

BAB I

PENDAHULUAN

1.1 Latar Belakang

Dalam data mining memiliki tahapan *pre processing* yang meningkatkan akurasi pada klasifikasi seperti seleksi fitur (*feature selection*). Seleksi fitur adalah salah satu tahap *pre processing* data yang meningkatkan akurasi, skalabilitas, efektif, dan efisiensi pada klasifikasi yang mempertimbangkan domain dan algoritma karakteristik agar dapat menentukan fitur yang sesuai untuk dipertimbangkan dalam proses pembelajaran.

Metode seleksi fitur yang digunakan dalam penelitian ini adalah *information gain*. *Information gain* merupakan metode yang menggunakan algoritma yang terdapat pada seleksi fitur yang menentukan batas untuk kepentingan sebuah atribut pada klasifikasi. Untuk mengetahui *information gain* dengan menghitung nilai *entropy* sebelum pemisahan dan nilai *entropy* setelah pemisahan kemudian dapat menghitung nilai pada *information gain*.

Klasifikasi adalah suatu proses pembagian data menurut kelas-kelas yang memiliki perbedaan model klasifikasinya dengan menentukan data testing dan data training kedalam atribut data yang sesuai. Ada beberapa metode yang terdapat pada klasifikasi yakni, *Decision Tree*, *Classification And Regression Tree*, *Naive Bayes*, *Neural Network*, *Support Vector Machine*, *K-Nearest Neighbors*, *Random Forest* dan masih banyak lagi. Pada penelitian ini menggunakan metode *random forest*.

Random forest adalah metode klasifikasi yang memiliki beberapa parameter, mudah digunakan dan telah diterapkan pada banyak masalah dengan sampel yang besar. *Random forest* menggunakan *decision tree* sebagai pengklasifikasi dasar. Metode pembelajaran *ensemble* ini digunakan untuk klasifikasi dan regresi data. *Bagging* yang

diajukan oleh Breiman dan *Boosting* yang diajukan oleh Freund dan Schapire (1997) adalah dua metode utama untuk metode *ensemble learning*. Teknik yang digunakan untuk pengambilan sampel ulang data untuk menghasilkan *training* set yang berbeda untuk setiap penggantian pengklasifikasi (Fawagreh *et al.* 2014).

Lama studi seorang mahasiswa merupakan salah satu indikator kewajiban bagi setiap kampus karena sangat berpengaruh pada akreditasi sebagai dasar efektifnya suatu perguruan tinggi baik kampus swasta maupun negeri. Akreditasi sebuah kampus merupakan tolak ukur serta mutu untuk kelayakan pada suatu lembaga pendidikan agar dapat melibatkan mahasiswa dalam persaingan kerja. Data PDDIKTI pada tahun 2018/2019 jumlah mahasiswa UNG mencapai 30.942 orang mahasiswa dan 735 orang dosen yang memiliki rasio 42,1 : 1.

Dari data tata usaha (TU) Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam untuk angkatan 2013 terdapat 104 mahasiswa aktif dengan 85 mahasiswa yang telah lulus dan terdapat 15 variabel dalam penelitian ini yakni, Lama studi, Jenis kelamin, Tempat lahir, Jenis sekolah, Seleksi, Predikat, Dosen PA, Pekerjaan orang tua, Penghs. orang tua, IPK, SKS, Jumlah cuti/nonaktif, Jumlah MK buruk, Jumlah MK bagus dan Waktu studi. Dalam penelitian ini akan menggunakan seleksi fitur *information gain* untuk mencari atribut yang berpengaruh, setelah itu akan digunakan pada pengklasifikasian random forest untuk lama studi mahasiswa.

1.2 Rumusan Masalah

Berdasarkan uraian pada latar belakang, maka dirumuskan permasalahan penelitian adalah sebagai berikut :

1. Bagaimana menentukan atribut terbaik dengan menggunakan metode seleksi fitur *information gain* ?
2. 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 yang menjadi tujuan pada penelitian ini adalah sebagai berikut :

1. Mengetahui atribut terbaik dengan menggunakan metode seleksi fitur *information gain*.
2. 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 atau Keilmuan

Hasil dari penelitian ini diharapkan dapat dikembangkan dan dijadikan sebagai bahan tinjauan pustaka yang berguna bagi pihak yang melakukan penelitian terkait tentang teori seleksi fitur *information gain* dan *random forest*.

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.