

BAB I

PENDAHULUAN

1.1. Latar Belakang

Perguruan tinggi merupakan suatu instansi pengelola pendidikan akademik tingkat lanjut bagi mahasiswa. Lima lembaga perguruan tinggi diantaranya adalah universitas, institut, sekolah tinggi, akademi dan politeknik. Pusat Statistik Pendidikan Badan Penelitian dan Pengembangan Departemen Pendidikan Nasional Republik Indonesia memiliki data yang menyebutkan bahwa jumlah lembaga penyelenggara perguruan tinggi mengalami peningkatan setiap tahunnya. Pada tahun 2018 tercatat 3171 perguruan tinggi diselenggarakan di Indonesia. Perguruan tinggi diharapkan mampu menghasilkan intelektual muda yang berkualitas dan menghasilkan sumber daya manusia yang berpengetahuan juga kreatif, dan mahasiswa dapat menyelesaikan studi tepat waktu. Semakin meningkatnya jumlah perguruan tinggi, semakin meningkat jumlah sumber daya manusia yang berkualitas yang menjadi *outcome* bagi Institusi perguruan tinggi. Salah satu faktor penentu penilaian kualitas perguruan tinggi adalah persentase kemampuan mahasiswa untuk menyelesaikan studi tepat waktu. Berdasarkan matriks penilaian instrument akreditasi program studi (BAN-PT) [1], bahwa persentase mahasiswa lulus tepat waktu adalah salah satu unsur penilaian akreditasi Perguruan Tinggi. Data yang didapat dari Pusat Statistik Pendidikan Badan Penelitian dan Pengembangan Departemen Pendidikan Nasional Republik Indonesia [2] pada tahun akademik 2014/2015 sampai dengan 2017/2018 menunjukkan bahwa perguruan tinggi swasta di Indonesia menerima rata-rata sebanyak 966.587 mahasiswa baru dan meluluskan rata-rata 691.124 mahasiswa setiap tahunnya atau hanya mencapai 72% saja dari data tersebut diketahui bahwa 28% mahasiswa tidak diketahui keberadaannya. status tersebut bisa jadi karena mahasiswa menempuh studi tidak tepat waktu, memiliki status tidak aktif, atau bahkan *student loses*.

Saat ini, masalah *student loses* dan faktor-faktor yang berkontribusi telah menjadi topik yang menarik untuk diteliti. Institusi perlu mendeteksi kinerja

mahasiswa yang berstatus "*student loses*" sehingga institusi dapat mengidentifikasi faktor-faktor yang berkontribusi terhadap kegagalan dalam menempuh studi. Dan beberapa penyebab *student loses* mahasiswa diantaranya, faktor pembiayaan rendahnya kemampuan akademik, domisili saat menempuh studi dan faktor lainnya. Universitas Pembangunan Pancabudi merupakan salah satu perguruan tinggi swasta di Sumatera Utara yang memiliki 19.881 mahasiswa. Pada tahun ajaran 2014/2015 sampai dengan data 2018/2019 meluluskan 17%, terdapat 21 % atau sekitar 4.191 mahasiswa dengan status tidak aktif. Tingginya persentase mahasiswa yang memiliki status tersebut menyebabkan tingginya persentase mahasiswa lulus tidak tepat waktu. Hal ini menjadi sangat penting bagi Institusi mengingat persentase mahasiswa lulus tepat waktu adalah merupakan unsur penting penilaian akreditasi yang ditetapkan oleh Badan Akreditasi Nasional. Dan Institusi perlu melakukan tindakan untuk mengetahui faktor-faktor penyebab mahasiswa memiliki status *student loses*.

Institusi perlu melakukan prediksi perilaku terhadap mahasiswa dalam mencegah kegagalan akademik secara dini untuk mencegah *student loses*. Database institusi perguruan tinggi menyimpan data akademik, administrasi biodata yang terkait dengan mahasiswa dan Data tersebut apabila digali dengan tepat maka dapat diketahui pola atau *knowledge* untuk mengambil keputusan (El-Halees, 2015). Serangkaian proses mendapatkan *knowledge* atau pola dari data base disebut dengan data mining (Witten, Ian 2016). Data mining dapat membantu mengatasi masalah dengan menganalisis data yang telah ada dalam *database*.

Masalah tersebut akan menjadi topik menarik untuk menjadi pokok pembahasan untuk diteliti. Beberapa penelitian sebelumnya telah menjelaskan beberapa metode untuk mengurangi *student loses*. Diantaranya, Penelitian yang dilakukan oleh (Kelly J. de O. Santos 2019) menyebutkan bahwa sangat penting bagi dosen untuk mendeteksi mahasiswa yang cenderung *loses* sebelum mereka memasuki pertengahan masa studi. Beberapa algoritma klasifikasi data mining telah digunakan untuk memprediksi perilaku mahasiswa yang berpotensi *loses* diantaranya (*Decision Tree, KNN, NN, SVM, NB, RF*). berdasarkan kontribusi

akurasi dengan hasil *average* akurasi 70-72%. Peneliti menyebutkan bahwa monitoring dan dukungan terhadap mahasiswa di tahun pertama sangat penting dilakukan di Federal University of Sergipe (UFS) Brazil. Mahasiswa gagal pada semester 4. Prodi *System Information* dan *Computer Science* dengan menggunakan 23,690 data set mahasiswa,. Pada penelitian selanjutnya, **(Nagy, Marcell 2018)**, melakukan perbandingan untuk memprediksi mahasiswa *student losses* dengan jumlah 15.825 data akademik mahasiswa dari Universitas Teknologi dan Ekonomi Budapest terdaftar antara 2010 sampai dengan 2017 dengan beberapa algoritma diuji dengan 10-fold cross-validation dengan hasil Decision Tree 63%, Random Forest 65%, Generalized Linear Model 67%, Naive Bayes 68.3%, k-NN 69%, Logistic Regression 70.3%, Gradient Boosted Trees 70.6%, Deep Learning 73.5%. Pada penelitian selanjutnya, **(Mayra Alban; Mauricio 2018)** dalam penelitiannya memprediksi mahasiswa yang cenderung *losses* dengan Dataset yang digunakan 3000 data akademik mahasiswa di University Cotopaxi, kota *Latacunga* Ekuador. terdaftar dari 2011 sampai dengan 2016 prediksi *student losses* sesuai dengan evaluasi algoritma yang digunakan (Decision Tree, SVM, Logistic Regression) dan DT paling tinggi akurasi 95%. Pada penelitian selanjutnya, **(Tasnim, Nafisa 2019)** dalam penelitiannya memprediksi mahasiswa yang cenderung *losses* dengan menggunakan Dua dataset terintegrasi dalam “Analisis Kinerja Mahasiswa”. Setiap dataset memiliki 395 instance dan yang lainnya memiliki 649 instance. Kedua dataset memiliki 33 atribut. dataset akademik dari UCI MachineLearning Repository algoritma *Logistic Regression, Naive Bayes, SVM*. Dengan akurasi *average* 83%.

Pada penelitian diatas tidak menjelaskan atribut yang paling berpengaruh terhadap kegagalan mahasiswa, maka penulis mengusulkan penggunaan data akademik dan beberapa atribut yang berkaitan dengan faktor kegagalan diantaranya faktor pembiayaan rendahnya kemampuan akademik, domisili saat menempuh studi dan faktor lainnya dengan metode klasifikasi algoritma C4.5 dalam mengambil keputusan yang dianggap memiliki kinerja sangat baik, tetapi algoritma C4.5 memiliki beberapa kelemahan terhadap data yang tidak seimbang dan tidak lengkap dengan menggunakan pendekatan *Split Feature Reduction*

Model metode ini dianggap bisa meningkatkan akurasi klasifikasi yang terdiri atas 22 atribut dataset akademik mahasiswa, sehingga lebih akurat untuk memprediksi *student loses*. Pada uraian diatas maka penelitian ini diberi judul : ANALISIS KINERJA DALAM MENDETEKSI *STUDENT LOSES* BERDASARKAN NILAI *GAIN* DENGAN *SPLITE FEATURE REDUCTION MODEL* PADA ALGORITMA C4,5.

1.1.1. Masalah Penelitian

Berdasarkan uraian latar belakang di atas, maka masalah dalam penelitian ini dapat diuraikan menjadi dua bagian yaitu identifikasi masalah dan rumusan masalah

1.2.1 Identifikasi Masalah

1. Perguruan tinggi swasta merupakan pilihan kedua bagi para mahasiswa untuk melanjutkan studi tingkat lanjut.
2. Data akademik mahasiswa di perguruan tinggi swasta menjadi salah satu faktor penilaian akreditasi nasional.
3. Tingginya data akademik dengan status Tidak aktif terhadap data terdaftar, dan lulusan menjadi faktor penting bagi institusi untuk diteliti, dan peneliti sebelumnya tidak menjelaskan faktor paling berpengaruh terhadap *student loses*.
4. Algoritma C4,5 mempunyai kelemahan dalam mengklasifikasi data, ketika atribut data tidak seimbang atau data kosong yang menyebabkan error atau akurasi kurang baik, sedangkan penelitian ini menggunakan jumlah 22 atribut dan sub atribut dataset akademik mahasiswa sehingga perlu pendekatan metode tambahan, yaitu Split Feature Reduction Model.

1,2,2. Rumusan Masalah

Rumusan masalah pada penelitian ini adalah sulitnya menentukan faktor yang berkontribusi terhadap kegagalan mahasiswa dalam menempuh studi serta membandingkan dengan pola dari data akademik.

1.3. Tujuan dan Manfaat Penelitian

1.3.1. Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah sebagai berikut :

1. Menganalisis dalam pendeteksian masalah *student loses* dengan pendekatan algoritma C4,5 berdasarkan pohon keputusan dalam peningkatan kinerja dari perspektif akurasi, presisi dan sensitivitas pada validasi pengukuran dataset Mahasiswa Universitas Pembangunan Pancabudi Medan.
2. Menganalisis dalam mengetahui cara pemanfaatan Algoritma *cross validation* dan dapat memberikan solusi bagi ketidakseimbangan data training pada dataset mahasiswa Universitas Pembangunan Pancabudi Medan
3. Untuk mengetahui cara melakukan teknik reduksi data dengan menggunakan *Split Feature Reduction Model*

1.3.2. Manfaat Penelitian

Hasil penelitian ini diharapkan dapat digunakan sebagai referensi alternatif untuk menangani masalah *student loses* dan meminimalisir kerugian institusi baik materil maupun *image building*. Selain itu dapat membangun model alternatif untuk memecahkan masalah dengan meningkatkan angka kelulusan mahasiswa dan diharapkan untuk memberikan kontribusi bagi ilmu pengetahuan masa depan terutama di bidang klasifikasi data.

1.4. Ruang Lingkup

Batasan masalah yang digunakan dalam penelitian ini bertujuan agar pembahasan pada penelitian ini lebih terarah. Adapun batasan masalah pada penelitian ini adalah:

1. Klasifikasi menggunakan algoritma C4.5 dan Split Feature Reduction Model
2. Menggunakan *confusion matriks* untuk mengukur akurasi, *presisi*, dan *recall* dalam menganalisis kinerja klasifikasi.
3. Penjelasan perkiraan kinerja klasifikasi dengan area di bawah ROC dan AUC
4. Menggunakan dataset asli dari sistem informasi mahasiswa Biro akademik, program studi manajemen dan program studi Ilmu Komputer tahun ajaran 2017/2018 Universitas Pembangunan Pancabudi.
5. Analisis dan pengujian data menggunakan Rapid Miner studio 9.3

1.5. Metodologi Penelitian

Metodologi yang digunakan dalam penelitian ini adalah sebagai berikut:

1. Studi Literatur

Pada tahap ini dilakukan pendalaman mengenai bagaimana menangani mahasiswa yang cenderung loses (gagal) dan cara kerja data mining dan perbandingan algoritma serta proses klasifikasi pada beberapa penelitian sebelumnya.

2. Tahap Analisis

Pada tahap ini dilakukan proses untuk mengidentifikasi data yang dibutuhkan masalah dan tantangan yang harus diselesaikan dan menjelaskan solusi yang diusulkan untuk menyelesaikan masalah yang ada.

3. Proses dalam data mining akan digambarkan dalam bentuk gambar *flowchart*.

4. Perancangan Model

Perancangan model dimulai dari mengumpulkan dataset akademik menerapkan preprocessing kemudian diolah dan pemilihan fitur berdasarkan nilai gain tertinggi sampai diperoleh pembobotan dan klasifikasi.

5. Pengujian

Pengujian tingkat akurasi teknik klasifikasi dengan rating serta pengujian dengan confusion matrix untuk memperoleh nilai akurasi, *presisi* dan *recall*.

6. Evaluasi

Pada tahap ini, dilakukan evaluasi terhadap hasil pengujian yang sudah dikerjakan untuk membuat kesimpulan dan saran.

3.5. Sistematika Penulisan

Hasil penelitian ini diharapkan dapat digunakan sebagai referensi alternatif untuk menangani masalah *student loses* dan meminimalisir kerugian institusi baik materil maupun *image building*. Selain itu dapat membangun model alternatif untuk memecahkan masalah dengan meningkatkan angka kelulusan mahasiswa dan diharapkan untuk memberikan kontribusi bagi ilmu pengetahuan masa depan terutama di bidang klasifikasi data.

Adapun Untuk dapat memahami laporan ini, materi yang disajikan dalam laporan Tesis ini dikelompokkan ke dalam beberapa sub-bagian dengan penyampaian sebagai berikut.

BAB I PENDAHULUAN

Pada bab ini berisi Berisi tentang latar belakang, perumusan masalah, tujuan dan manfaat penelitian, ruang lingkup penelitian, dan sistematika penulisan.

BAB II KAJIAN LITERATUR

Bab ini berisikan teori yang berupa pengertian dan definisi yang diambil dari kutipan buku, jurnal internasional yang berkaitan dengan penyusunan laporan tesis serta beberapa literature review yang berhubungan dengan penelitian.

BAB III METODOLOGI PENELITIAN

Bab ini berisikan analisis masalah, metode penelitian, metode pengumpulan data, metode pemilihan sampel, pengolahan data metode yang diusulkan validasi dan evaluasi.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menjelaskan analisis sistem yang diusulkan dengan menggunakan flowchart dan mind map dari sistem yang diimplementasikan, serta pembahasan secara detail final elisitasi yang ada di bab sebelumnya, dijabarkan secara satu persatu dengan menerapkan konsep sesudah adanya sistem yang diusulkan.

BAB V PENUTUP

Bab ini berisi kesimpulan dan saran yang berkaitan dengan analisa dan optimalisasi sistem berdasarkan yang telah diuraikan pada bab-bab sebelumnya.



UNIVERSITAS
MIKROSKIL