BAB I

PENDAHULUAN

1.1 Latar Belakang

Pertumbuhan jumlah data yang tersimpan pada masing-masing organisasi atau institusi saat ini sangat berkembang pesat dan terus menerus perkembang (Göker, Bülbül And Irmak, 2013). Pemanfaatan teknik data mining merupakan langkah konkrit dalam menghadapi pertumbuhan data yang semaking berkembang pesat dengan teknik pengekstraksian data tersembunyi di dalam gudang data(N, P and Sathiaseelan, 2015).

Dili Institute of Technology (DIT) merupakan salah satu perguruan tinggi swasta yang berada di Timor Leste dan telah terdaftar di pengadilan Dili, serta mendapatkan pengakuan secara Internasional dengan persentasi akreditasi kampus 92.88% yang telah dilakukan oleh kementrian Pendidikan Timor Leste ((ANAAA), 2017), (Tatoli Agência Nacional Timor Leste, 2017). DIT juga memberikan perhatian khusus pada kualitas belajar-mengajar, penelitian dan pengabdian pada masyarakat. Selain itu memberikan pelatihan untuk pendidikan kejuruan di antaranya adalah pelatihan di bidang *auto mechanic, administration and finance, civil construction, certificate IV in training and assessment, hause keeping, food and beverage, front office, tour and travel* serta *library and information service* (DIT, 2011).

Berdasarkan salah satu visi dari DIT bahwa pada tahun 2020 institutsi ini akan diakui secara internasional sebagai lembaga professional yang unggul untuk

pendidikan dan pelatihan dan juga sebagai penunjuk jalan untuk sumber daya yang inovatif dan kreatif bagi masyarakat yang dilayaninya. Kemudian berdasarkan *strategic plan* DIT 2008-2020 dimana pada bagian pengajaran dan pembelajaran untuk menyediakan pendidikan dan pelatihan kejuruan dengan kualitas yang tinggi yang diakui secara nasional dan internasional (DIT, 2008).

Dili Institute of Technology merupakan salah satu perguruan tinggi swasta di Timor-Leste yang memiliki jumlah mahasiswa yang sangat banyak yang masih terkendala dalam mengontrol lama studi setiap mahasiswanya. Hal ini terlihat dari peningkatan jumlah calon mahasiswa baru pada setiap tahun ajaran baru sehingga pencatatan data mahasiswapun semakin menumpuk. Selain itu, masih ada pula mahasiswa yang lulus dengan lama masa studi melampaui waktu yang telah ditetapkan dengan perolehan Indeks Prestasi Kumulatif (IPK) yang rendah. Faktor-faktor yang dapat mempengaruhi lama studi dan perolehan IPK mahasiswa dapat berasal dari prestasi akademik mahasiswa maupun nonakademik.

Seiring dengan terus bertambahnya jumlah mahasiswa di (DIT) yang begitu banyak sehingga terjadi penumpukan data yang belum diolah dengan optimal untuk menggali informasi dan pengetahuan baru melalui pola-pola yang terbentuk dari penumpukan data tersebut. Dengan demikian perlu adanya suatu teknik atau metode untuk mengolah data tersebut menjadi sebuah informasi dan pengetahuan yang dapat digunakan sebagai bahan pertimbangan pimpinan dalam proses pengambilan kebijakan dan keputusan juga sebagai peringatan dini *(early warning)* bagi mahasiswa tertentu yang berdasarkan hasil prediksi dinyatakan berpotensi lulus

dengan melampaui ketentuan lama studi ataupun berpotensi lulus dengan perolehan IPK yang rendah.

Berkaitan dengan kendala yang ada, maka penulis melakukan penelitian yang berkaitan dengan penerapan data mining untuk prediksi lama masa studi mahasiswa pada Dili Institute of Technology. Beberapa algoritma klasifikasi data mining telah digunakan untuk memprediksi lama masa studi mahasiswa adalah *Neural Network, Naïve Bayes, Instance-Based Learning, Logistic Regression* dan *Support Vector Machine*. Hasilnya, *Naïve Bayes* menunjukkan hasil yang paling akurat. (Hastuti, 2012). Teknik yang dapat digunakan untuk menemukan pola guna mendapatkan informasi dan pengetahuan baru dari kumpulan data yang tersedia adalah dengan teknik Data Mining, (Witten, 2011).

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, dapat dirumuskan permasalahanya yaitu: Bagaimana kinerja kedua metode yaitu, *Naïve Bayes* dan *Adaboost* untuk prediksi lama studi mahasiswa?

1.3 Batasan Masalah

Agar pembahasan masalah ini dapat lebih terarah, maka perlu adanya suatu batasan masalah, yaitu:

a. Dalam penelitian ini hanya dibatasi masalah yang terkait dengan metode *Naïve Bayes* dan *Adaboost* menggunakan teknik *Data Mining* dengan cara menganalisis sejumlah atribut yang menjadi parameter dalam

prediksi lama studi yaitu: pendapatan orang tua, Asal sekolah, Nilai tes seleksi, Jenis kelamin, jumlah SKS yang sudah ditempuh dari semester satu hingga semester empat, IP semester satu sampai dengan semester empat dan IPK yang dinyatakan lulus dan digunakan sebagai data latih dan data uji.

b. Penelitian ini hanya dilakukan di kampus Dili Institute of Technology (DIT)

c. Data yang diproses dari angkatan tahun 2010 sampai dengan 2014.

1.4 Tujuan Penelitian

Dalam penelitian ini, Tujuan yang ingin dicapai yaitu:

a. Dapat memprediksi lama studi dari setiap mahasiswa atas dasar data kelulusan sebelumnya.

b. Hasilnya dapat digunakan sebagai bahan pertimbangan pimpinan dalam proses pengambilan kebijakan dan keputusan sebagai peringatan dini *(early warning)* bagi mahasiswa yang berdasarkan hasil prediksi dinyatakan berpotensi lulus dengan melampaui ketentuan lama studinya.

1.5 Manfaat Penelitian

Adapun manfaat yang diperoleh dari penelitian ini adalah:

a. Dari hasil prediksi yang dilakukan oleh peneliti maka, pihak DIT dapat mengetahui jumlah mahasiswa tertentu yang berdasarkan hasil

prediksinya dinyatakan berpotensi lulus dengan melampaui ketentuan lama studi yang ditentukan.

b. Bagi program studi dapat melakukan tindakan preventif yang berkaitan dengan lama studi setiap mahasiswa tersebut.

1.6 Sistematika Penulisan

Penulisan Tesis ini dibagi dalam enam bagian, yang terdiri dari:

BAB I: Pendahuluan:

Bab ini berisi latar belakang masalah, rumusan masalah, batasan masalah, tujuan, metode yang digunakan, serta sistematika penulian yang digunakan dalam penyusunan Tesis ini.

BAB II: Tinjauan Pustaka:

Tinjauan pustaka memuat uraian sistematis tentang hasil-hasil penelitian yang didapat oleh peneliti terdahulu serta pustaka yang ada hubungannya dengan penelitian yang akan dilakukan.

BAB III: Dasar Teori

Berisi teori dan rumusan awal yang dari berbagai sumber yang tentang Data mining, Klasifikasi, Prediksi, *Naïve Bayes* dan *Adaboost.*

BAB IV: Metodologi Penelitian:

Metodologi penelitian pada umumnya mengandung uraian tentang: bahan atau materi, langkah-langkah penelitian, data yang akan dikumpulkan, dan metoda analisis hasil.

BAB V: Hasil dan Pembahasan:

Bab ini menguraikan hasil analisis dalam menyelesaikan proses klasifikasi pada himpunan data yang digunakan serta hasil pengujian terhadap perangkat lunak.

BAB VI: Kesimpulan dan Saran:

Berisi kesimpulan dari keseluruhan hasil yang dibuat, metode yang digunakan serta memberikan saran tentang kelebihan dan kekurangan dari metode tersebut.