

BAB I

PENDAHULUAN

I.1 Latar Belakang

Mahasiswa merupakan salah satu pilar penting dalam siklus hidup suatu perguruan tinggi. Dalam proses berkembangnya, suatu perguruan tinggi dapat dipengaruhi oleh berapa banyak jumlah lulusan strata-1 (S1) dari perguruan tinggi tersebut. Jumlah kelulusan suatu perguruan tinggi terkadang memiliki rasio yang rendah jika dibandingkan dengan jumlah mahasiswa yang diterima pada tahun ajaran yang sama. Tingkat kelulusan mahasiswa yang rendah ini dapat diakibatkan oleh beberapa faktor, seperti banyaknya kegiatan kemahasiswaan yang diikuti, faktor ekonomi, dan beberapa faktor tidak terduga lainnya. Hal ini membuat suatu perguruan tinggi harus memiliki suatu skema atau suatu rumus yang dapat memprediksi apakah mahasiswa tersebut dapat lulus tepat waktu. Normalnya, seorang mahasiswa strata-1 (S1) menempuh pendidikan selama 8 semester. Tetapi dengan adanya beberapa faktor yang telah disebutkan dapat membuat waktu menempuh pendidikan S1 menjadi lebih banyak, atau bahkan gagal lulus.

Penelitian ini akan mencoba untuk membandingkan hasil Analisa dua metode dalam algoritma klasifikasi untuk memprediksi kelulusan mahasiswa. Algoritma yang digunakan ialah Algoritma K-Nearest Neighbour dan Naïve Bayes. Penelitian ini juga bertujuan mengidentifikasi algoritma terbaik di antara dua pilihan algoritma klasifikasi tersebut.

Algoritma K-NN adalah suatu metode yang menggunakan algoritma *supervised*. Perbedaan antara *supervised learning* dengan *unsupervised learning* adalah pada *supervised learning* bertujuan untuk menemukan pola baru dalam data dengan menghubungkan pola data yang sudah ada dengan data yang baru (Giovani, dkk, 2020). Sedangkan pada *unsupervised learning*, data belum memiliki pola apapun, dan tujuan *unsupervised learning* untuk menemukan pola dalam sebuah data. Tujuan dari algoritma K-NN adalah untuk mengklasifikasi objek baru berdasarkan atribut dan *training samples* (Saadatfar, dkk, 2020). Dimana hasil dari sampel uji yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada K-NN (Farokhah, 2020). Pada proses pengklasifikasian, algoritma ini tidak menggunakan model apapun untuk dicocokkan dan hanya berdasarkan pada memori (Muhathir, dkk, 2020). Algoritma *K-Nearest Neighbor* merupakan metode klasifikasi yang mengelompokkan data baru berdasarkan jarak data baru itu kebeberapa data/tetangga (*neighbor*) terdekat (Rohman, 2016). K-NN adalah rumus yang paling sederhana yang sering digunakan dalam pengimplementasian pencarian jarak. Rumus yang digunakan dalam perhitungan K-NN adalah rumus Euclidean Distance (Suwirmayanti, 2017).

Algoritma Naïve Bayes merupakan algoritma data mining yang dapat digunakan untuk melakukan klasifikasi, dan merupakan algoritma yang akan memberikan hasil kemungkinan yang akan terjadi di masa akan datang dari hasil klasifikasi. Naive Bayes merupakan teknik prediksi berbasis probabilitas sederhana yang berdasar pada penerapan aturan bayes dengan asumsi ketidak

tergantungan yang kuat. Selain itu naive bayes juga dapat menganalisa variabel-variabel yang paling mempengaruhinya dalam bentuk peluang (Prasetyo, 2012).

Menurut Olson Delen (2008) menjelaskan Naïve Bayes untuk setiap kelas keputusan, menghitung probabilitas dengan syarat bahwa kelas keputusan adalah benar, mengingat vektor informasi obyek. Algoritma ini mengasumsikan bahwa atribut obyek adalah independen. Probabilitas yang terlibat dalam memproduksi perkiraan akhir dihitung sebagai jumlah frekuensi dari "master" tabel keputusan. Naïve Bayes berasumsi bahwa efek dari suatu pada kelas yang diberikan adalah independen terhadap nilai atribut yang lainnya. Asumsi ini biasa disebut dengan class conditional independence. Itu dibuat untuk menyederhanakan komputasi yang terkait dan dalam hal ini disebut sebagai "naïve".

Naïve Bayes Classifier merupakan sebuah pengklasifikasian probabilitas sederhana yang mengaplikasikan Teorema Bayes dengan asumsi ketidaktergantungan (independent) yang tinggi (Vangara, Vangara, 2020). Keuntungan penggunaan Naïve Bayes Classifier adalah metode ini hanya membutuhkan jumlah data pelatihan (training data) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian.

Berdasarkan masalah yang diuraikan maka penelitian ini membahas analisis komparasi algoritma naive bayes dan k-nearest neighbor. Tujuan diadakannya komparasi ini adalah diharapkan akan diketahui metode yang mendapatkan akurasi lebih besar. Berikutnya, metode yang terbaik diantara keduanya dipakai dalam menentukan prediksi kelulusan mahasiswa, yaitu antara tepat waktu atau terlambat, hasil pengujian dengan mengukur kinerja kedua

metode tersebut menggunakan metode pengujian Confusion Matrix dalam tool weka.

Penelitian dilakukan dengan mengambil data dari indeks prestasi mahasiswa yang diperoleh dari pengambilan matakuliah yang ditawarkan pada setiap semesternya. Kelulusan tepat waktu merupakan indikator keberhasilan mahasiswa dalam proses pendidikan dilihat dari nilai yang diperoleh, indeks prestasi serta faktor lain yang dapat menyebabkan kegagalan masa studi mahasiswa sehingga perlu adanya tool yang mampu memprediksi. Penekanan pada penelitian ini adalah mahasiswa Fakultas Kedokteran Universitas Muhammadiyah Sumatera Utara sebagai training set yang akan diprediksi kelulusannya ditinjau dari indeks prestasi selama 5 semester pertama yaitu semester 1, semester 2, semester 3, semester 4 dan semester 5. Selanjutnya dilakukan analisis komparasi algoritma klasifikasi yaitu Algoritma K-Nearest Neighbor dan Naïve Bayes sehingga dapat diketahui algoritma yang paling akurat untuk memprediksi kelulusan mahasiswa tepat waktu pada Fakultas Kedokteran Universitas Muhammadiyah Sumatera Utara.

Berdasarkan latar belakang dan permasalahan di atas, maka penelitian ini mengangkat judul penelitian “**Analisis Komparasi Algoritma Naïve Bayes dan K-Nearest Neighbor Untuk Memprediksi Kelulusan Mahasiswa Tepat Waktu**”.

I.2 Rumusan Masalah

Pada penelitian ini, rumusan masalah penelitian adalah melakukan perbandingan Algoritma Naïve Bayes dan K-Nearest Neighbor untuk

memprediksi kelulusan mahasiswa tepat waktu, melihat tingkat akurasi dari hasil analisis Algoritma *Naïve Bayes* dan Algoritma *K-Nearest Neighbor*, dan menggunakan aplikasi Weka dalam penerapan Algoritma *Naïve Bayes* dan Algoritma *K-Nearest Neighbor*.

I.3 Tujuan

Adapun tujuan dan harapan yang penulis lakukan dalam penelitian ini agar penelitian ini bermanfaat nantinya adalah:

1. Menghasilkan tingkat akurasi dari Algoritma *Naïve Bayes* dan *K-Nearest Neighbor* dalam memprediksikan kelulusan tepat waktu.
2. Menghasilkan pengambilan keputusan dalam menentukan kebijakan terhadap mahasiswa yang memiliki nilai rendah
3. Menerapkan Algoritma *Naïve Bayes* atau *K-Nearest Neighbor* dalam melakukan prediksi terhadap kelulusan tepat waktu.
4. Menguji validitas dari Algoritma *Naïve Bayes* atau *K-Nearest Neighbor* dengan menggunakan Aplikasi Weka.

I.4 Manfaat

Manfaat yang ingin dicapai Penulis dengan penelitian ini adalah:

1. Manfaat praktis dari hasil penelitian ini diharapkan agar dapat digunakan oleh Fakultas Kedokteran Universitas Muhammadiyah Sumatera Utara dalam melakukan prediksi kelulusan tepat waktu.

2. Manfaat berikutnya sebagai early warning bagi pengelola institusi, jika diketahui adanya mahasiswa yang akan tidak tepat lulus waktu untuk diberikan pembinaan.
3. Data yang sudah dimining dan menghasilkan prediksi kelulusan tepat waktu, maka informasi tersebut dapat dipergunakan sebagai patokan dalam menentukan jumlah mahasiswa pada Penerimaan Mahasiswa Baru.

I.5 Batasan Masalah

Berdasarkan latar belakang di atas untuk lebih focus, maka permasalahan dibatasi pada :

1. Penelitian dilakukan di Fakultas Kedokteran Universitas Muhammadiyah Sumatera Utara.
2. Prediksi kelulusan tepat waktu dengan menganalisis komparasi algoritma Naïve Bayes dan K-Nearest Neighbor
3. Data yang digunakan dalam menentukan prediksi kelulusan tepat waktu adalah indeks prestasi sementara dari semester 1 hingga semester 5 Alumni Mahasiswa Fakultas Kedokteran Angkatan Tahun 2015.
4. Aplikasi yang digunakan untuk analisis ialah Weka.

I.6 Sistematika Penulisan

Penelitian ini mencoba untuk membahas pokok permasalahan secara cermat dan sistematis. Untuk itu pembahasan dibuat dan disusun dengan sistematika penulisan sebagai berikut :

BAB I PENDAHULUAN

Pada bab pendahuluan ini berisi pendahuluan yang mencakup latar belakang, perumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Pada Bab ini berisi tinjauan tentang pustaka terhadap Naïve Bayes, K-Nearest Neighbor, Komparasi Prediksi, dan Aplikasi Weka.

BAB III METODOLOGI PENELITIAN

Metodologi penelitian menjelaskan rangkaian tahapan dalam pembuatan sistem yang terdiri dari pengumpulan data, analisis data, perancangan sampai implementasi beserta analisa yang akan dilakukan.

BAB IV HASIL

Hasil penelitian yaitu berupa laporan hasil pelaksanaan metode/teknik penelitian dan menyajikan data yang mendukung hasil tersebut. Penyajian data dan penjelasannya dilakukan secara terurut dan logis menggunakan teks dan ilustrasi lainnya (misalnya, tabel dan gambar).

BAB V PEMBAHASAN

Pembahasan penelitian yaitu menerjemahkan makna dari hasil yang diperoleh untuk menjawab pertanyaan atau masalah penelitian. Fungsi lainnya adalah untuk menjelaskan pemahaman baru yang didapatkan dari hasil penelitian, yang diharapkan

berguna dalam pengembangan keilmuan. Dalam penelitian ini, hasil pembahasan bukan pengembangan keilmuan secara substansial, tetapi cukup terhadap pemahaman personal dalam implementasi konsep atau teori.

BAB VI KESIMPULAN DAN SARAN

Pada Bab ini memuat tentang kesimpulan dari penelitian yang telah dilakukan dan saran yang dapat digunakan oleh pihak-pihak yang terkait maupun pihak lain yang membahas permasalahan yang sama.