaan.

Kelulusan mahasiswa merupakan faktor untuk meningkatkan kualitas perguruan tinggi. Namun kelulusan mahasiswa tidak selalu dapat diprediksi, sehingga mengakibatkan keterlambatan kelulusan. Oleh karena itu untuk mengatasinya perlu dilakukan suatu teknik untuk bisa melakukan prediksi terhadap kelulusan mahasiswa. Walaupun prediksi juga tidak harus memberikan jawaban secara pasti kejadian yang akan terjadi, melainkan berusaha untuk mencari jawaban sedekat mungkin yang akan terjadi (Rohmawati et al., 2017). Adapun teknik yang digunakan yaitu dengan menggunakan teknik *data mining*.

Data mining merupakan analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga dan meringkas data dengan cara yang berbeda dengan sebelumnya yang dapat dipahami dan bermanfaat bagi pemilik data (Widaningsih, 2019). *Data mining* digunakan untuk proses mendapatkan pengetahuan atau pola dari basis data, yang nantinya pola yang diperoleh dapat dijadikan solusi pengambil keputusan. Dalam penelitian ini akan memanfaatkan *data mining* untuk memprediksi kelulusan mahasiswa dengan mengkomparasi 2 algoritma klasifikasi yaitu *K Nearest Neighbor* dan *Neural Network*.

Penelitian dalam lingkup *data mining* yang menggunakan metode *K Nearest Neighbor* dan *Neural Network* untuk prediksi kelulusan mahasiswa telah banyak dilakukan, sebagaimana yang pernah dilakukan oleh (Rohman, 2015) menerapkan perhitungan algoritma *K-Nearest Neighbor* untuk prediksi kelulusan mahasiswa. Hasil akurasi prediksi menggunakan algoritma *K-Nearest Neighbor* sebesar 85,15%. Pada penelitian ini data dikluster sebanyak 5, yaitu data $k=1$, $k=2$, $k=3$, $k=4$, dan $k=5$. Hasil yang diperoleh dengan cluster data $k=5$ accuracy adalah 85,15% dan nilai AUC adalah 0.888.

Kemudian penelitian oleh (Kartarina et al., 2021) membuat sistem untuk memprediksi kelulusan mahasiswa dengan membandingkan metode *Naïve Bayes* dan *K-Nearest Neighbor*. Hasil perhitungan akurasi menggunakan *Rapidminer*, metode *Naïve Bayes* sebesar 91,94 dan *K-Nearest Neighbor* sebesar 96,18%. Kemudian hasil perhitungan sistem menggunakan metode *K-Nearest Neighbor* dengan nilai $K=7$ didapatkan hasil keterangan “TEPAT WAKTU”, dan dari hasil perhitungan *Naïve Bayes*, didapatkan hasil keterangan “TIDAK TEPAT WAKTU”.

Begitu pula hasil dari penelitian yang dilakukan (Rohman & Rochcham, 2019) dengan mengkomparasi 3 algoritma yaitu *Decision Tree*, *K-Nearest Neighbor* dan *Neural Network*. Pada penelitian ini menggunakan *Confusion Matrix* dan kurva *ROC* untuk mengukur kinerja algoritma. Jumlah nilai k pada pengujian ini sebanyak 10 kali. Hasil yang didapatkan yaitu *Neural Network* memiliki nilai akurasi yang paling tinggi yakni sebesar 87.32% dengan nilai *AUC* 0,917 diikuti *Decision Tree* sebesar 83.57% nilai *AUC* 0,844 dan *K-Nearest Neighbor* sebesar 83.66% nilai *AUC* 0,842.

Adanya perbedaan nilai akurasi dari penelitian diatas ini disebabkan oleh pemilihan jumlah *dataset*, pembagian jumlah *data training* dan *data testing*, parameter serta cara perhitungannya. Jumlah nilai k sangat berpengaruh dalam pengujian, karena digunakan untuk memperkirakan akurasi estimasi. Beberapa kelemahan yang ada pada penelitian sebelumnya yaitu jumlah *data training*. Dalam proses prediksi harus menggunakan *data training* dalam jumlah banyak karena *data training* berpengaruh pada akurasi prediksi, terutama pada metode *K-Nearest Neighbor* yang akan efektif apabila *data training*-nya besar.

Berangkat dari paparan latar belakang tersebut, maka diusulkan sebuah penelitian dengan judul “Komparasi Metode *K-Nearest Neighbor* dan *Neural Network* Untuk Menentukan Tingkat Akurasi Dalam Prediksi Kelulusan Mahasiswa Studi kasus Program Studi Informatika Universitas Siliwangi.”

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, maka dapat ditarik rumusan masalah sebagai berikut:

1. Bagaimana mengukur tingkat akurasi prediksi kelulusan mahasiswa menggunakan metode *K-Nearest Neighbor* dan *Neural Network*?
2. Seberapa akurat metode *K-Nearest Neighbor* dan *Neural Network* dalam memprediksi kelulusan mahasiswa bila dibandingkan dengan metode sejenis lainnya ?

1.3 Batasan Masalah

Batasan batasan masalah dalam penulisan tugas akhir ini adalah sebagai berikut:

1. Metode yang digunakan dalam penelitian ini, yaitu metode *K-Nearest Neighbor* dan *Neural Network*.
2. *Software* yang dipakai didalam penelitian tentang *data mining* ini menggunakan *rapidminer* dan *Google colab* dengan menggunakan bahasa pemrograman *python*..
3. Akan dilakukan komparasi terhadap dua metode untuk menghasilkan nilai akurasi yang tertinggi.
4. Dalam penelitian ini metode *backpropagation* digunakan untuk melatih jaringan *Neural Network*.
5. Perbandingan pembagian data 90% data *training* dan 10% data *testing*.
6. Penelitian tentang prediksi kelulusan mahasiswa ini hanya menggunakan data kelulusan mahasiswa prodi informatika angkatan tahun 2014.

1.4 Tujuan Penelitian

Tujuan penelitian tugas akhir ini adalah sebagai berikut:

1. Menerapkan metode *K-Nearest Neighbor* dan *Neural Network* untuk mengukur tingkat akurasi prediksi kelulusan mahasiswa.

2. Mengetahui metode manakah yang mempunyai nilai akurasi tinggi dan paling akurat dalam memprediksi kelulusan mahasiswa program studi informatika dibandingkan dengan metode sejenis lainnya?

1.5 Manfaat Penelitian

Manfaat dari penelitian tugas akhir ini adalah sebagai berikut :

1. Mengetahui keakuratan metode yang digunakan dalam prediksi kelulusan mahasiswa.
2. Sebagai bahan evaluasi dasar untuk penelitian selanjutnya.
3. Memberikan kontribusi keilmuan pada penelitian bidang klasifikasi data mining khususnya untuk prediksi kelulusan mahasiswa.
4. Membantu perguruan tinggi dalam menciptakan kebijakan supaya dapat meningkatkan kelulusan mahasiswa.

1.6 Metodologi Penelitian

Langkah – langkah metodologi penelitian yang dilakukan dalam penelitian ini adalah sebagai berikut :

1. Studi Literatur

Metode ini dilakukan dengan mempelajari dan mengkaji teori-teori melalui buku, jurnal ilmiah, laporan penelitian serta dokumen-dokumen yang berkaitan dengan penelitian ini.

2. Pemodelan kebutuhan

a. Pengumpulan data untuk penelitian

Melakukan permintaan data mahasiswa informatika Angkatan 2014 ke Unit Pelaksana Teknis Teknologi Informasi dan Komunikasi (UPT TIK) Universitas Siliwangi dengan melalui tahap proses perizinan.

b. Pengumpulan kebutuhan perangkat untuk penelitian

Perangkat penelitian disesuaikan dengan analisis kebutuhan perangkat. Pengumpulan kebutuhan untuk penelitian mencakup laptop yang mendukung, serta penggunaan *software Rapidminer* dan bahasa pemrograman *python* untuk proses analisis data.

3. Pengolahan dan pengujian data

Pengolahan dan pengujian data yang dilakukan dalam penelitian ini mengikuti tahapan dalam *Knowledge Discovery in Database (KDD)*, yaitu *data selection, preprocessing, transformation, data mining* dan *interpretation*. Tahap ini data diuji menggunakan metode *K-Nearest Neighbor* dan *Neural Network*. Kemudian *tools* yang digunakan untuk membantu dalam pengujian data yaitu menggunakan *software Rapidminer* dan *Google colab* dengan menggunakan bahasa pemrograman *python*. Dilakukan juga pengujian menggunakan *k-fold cross validation*.

4. Evaluasi hasil pengujian

Evaluasi hasil pengujian algoritma dilakukan untuk menilai kinerja dan efektivitas suatu algoritma yang digunakan. Evaluasi hasil pengujian membantu mengukur sejauh mana algoritma mampu menghasilkan hasil yang akurat dan relevan dalam konteks masalah yang sedang diselesaikan. Evaluasi hasil pengujian meliputi *accuracy, precision, recall, F1-Score, confusion matrix* dan kurva ROC (*Receiver Operating Characteristic*).

1.7 Sistematika Penulisan

Sistematika penulisan yang digunakan dalam laporan tugas akhir ini adalah sebagai berikut:

BAB I PENDAHULUAN

Bab ini menguraikan tentang gambaran umum dari penelitian. Terdiri dari latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, metodologi penelitian dan sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini berisi penelitian – penelitian terkait yang pernah dilakukan dalam penelitian sebelumnya dan hubungannya dengan masalah penelitian yang sedang dilakukan ini. Kemudian memuat teori – teori yang berhubungan dengan penelitian seperti tentang dataset, metode *K – Nearest Neighbour*, *Neural Network* dan penjelasan lainnya yang berhubungan dengan penelitian.

BAB III METODOLOGI

Bab ini berisi tentang metode yang digunakan dalam pembahasan serta langkah-langkah dalam penyelesaian masalah selama melakukan penelitian. Langkah-langkah tersebut memuat konsep dari metode yang digunakan, analisis kebutuhan, serta uraian-uraian lainnya yang berkaitan dengan penelitian yang akan dibahas.

BAB IV HASIL DAN PEMBAHASAN

Bab ini memuat uraian tentang hasil proses analisis yang telah dilakukan serta pembahasan yang meliputi penjelasan mengenai data yang akan dianalisis.

BAB V KESIMPULAN DAN SARAN

Bab ini berisi tentang kesimpulan yang merupakan rangkuman dari pembahasan masalah pada penelitian dan saran yang perlu diperhatikan berdasarkan keterbatasan-keterbatasan yang ada.