

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Persaingan antar perguruan tinggi semakin hari semakin ketat. Masing masing perguruan tinggi berlomba-lomba untuk mendapatkan mahasiswa sebanyak banyaknya. Semakin banyak mahasiswa yang masuk, tentu akan semakin membanggakan bagi pihak perguruan tinggi. Namun, kualitas perguruan tinggi tidak dapat ditentukan dari jumlah mahasiswa yang masuk saja, tetapi juga jumlah mahasiswa yang lulus. Salah satu faktor yang terpengaruhi oleh kondisi ini adalah akreditasi perguruan tinggi.

Menurut Nasiri, mahasiswa yang tidak lulus tepat waktu, akan mengakibatkan penumpukan jumlah mahasiswa. Semakin banyak mahasiswa yang menumpuk, semakin tidak seimbang pula antara jumlah mahasiswa dengan jumlah dosen yang ada. Hal ini dapat menimbulkan berbagai masalah bagi perguruan tinggi, khususnya dalam hal belajar mengajar. Sehingga, pihak perguruan tinggi sangat mengupayakan agar seluruh mahasiswanya dapat lulus tepat waktu (Pembicaraan pribadi, 30 Mei 2018).

Universitas AMIKOM Yogyakarta adalah suatu perguruan tinggi swasta yang memiliki prodi Informatika dengan standar kelulusan tepat waktu 48 bulan atau 8 semester. Setiap tahunnya AMIKOM menerima kurang lebih 1.000 mahasiswa baru pada prodi Informatika. Pada tahun 2012, jumlah mahasiswa baru yang diterima pada prodi Informatika mencapai 1009 mahasiswa. Namun sayangnya, dari 1009 mahasiswa, hanya 527 mahasiswa yang dapat lulus tepat waktu.

Sedangkan pada tahun 2013, jumlah mahasiswa baru yang diterima mencapai 859 mahasiswa dan yang berhasil lulus tepat waktu hanya 435 mahasiswa. Angka ini jauh mendekati target kelulusan mahasiswa yang ditargetkan oleh AMIKOM.

Perguruan tinggi saat ini dituntut untuk memiliki keunggulan bersaing dengan memanfaatkan semua sumber daya yang dimiliki. Selain sumber daya sarana, prasarana, dan manusia, sistem informasi adalah salah satu sumber daya yang dapat digunakan untuk meningkatkan keunggulan bersaing [1]. Data mahasiswa yang menumpuk merupakan salah satu sumber daya yang dapat dimanfaatkan dengan menggunakan teknik *Data Mining*. Teknik *Data Mining* dapat dipakai untuk klasifikasi waktu kelulusan mahasiswa. Salah satu algoritma yang dapat digunakan untuk klasifikasi adalah *Naïve Bayes Classifier*. Proses *Data Mining* dapat menghasilkan pola tertentu yang dapat digunakan untuk mengklasifikasikan mahasiswa lulus tepat waktu, atau lulus tidak tepat waktu berdasarkan data mahasiswa yang sudah lulus.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang ada, dapat diambil poin-poin masalah sebagai berikut:

1. Bagaimana melakukan penerapan algoritma *Naïve Bayes* untuk dapat memprediksi waktu masa studi mahasiswa?
2. Bagaimana akurasi algoritma *Naïve Bayes* dalam melakukan prediksi waktu masa studi mahasiswa?

1.3 Batasan Masalah

Agar penulisan skripsi ini tidak menyimpang dari tujuan yang direncanakan, maka penulis menetapkan batasan-batasan sebagai berikut:

1. Subjek penelitian berfokus pada jurusan Informatika untuk jenjang S1.
2. Platform yang digunakan oleh sistem adalah website.
3. Penelitian ini menggunakan algoritma *Naïve Bayes*.
4. Variabel yang digunakan dalam penelitian ini adalah: Jenis Kelamin, Tipe Sekolah, Kota Sekolah, IP Semester 1, IP Semester 2, IP Semester 3, IP Semester 4, IPK, dan Keterangan Kelulusan.
5. Pengukuran akurasi dan performa dilakukan menggunakan *Confussion Matrix*, dengan *K-Fold Cross Validation* sebagai metode untuk membagi data *training*, dan data *testing*.
6. Dataset yang digunakan adalah mahasiswa Universitas AMIKOM Yogyakarta Jurusan Informatika angkatan 2012 dan 2013 yang sudah lulus masing-masing sebanyak 150.
7. Jumlah data training 300 data.
8. Data target yang digunakan adalah mahasiswa Universitas AMIKOM Yogyakarta Jurusan Informatika angkatan 2016 yang sudah menempuh perkuliahan hingga tahun ke-2 dan belum lulus sebanyak 50 mahasiswa.

1.4 Maksud dan Tujuan Penelitian

Tujuan dari pembuatan skripsi ini adalah:

1. Menerapkan algoritma Naïve Bayes pada klasifikasi waktu kelulusan mahasiswa terhadap data alumni Universitas AMIKOM Yogyakarta.
2. Mencari nilai akurasi algoritma Naïve Bayes dalam mengklasifikasi waktu kelulusan mahasiswa terhadap data alumni Universitas AMIKOM Yogyakarta.
3. Mencari pola yang berhubungan dengan waktu kelulusan mahasiswa dari data induk, dan data akademik alumni Universitas AMIKOM Yogyakarta menggunakan algoritma Naïve Bayes.

1.5 Manfaat Penulisan

1. Bagi Universitas AMIKOM Yogyakarta
Mendapatkan hasil klasifikasi dari mahasiswa tahun ke-3 mengenai waktu kelulusan guna melakukan tindakan lanjutan untuk meningkatkan jumlah mahasiswa yang dapat lulus tepat waktu.
2. Bagi Ilmu Pengetahuan
 - a. Mengetahui bagaimana implementasi algoritma naïve bayes pada kasus *data mining* untuk klasifikasi waktu kelulusan.
 - b. Mengetahui cara mengolah data untuk klasifikasi waktu kelulusan.

1.6 Metode Penelitian

Dalam penelitian ini akan menjelaskan tahapan-tahapan *data mining* untuk klasifikasi waktu kelulusan mahasiswa sebagai berikut.

1.6.1 Metode Pengumpulan Data

Pada penelitian ini, dilakukan beberapa metode yang akan digunakan untuk mengumpulkan data penelitian, berikut merupakan metode pengumpulan data yang digunakan.

1. Wawancara

Wawancara dilakukan dengan Ketua Jurusan Informatika, Sekretaris Program Studi Informatika, dan Direktur Innovation Center Universitas AMIKOM Yogyakarta untuk mendapatkan data terkait kriteria kelulusan mahasiswa di Universitas AMIKOM Yogyakarta.

2. Dokumentasi

Data yang digunakan adalah data mahasiswa yang berasal dari database Universitas AMIKOM Yogyakarta. Data didapat dengan mengajukan permintaan ke pihak DAAK dan Direktorat Innovation Center Universitas AMIKOM Yogyakarta.

3. Kepustakaan

Data mengenai teori dan hasil percobaan sejenis didapat melalui membaca jurnal ilmiah internasional, dan jurnal ilmiah nasional dengan tema *data mining*, *machine learning*, *cross validation*, dan *confusion matrix* dari perpustakaan dan internet.

1.6.2 Tahapan Penelitian

Untuk melakukan penelitian ini, dilakukan beberapa tahapan penelitian sebagai berikut:

1. Metode Analisis

Dalam metode analisis ini terdapat beberapa tahapan sebagai berikut:

a. Data Collection

Data yang digunakan dalam penelitian adalah data alumni yang diperoleh dari Direktorat Innovation Center Universitas AMIKOM Yogyakarta.

b. Data Preprocessing

Tahap preprocessing adalah tahap awal pengolahan data dalam *data mining*.

Berikut tahap-tahapannya:

- a) Data Cleaning merupakan tahap untuk membersihkan data pada variabel tipe sekolah, kota sekolah, jenis kelamin, IP Semester 1, IP Semester 2, IP Semester 3, IP Semester 4, keterangan lulus dari data kosong, dan data yang tidak relevan dengan cara mengisi data tersebut dengan nilai mediannya.
- b) Data Selection merupakan tahap untuk membuang data dengan variabel tanggal lulus, dan lama studi karena variabel tersebut tidak memiliki informasi yang dapat mempengaruhi kelulusan mahasiswa. Variabel gaji orang tua juga dihilangkan karena banyak memiliki nilai kosong.
- c) Data Transformation merupakan tahap untuk mengubah bentuk data pada variabel tipe sekolah, dan kota sekolah ke dalam beberapa kelompok. Ditambahkan pula variabel IPK dari hasil perhitungan rata-rata variabel

nilai IP Semester 1 hingga IP Semester 4. Pada tahap ini juga dilakukan perubahan bentuk data pada variabel nilai yang meliputi IP Semester 1 hingga IPK ke dalam bentuk huruf (diskret).

c. Klasifikasi Naïve Bayes

Pada tahap ini data akan dilatih menggunakan algoritma Naïve Bayes untuk menemukan pola yang dapat digunakan untuk proses klasifikasi data baru. Proses pelatihan data dilakukan dengan menggunakan sistem. Hasil dari proses ini adalah sebuah model Naïve Bayes yang dapat digunakan untuk mengklasifikasi data yang belum memiliki kelas.

d. Menghitung akurasi model

Untuk menguji seberapa akurat akurasi dari model Naïve Bayes maka akan dilakukan pengujian menggunakan metode *confusion matrix* untuk menghasilkan performa model berupa nilai akurasi, *precision*, *recall*, dan *f1-score*, dengan *10-Fold Cross Validation* sebagai metode untuk membagi data *training*, dan data *testing*. Masing-masing data *testing* akan diuji menggunakan data *training* kemudian hasil klasifikasi dari data *testing* akan didokumentasi menggunakan metode *confusion matrix*. Hasil akhir yang diperoleh adalah hasil dari nilai rata-rata seluruh *fold*.

2. Metode Perancangan

Perancangan yang perlu dilakukan adalah perancangan untuk mengimplementasikan algoritma ke dalam aplikasi menggunakan *flowchart*. Untuk menerapkan algoritma Naïve Bayes pada suatu sistem, diperlukan perancangan

sistem mulai dari tahap *preprocessing*, *training*, hingga *testing*, serta perancangan algoritma Naïve Bayes untuk mengetahui bagaimana alur proses algoritma bekerja.

3. Metode Implementasi

Proses implementasi dilakukan menggunakan metode *One Hot Encoding* yang berfungsi untuk mengubah bentuk data ke dalam bentuk biner sehingga dapat digunakan untuk perhitungan algoritma Naïve Bayes pada sistem. Pembuatan sistem dilakukan menggunakan bahasa pemrograman Python dengan memanfaatkan *library* sklearn, numpy, pandas, dan matplotlib.

4. Metode Pengujian

Pengujian dilakukan dengan membagi dataset ke dalam kelompok data *training*, dan data *testing* menggunakan metode *10-Fold Cross Validation*. Masing-masing *fold* akan dihitung performanya menggunakan metode *confusion matrix* dengan ukuran performanya adalah akurasi, *precision*, *recall*, dan *f1-score*.

1.7 Sistematika Penulisan

Laporan penelitian ini akan disusun secara sistematis yang akan diurutkan sebagai berikut:

BAB I PENDAHULUAN

Pada bab ini akan menjelaskan mengenai latar belakang masalah, perumusan masalah, batasan masalah, tujuan, manfaat, dan sistematika penulisan penelitian.

BAB II LANDASAN TEORI

Pada bab ini menjelaskan tentang teori yang digunakan sebagai dasar dalam analisis penelitian. Mencakup tinjauan pustaka meliputi Pengertian *Data Mining*, Algoritma *Bayes*, *Naïve Bayes*, *Multinomial Naïve Bayes*, *Laplace Correction*, *One Hot Encoding*, *K-Fold Cross Validation*, dan *Confusion Matrix*.

BAB III ANALISIS DAN PERANCANGAN

Pada bab ini menjelaskan mengenai perancangan model, dan cara kerja yang akan dilakukan di dalam klasifikasi waktu kelulusan mahasiswa.

BAB IV IMPLEMENTASI DAN PEMBAHASAN

Pada bab ini akan menjelaskan kode program dari proses *Data Mining* mulai dari proses *preprocessing* hingga proses untuk mendapatkan model *Naive Bayes*. Ditampilkan pula hasil dari implementasi program berupa *screenshot*.

BAB V PENUTUP

Pada bab ini akan menjelaskan kesimpulan yang didapat dari hasil analisis, perancangan, dan implementasi serta saran untuk pengembangan penelitian yang lebih baik.