

BAB I

PENDAHULUAN

1.1 Latar Belakang

Akreditasi merupakan salah satu bentuk penilaian mutu dan kelayakan program studi di perguruan tinggi. Ketepatan waktu mahasiswa dalam menyelesaikan studi dan proporsi mahasiswa yang menyelesaikan studi dalam batas masa studi termasuk dalam elemen penilaian akreditasi. Selain itu, ketepatan lama studi mahasiswa merupakan isu yang penting karena ketepatan tersebut menjadi dasar efektifnya suatu perguruan tinggi (Untari, 2014). Mahasiswa dikatakan tepat waktu jika menempuh masa studi sarjana/S1 maksimal selama 4 tahun atau kurang (PERMENDIKNAS No. 232, 2000).

Setiap perguruan tinggi berusaha untuk terus memperbaiki menejemennya untuk meningkatkan mutu pendidikan dan meningkatkan akreditasi. Salah satu elemen penilaian akreditasi perguruan tinggi adalah lulus tepat waktu. Kurangnya informasi dan analisa yang diperoleh Bidang Akademik mengakibatkan sulitnya melakukan klasifikasi lama studi mahasiswa. Klasifikasi lama studi mahasiswa dapat membantu

Bidang Akademik dalam menyusun strategi yang tepat untuk menekan atau memperpendek lama studi mahasiswa. Oleh karena itu perlu dilakukan suatu analisis klasifikasi dalam memprediksi bahwa seorang mahasiswa dikatakan lulus tepat waktu atau tidak berdasarkan data yang diperoleh dari mahasiswa itu sendiri. Klasifikasi adalah metode untuk memprediksi suatu kejadian atau keputusan yang akan datang berada di suatu titik. Klasifikasi merupakan suatu pekerjaan yang melakukan penilaian terhadap suatu objek data untuk masuk dalam suatu kelas tertentu dari sejumlah kelas yang tersedia (Prasetyo, 2012).

Metode KNN dan Random Forest merupakan bagian dari metode klasifikasi. Menurut Han dan Kember (2006) algoritma *K-NN* merupakan salah satu algoritma yang termasuk dalam *supervised*. Konsep dasar dalam algoritma *K-NN* yakni dengan mencari jarak yang paling terdekat diantara data yang terevaluasi dengan sejumlah *K* (*Neighbor*) paling dekat dengan data *test* (Sonatha, 2013). *K-NN* digunakan karena lebih sederhana, efektif dan juga dapat diaplikasikan pada jumlah *training data* yang sedikit dan mudah dalam mengelompokkan kemudian mudah dimengerti (Hardjianto, M., & Winarko, E. 2013). *Random Forest* adalah algoritma yang dapat digunakan dalam masalah klasifikasi pada *machine learning* dan *data mining* (Larose, 2013). Dalam hal ini memetakan dari kelas sehingga bisa digunakan dalam menemukan prediksi terhadap data belum muncul. Pada pendekatan pohon keputusan yaitu "*divide and conquer*" dalam mempelajari masalah dari kumpulan data independen yang digambar pada bagan pohon (Witten, dkk 2011).

Adapun penelitian menggunakan algoritma *K-NN* diantaranya dilakukan oleh Badu (2016) yang mengklasifikasikan dana desa dengan tingkat akurasi 78,95% dengan nilai $K = 2$. Pada penelitian Subrata, dkk (2017) *K-NN* digunakan untuk mengklasifikasikan penggunaan protokol komunikasi pada trafik jaringan dengan tingkat akurasi 99,14 %. Selanjutnya, penelitian tentang *Random forest* diantaranya dilakukan oleh Ratnawati dan Sulistyaningrum (2019) yang menerapkan *Random Forest* untuk mengukur tingkat keparahan penyakit pada daun apel. Dalam hal ini diperoleh hasil akurasi sebesar 75,3191%. Lebih lanjut Hanun dan Zailani (2020) menerapkan klasifikasi *Random Forest* dalam menentukan kelayakan pemberian kredit. Penelitian tersebut menganalisis debitur yang bermasalah dan debitur tidak bermasalah dengan tingkat akurasi sebesar 87,88%. Dari hasil penelitian yang telah disebutkan menunjukkan tingkat akurasi yang baik dari algoritma *K-NN* maupun *Random Forest*. Namun sejauh ini, belum dapat dikonfirmasi secara pasti, mana yang terbaik diantara kedua metode tersebut.

Pada penelitian ini, Algoritma *K-NN* dan *Random Forest* digunakan untuk mengklasifikasikan lama studi mahasiswa, namun digunakan *Information Gain* untuk menyeleksi fitur-fitur yang tidak memiliki pengaruh. Hal ini sesuai dengan penelitian yang dilakukan Bimantoro dan Uyun (2017) yang menggunakan *Information Gain* untuk seleksi fitur citra dalam menilai kesesuaian lahan pada tanaman cengkeh. Tingkat akurasi penggunaan fitur tanpa proses seleksi hanya 50%, sedangkan fitur terpilih dari hasil seleksi menggunakan *Information Gain* dengan nilai *threshold* 0,7 naik menjadi 88%. Selanjutnya, untuk melihat evaluasi dari sebuah model yang dibangun akan digunakan *Confusion Matrix* dengan tabel matriks.

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, maka rumusan masalah yang diangkat dalam penelitian kali ini yaitu sebagai berikut :

1. Bagaimana menentukan atribut terbaik dengan menggunakan metode seleksi fitur *Information gain* ?
2. Bagaimana akurasi pengklasifikasian pada metode *K-NN* untuk lama studi mahasiswa dengan menggunakan atribut terbaik dari metode seleksi fitur *Information Gain* ?
3. Bagaimana akurasi pengklasifikasian pada metode *random forest* untuk lama studi mahasiswa dengan menggunakan atribut terbaik dari metode seleksi fitur *Information Gain* ?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah diatas, maka tujuan yang ingin dicapai dalam penelitian ini yaitu sebagai berikut :

1. Mengetahui atribut terbaik dengan menggunakan metode seleksi fitur *Information Gain*
2. Mengetahui akurasi pengklasifikasian pada metode *K-NN* untuk lama studi mahasiswa dengan menggunakan atribut terbaik dari metode seleksi fitur *Information Gain*
3. Mengetahui akurasi pengklasifikasian pada metode *random forest* untuk lama studi mahasiswa dengan menggunakan atribut terbaik dari metode seleksi fitur *Information Gain*

1.4 Manfaat Penelitian

Peneliti berharap hasil dari penelitian ini memberikan manfaat dan mempunyai kegunaan bagi pihak yang membutuhkan diantaranya:

1. Manfaat Teoritis

Hasil penelitian ini diharapkan dapat dikembangkan dan dijadikan sebagai bahan tinjauan pustaka yang berguna bagi pihak yang melakukan penelitian terkait tentang metode *K-Nearest Neighbor* Dan *Random Forest* Dengan Seleksi Fitur *Information Gain*

2. Manfaat praktis

Penelitian ini memfokuskan pada lama studi mahasiswa sehingga diharapkan dapat dijadikan sebagai bahan pertimbangan bagi para pengambil kebijakan agar dapat menentukan langkah yang strategis dalam meningkatkan mutu lulusan Universitas.

