

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Perguruan Tinggi merupakan tempat dimana mahasiswa menuju ke jenjang pendidikan yang lebih tinggi sesuai dengan jurusan yang diambil. Setiap perguruan tinggi memiliki waktu maksimal yang diberikan kepada mahasiswa dalam menyelesaikan studinya, jika mahasiswa tersebut sudah melewati batas waktu yang telah ditentukan maka mahasiswa tersebut akan dikeluarkan dari perguruan tinggi. Di Universitas Tanjungpura terutama di jurusan Informatika masa studi yang diperlukan untuk menyelesaikan studi adalah selama 4 tahun atau sebanyak 8 semester dan batas waktu maksimal yang diberikan untuk mahasiswa dalam menyelesaikan studinya adalah selama 7 tahun atau sebanyak 14 semester. Mahasiswa diharapkan mampu menyelesaikan studinya dalam waktu yang telah diberikan agar bisa mendapatkan gelar sarjana.

Jurusan Informatika Universitas Tanjungpura memiliki tujuan yang salah satunya adalah menyelenggarakan program studi yang terakreditasi sangat baik, yang memiliki sistem pendidikan dan pembelajaran yang didukung oleh sumber daya manusia, fasilitas pendidikan dan sistem manajemen akademik yang baik. Seperti yang diketahui bahwa mahasiswa merupakan salah satu faktor penentu yang dapat mempengaruhi kualitas, mutu, dan akreditasi perguruan tinggi maupun jurusan. ketepatan waktu mahasiswa dalam menyelesaikan studinya merupakan salah satu aspek penilaian dalam menentukan akreditasi perguruan tinggi dan jurusan. Akan tetapi permasalahan yang terjadi adalah masih banyaknya jumlah mahasiswa yang lulus tidak tepat waktu, sehingga akan memberikan dampak yang buruk untuk akreditasi perguruan tinggi maupun jurusan. Selain itu dampak buruk juga bisa terjadi kepada mahasiswa itu sendiri karena mahasiswa tersebut dapat di berhentikan jika waktu yang telah diberikan dalam menyelesaikan studinya telah habis. Dengan memanfaatkan data akademik mahasiswa yang tersimpan dalam database perguruan tinggi diharapkan

nantinya bisa memberikan informasi yang bermanfaat dengan dilakukan pengolahan pada data akademik mahasiswa tersebut.

Pada penelitian ini penulis akan membuat sebuah sistem yang dapat mengklasifikasikan masa studi mahasiswa Informatika Universitas Tanjungpura, sehingga nantinya dapat mempermudah dalam menampilkan informasi mahasiswa mana saja yang termasuk ke dalam kelas lulus tidak tepat waktu dan kelas lulus tepat waktu. Pada penelitian yang telah dilakukan sebelumnya oleh (Budiman et al., 2015) atribut yang digunakan untuk proses klasifikasi masa studi mahasiswa yaitu Indeks Prestasi Semester 1 sampai Indeks Prestasi Semester 4, pada penelitian ini penulis menambahkan atribut Indeks Prestasi Semester 5 sampai Indeks Prestasi Semester 7, jalur masuk, dosen pembimbing, asal kabupaten dan Jenis sekolah yang merupakan data yang didapatkan dari data akademik mahasiswa Informatika Universitas Tanjungpura. Dalam melakukan proses klasifikasi terdapat berbagai macam algoritma yang dapat digunakan, antara lain K-Nearest Neighbor (Yustanti, 2012)(Fatahna et al., 2017) (Noviansyah et al., 2018), Naive Bayes(Indriani, 2014)(Saleh, 2015)(Qodrat, 2017), C4.5(Fajri, 2017)(Larose, 2005)(Latifah et al., 2018), Support Vector Machine (Gorunescu, 2011)(Yang, 2019) dan lainnya. Algoritma yang akan digunakan untuk proses klasifikasi adalah algoritma K-Nearest Neighbor dan Algoritma Naive Bayes. Penggunaan algoritma Naive Bayes pada penelitian ini adalah untuk membandingkan performa algoritma K-Nearest Neighbor untuk membuktikan bahwa algoritma K-Nearest Neighbor lebih baik dalam proses klasifikasi. Pada penelitian yang dilakukan oleh (Budiman et al., 2015) algoritma K-Nearest Neighbor dan algoritma Naive Bayes merupakan algoritma yang memberikan akurasi yang tinggi dibandingkan algoritma SVM dan Decision Tree.

Terdapat 3 penelitian terkait pada penelitian ini, pertama dari penelitian (Srianto & Mulyanto, 2016) yang melakukan penelitian mengenai “Perbandingan K-Nearest Neighbor dan Naive Bayes untuk Klasifikasi Tanah Layak Tanam Pohon Jati”. Penelitian ini melakukan perbandingan algoritma dengan melakukan pengujian pada setiap algoritma klasifikasi. Penelitian ini melakukan pengujian pada algoritma K-

Nearest Neighbor dengan menguji nilai dari parameter k (jarak antar data) dengan nilai 1, 3, 5, 7, dan 9. Untuk algoritma Naive Bayes dilakukan pengujian sebanyak 5 kali dengan percentage split yang dibagi menjadi data training dan data testing yaitu 95% dan 5%, 85% dan 15%, 75% dan 25%, 65% dan 35%, 55% dan 45%. Hasil yang didapat dari penelitian ini adalah bahwa akurasi paling tinggi pada algoritma K-Nearest Neighbor yaitu sebesar 97.14% dengan nilai parameter k = 3 dan akurasi paling tinggi pada algoritma Naive Bayes yaitu sebesar 88.46% dengan percentage split 75% dan 25%. Penelitian kedua dari penelitian (Muqorobin, Kusriani, 2019) yang berjudul “Optimasi Metode Naive Bayes dengan *Feature Selection Information Gain* Untuk Prediksi Keterlambatan Pembayaran Sumbangan Pembinaan Pendidikan Sekolah”. Penelitian ini menguji pengaruh terhadap performa pada algoritma Naive Bayes dengan menambahkan *Feature Selection Information Gain*. Hasil dari penelitian ini adalah dengan menambahkan *Feature Selection Information Gain* dapat meningkatkan tingkat akurasi pada algoritma Naive Bayes dengan tingkat akurasi sebesar 90% dan tanpa menambahkan *Feature Selection Information Gain* pada algoritma Naive Bayes mendapatkan akurasi sebesar 80%. Penelitian ketiga dari penelitian (Ivandari et al., 2017) yang berjudul “*Data Attribute Selection With Information Gain to Improve Credit Approval Classification Performance using K-Nearest Neighbor Algorithm*”. Penelitian ini melakukan pengujian algoritma K-Nearest Neighbor dengan menguji nilai dari parameter k (jarak antar data) dengan nilai 1, 3, 5, 7, 9, 11, 13, 15, dan 17 serta melakukan pengujian atribut yang telah dilakukan *Feature Selection Information Gain*. Hasil yang didapatkan bahwa dengan menambahkan Information Gain nilai akurasi algoritma K-Nearest Neighbor menjadi meningkat pada dua dataset. pada *dataset UCI credit approval* nilai akurasi meningkat sebesar 7.53% dan pada dataset *credit card customer* nilai akurasi meningkat sebesar 3.26%.

Berdasarkan uraian di atas, maka peneliti akan menambahkan *Feature Selection Information Gain* untuk mengetahui apakah akan meningkatkan performa algoritma klasifikasi. Pada penelitian ini juga nantinya akan membandingkan algoritma K-Nearest Neighbor dan algoritma Naive Bayes untuk menganalisa performa dari setiap algoritma klasifikasi.

## **1.2 Perumusan Masalah**

Berdasarkan latar belakang permasalahan, maka rumusan masalah pada penelitian ini adalah bagaimana mengimplementasikan dan menganalisa algoritma K-Nearest Neighbor dengan membandingkan performa algoritma Naive Bayes serta pengaruh penambahan *Feature Selection Information Gain* dalam mengklasifikasi masa studi mahasiswa Informatika Universitas Tanjungpura.

## **1.3 Tujuan Penelitian**

Tujuan dalam penelitian yang akan dilakukan untuk tugas akhir ini sebagai berikut :

1. Menentukan parameter yang baik untuk digunakan dalam proses klasifikasi masa studi mahasiswa Informatika Universitas Tanjungpura dengan menggunakan *Feature Selection Information Gain*.
2. Menentukan nilai parameter k (jarak terdekat antar data) paling optimal yang memberikan tingkat akurasi yang paling tinggi pada algoritma K-Nearest Neighbor dalam menentukan kelas pada data.
3. Membandingkan performa algoritma klasifikasi K-Nearest Neighbor dengan algoritma klasifikasi Naive Bayes.
4. Menghasilkan sistem klasifikasi masa studi mahasiswa Informatika Universitas Tanjungpura.

## **1.4 Pembatasan Masalah**

Batasan masalah dari penelitian yang akan dilakukan, agar nantinya tidak terjadinya penyimpangan inti permasalahan yang akan di lakukan yaitu sebagai berikut :

1. Data yang digunakan untuk dijadikan data training merupakan data akademik mahasiswa jurusan Informatika Universitas Tanjungpura dari angkatan 2010 sampai angkatan 2016 yang telah menyelesaikan studinya.
2. Data yang digunakan berasal dari UPT Teknologi Informasi dan Komunikasi Universitas Tanjungpura.

3. *Feature selection* yang digunakan untuk menentukan parameter yang paling pengaruh menggunakan *Information Gain*.
4. Algoritma yang akan digunakan untuk klasifikasi adalah algoritma K-Nearest Neighbor dan Algoritma Naive Bayes.
5. Proses klasifikasi masa studi mahasiswa menggunakan beberapa parameter yaitu Indeks Prestasi Semester 1, Indeks Prestasi Semester 2, Indeks Prestasi Semester 3, Indeks Prestasi Semester 4, Indeks Prestasi Semester 5, Indeks Prestasi Semester 6, Indeks Prestasi Semester 7, jalur masuk, dosen pembimbing akademik, Jenis Sekolah dan asal kabupaten. Pelabelan data akan dibagi menjadi 2 kelas yaitu kelas mahasiswa lulus tepat waktu dan kelas mahasiswa lulus tidak tepat waktu.

### **1.5 Sistematika Penulisan**

Sistematika dalam penulisan untuk tugas akhir ini disusun dalam lima bab yang terdiri dari:

1. Bab I Pendahuluan adalah bab yang dimana terdiri dari latar belakang, perumusan masalah, tujuan penelitian, pembatasan masalah, dan sistematika penulisan.
2. Bab II Tinjauan Pustaka adalah bab yang berisikan landasan teori yang berkaitan dengan penelitian yang akan dilakukan nantinya.
3. Bab III Metodologi Penelitian adalah bab yang berisikan mengenai langkah – langkah penelitian yang terdiri dari pengumpulan data, preprocessing data dan perancangan sistem, .
4. Bab IV Hasil dan Analisis adalah bab yang berisikan implementasi pada sistem, tampilan antar muka sistem yang telah dibuat, dan analisis hasil uji coba sistem dan algoritma klasifikasi.
5. Bab V Penutup adalah bab yang berisikan kesimpulan dari hasil penelitian yang telah dilakukan serta saran untuk melakukan pengembangan maupun perbaikan pada penelitian yang telah dilakukan.