

BAB I

PENDAHULUAN

1.1. Latar Belakang

Perkembangan teknologi yang sangat pesat membuat kemudahan bagi pemakainya dalam berbagai jenis bidang dan aspek. Dalam penanganan data yang besar, kemajuan teknologi selalu dikaitkan dengan data mining dalam penanganannya.

Status mahasiswa merupakan hal terpenting dalam proses pendataan kategori mahasiswa dalam perkuliahan. Banyaknya mahasiswa yang ada dalam prodi Informatika Universitas Bhayangkara Jaya menjadi tantangan tersendiri dalam penanganan data status mahasiswa bagi prodi karena proses klasifikasi status mahasiswa dengan cara manual dirasa kurang efektif dan memakan waktu yang banyak. Proses penentuan status mahasiswa dalam Universitas Bhayangkara Jaya prodi Informatika dikategorikan kedalam 4 status yaitu, aktif, non-aktif, cuti, dan drop out. Untuk melakukan proses klasifikasi status mahasiswa tersebut prodi menganalisa status pengisian krs mahasiswa tiap semester. Prodi melakukan pengumpulan data dan penilaian dalam status pengisian krs mahasiswa. Kemudian data status pengisian krs tersebut diolah dan diproses sampai akhirnya prodi mengambil keputusan status mahasiswa tersebut masuk dalam kategori aktif, non-aktif, cuti dan dropout.

Data mining adalah proses yang menggunakan teknik statistika, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai basis data besar (Wahyudi et al., 2020). Dalam data mining, ada banyak metode yang digunakan untuk menangani banyaknya data. Salah satu metodenya adalah metode klasifikasi. Metode klasifikasi merupakan Pengelompokan berdasarkan hubungan antara variabel kriteria dengan variabel target (Bulolo, 2020). Algoritma C4.5 adalah salah satu metode untuk membuat decision tree berdasarkan training data yang telah disediakan (Wanto, Siregar, Windarto, et al., 2020). Algoritma C4.5 dapat digunakan dalam proses klasifikasi dan juga proses

prediksi. Algoritma C4.5 merupakan pengembangan dari ID3. Algoritma C4.5 dan decision tree merupakan dua model yang tidak terpisahkan, karena untuk membangun sebuah decision tree, dibutuhkan algoritma C4.5 (Aldo et al., 2021). Pada penerapan kasus klasifikasi, algoritma ini mampu menghasilkan performa yang baik

Selain itu penelitian tentang algoritma C4.5 pernah dilakukan oleh Darmawan (2017) mengimplementasikan *data mining* dengan metode algoritma C4.5 untuk memprediksi status mahasiswa dengan menggunakan 10 dataset secara acak. Dari dataset tersebut 9 bagian digunakan sebagai data latih dan 1 bagian digunakan sebagai data uji. Lalu hasil akurasi dengan *confusion matrix* menghasilkan 93,05%.

Penelitian lainnya dilakukan oleh Nasrullah (2018) mengimplementasikan *data mining* dengan metode algoritma C4.5 untuk klasifikasi mahasiswa berpotensi *drop out* dengan menggunakan dataset yang sudah diseleksi menghasilkan 9 atribut. Lalu dari hasil penentuan akar dengan menghitung entropi dan *Gain* menghasilkan 17 rule. Namun akurasi dan tingkat error tidak dihitung karena hanya melihat hasil klasifikasi saja.

Penelitian lainnya juga dilakukan oleh Safii (2018) mengimplementasikan *data Mining* metode algoritma *ID3* untuk menentukan status mahasiswa dengan 10 atribut didapatkan 8 rule sehingga didapatkan hasil faktor yang mempengaruhi mahasiswa aktif adalah apabila mahasiswa memiliki IPK tinggi antara 2,75-4,00. Status tidak aktif sementara apabila mahasiswa memiliki IPK sedang dan dipengaruhi oleh waktu kuliah, penghasilan orang tua dan kehadiran. Lalu status tidak aktif cenderung memiliki IPK rendah.

Selanjutnya penelitian dilakukan oleh Maulana & Kumalasari (2019) mengimplementasikan *data mining* dengan metode komparasi antar metode algoritma klasifikasi *decision tree*, *naïve bayes*, *k-nearest neighbor*, *random forest* dan *random tree* untuk menentukan status kelulusan mahasiswa dengan menggunakan 12 atribut dari dataset yang didapat. Penelitian ini menghasilkan hasil pengujian performa dengan T-Test, Algoritma k-NN memiliki nilai akurasi tinggi yaitu 99,98% dan akurasi terendah adalah metode *random tree* yaitu 87,56%.

Sedangkan pengujian akurasi menggunakan ROC curve menunjukan AUC terbaik diperoleh oleh algoritma Naïve Bayes dengan nilai 1.00 dan nilai terendah dihasilkan oleh algoritma K-NN dengan nilai 0,500.

Dari latar belakang dan beberapa penelitian diatas, implementasi *data mining* dan metode algoritma C4.5 dapat digunakan sebagai metode untuk klasifikasi status mahasiswa dengan menentukan atribut serta pembagian dataset berupa data uji dan data latih.

1.2. Identifikasi Masalah

Masalah yang dapat diidentifikasi dari latar belakang diatas, yaitu adanya kesulitan manajemen prodi Informatika dalam melakukan klasifikasi dan prediksi terhadap status mahasiswa yang berupa aktif, Non-aktif, cuti, dan *drop out*.

1.3. Rumusan Masalah

Berdasarkan latar belakang diatas maka rumusan masalah dari penelitian ini adalah:

1. Bagaimana proses klasifikasi status mahasiswa aktif, non-aktif, cuti, dan *drop out*.
2. Bagaimana proses penentuan status mahasiswa aktif, non-aktif, cuti, dan *drop out*.
3. Bagaimana cara kerja algoritma C4.5.
4. Bagaimana menghasilkan sistem yang dapat membantu proses klasifikasi dan penentuan status mahasiswa.

1.4. Batasan Masalah

Dikarenakan adanya keterbatasan waktu dalam penulisan dan kemampuan, maka penulis membatasi permasalahan-permasalahan yang akan dibahas dalam penelitian ini. Adapun batasan-batasan masalah dalam proposal skripsi ini antara lain:

1. Penelitian ini hanya mencakup klasifikasi data status mahasiswa di program studi Informatika Universitas Bhayangkara Jakarta Raya.
2. Objek yang diklasifikasi untuk menentukan status mahasiswa adalah status pengisian krs, status cuti, dan status mengisi krs per semesternya
3. .Algoritma yang digunakan dalam melakukan proses klasifikasi ini adalah algoritma C4.5.
4. Sistem yang dibangun menghasilkan data hasil klasifikasi dan proses prediksi.

1.5. Tujuan Penelitian

Berdasarkan rumusan masalah yang ada, maka tujuan penelitian ini adalah:

- a. Mampu melakukan klasifikasi data mahasiswa kedalam status aktif, nonaktif, cuti dan *drop out*.
- b. Mampu mengimplementasikan dan menerapkan algoritma C4.5 ke dalam sebuah system pengelompokan status mahasiswa.

1.6. Manfaat Penelitian

Adapun manfaat dari penelitian yang dilakukan ini adalah:

1. Bagi instansi
Penelitian ini dilakukan agar instansi perguruan tinggi dapat mengetahui seberapa banyak siswa yang aktif, non-aktif, cuti, dan *drop out* sebagai evaluasi dan memberi pemberitahuan mengenai status mahasiswa.
2. Bagi penulis
Penelitian ini dilakukan untuk mengetahui bagaimana cara mengimplementasikan metode klasifikasi menggunakan C4.5 serta mengetahui hasil perhitungannya.

3. Bagi perkembangan ilmu pengetahuan

Penelitian ini dilakukan sebagai bahan referensi atau rujukan bagi peneliti lain dalam menambah ilmu pengetahuan yang ada relevansinya.

1.7. Tempat dan Waktu Penelitian

1.7.1. Tempat penelitian

Penelitian ini dilakukan di Universitas Bhayangkara Jakarta Raya kampus II, yang beralamat di Jl. Perjuangan No.81, RT.003/RW.002, Marga Mulya, Kec. Bekasi Utara, Kota Bks, Jawa Barat 17143 program studi Informatika.

1.7.2. Waktu Penelitian

Penelitian yang dilakukan di Universitas Bahyankara Jakarta Raya dilakukan terhitung sejak tanggal 21 Maret 2021 sampai selesainya penelitian.

1.8. Metode Penelitian

1.8.1. Wawancara

Wawancara yang dilakukan adalah melakukan wawancara dengan kepala program studi Informatika Universitas Bhayangkara Jakarta Raya untuk mendapatkan informasi seputar masalah dan *data set* terhadap penelitian terkait.

1.8.2. Studi Literatur

Studi literatur yang dilakukan dalam penelitian ini adalah dengan mengumpulkan dan mempelajari bahan referensi baik dari buku, jurnal, makalah, internet, dan sumber lainnya.

1.8.3. Metode Pengumpulan Data

Pada tahap ini data yang dikumpulkan adalah data akademik mahasiswa program studi Informatika Universitas Bhayangkara Jaya pada tahun 2019.

1.8.4. Metode Analisis

Pada tahap ini dilakukan analisis terhadap hasil dari studi literatur untuk mendapatkan pemahaman mengenai metode klasifikasi menggunakan algoritma C4.5 untuk menyelesaikan masalah pengeompokan pada penentuan status mahasiswa melalui status pengisian krs, status cuti, dan status skripsi.

1.8.5. Metode Perancangan

Pada tahap perancangan system, dilakukan pengumpulan data dan merancang antar muka dari system yang akan dibuat. Proses perancangan dilakukan berdasarkan hasil analisis studi literatur yang telah didapatkan sebelumnya

1.9. Sistematika Penulisan

Tujuan dari sistematika penulisan ini yaitu, untuk mempermudah penulis menyusun proposal skripsi. Adapun sistematika penulisan terdiri dari:

BAB I PENDAHULUAN

Pada bab ini berisi tentang Latar Belakang Masalah, Identifikasi Masalah, Rumusan Masalah, Batasan Masalah, Tujuan Penelitian, Manfaat Penelitian, Tempat dan Waktu Penelitian, Metode Penelitian, serta Sistematika Penulisan.

BAB II LANDASAN TEORI

Pada bab ini berisi tentang beberapa teori yang penunjang dan pendukung serta berhubungan dengan pokok pembahasan dalam proposal skripsi ini yang secara garis besar berisi tentang sistem pengelompokan status mahasiswa.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan metode penelitian yang digunakan dan dilakukan dalam pengerjaan proposal skripsi dan perancangan sistem ini dibuat.

BAB IV PERANCANGAN SISTEM DAN IMPLEMENTASI

Bab ini menjelaskan langkah-langkah implementasi algoritma C4.5 baik berupa perhitungan dan juga implementasi dalam sistem serta pengujian sistem.

BAB V PENUTUP

Bab ini menampilkan kesimpulan dan saran dari hasil penelitian dari penelitian.

