

BAB I

PENDAHULUAN

1.1 Latar Belakang

Universitas Pendidikan Ganesha (Undiksha) merupakan universitas negeri yang berada di Kota Singaraja, Bali. Undiksha menerapkan 3 jalur penerimaan mahasiswa baru, yaitu jalur Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN), Seleksi Bersama Masuk Perguruan Tinggi Negeri (SBMPTN), dan Seleksi Mahasiswa Baru Jalur Mandiri (SMBJM) (Pratama dkk. 2020).

Berdasarkan penelitian Pratama (2020), masing-masing jalur yang disediakan oleh Undiksha memiliki kuota. Setiap kuota yang telah disediakan harus terpenuhi untuk mencapai banyaknya mahasiswa yang ideal. Hal ini memerlukan strategi dalam meluluskan calon mahasiswa baru. Mahasiswa yang lolos seleksi sebaiknya dapat diprediksi akan melakukan pendaftaran kembali sehingga dapat memenuhi kuota yang telah ditentukan. Hal ini pun sering menjadi masalah yang dialami oleh pihak panitia penerimaan mahasiswa baru. Misalnya setelah melakukan tahap daftar ulang, ternyata kuota tidak terpenuhi. Undiksha telah menangani masalah ini dengan cara melihat dari persentase daftar kembali tahun sebelumnya untuk mempertimbangkan kelulusan berikutnya.

Penelitian tersebut juga membangun sistem cerdas yang dapat memprediksi calon mahasiswa baru yang kemungkinan besar akan melakukan daftar kembali. Penelitian yang berjudul “Pengembangan Sistem Cerdas untuk Prediksi Daftar

Kembali Penerimaan Mahasiswa Baru dengan Metode *Naive Bayes* (Studi Kasus: Universitas Pendidikan Ganesha)”. Metode *Naive Bayes* digunakan untuk menentukan probabilitas daftar kembali mahasiswa baru. Berdasarkan jumlah mahasiswa yang mendaftar, dapat ditentukan jumlah mahasiswa baru yang diluluskan dan akan melakukan pendaftaran kembali. Hasil penelitian ini menunjukkan bahwa sistem dapat memprediksi dengan akurasi yang 71%. (Pratama dkk. 2020)

Namun dalam penelitiannya, diberikan saran kepada pengembang selanjutnya. Sarannya adalah menerapkan metode *miner* yang berbeda dan lebih efektif, memilih atribut data yang lebih relevan agar sistem dapat memperoleh tingkat akurasi yang lebih tinggi. (Pratama dkk. 2020)

Bedasarkan saran dari peneliti dengan angka akurasi sistem yang 71%, maka terdapat kemungkinan bahwa sistem memperoleh akurasi yang berbeda. Hal ini disebabkan bilamana menerapkan algoritma lain. Algoritma lain yang akan diterapkan sebagai pembanding algoritma sebelumnya yaitu metode *C 4.5* dan *C 4.5* dengan *Adaptive Boosting (Adaboost)*. Untuk mendukung pemilihan algoritma ini terdapat beberapa penelitian terkait.

Penelitian pertama adalah penelitian yang dilakukan oleh (Agustin dkk. 2019). Penelitian yang dilakukan adalah memprediksi calon mahasiswa baru pada STIMIK XYZ yang akan lulus tepat waktu atau terlambat, sehingga dapat dijadikan dasar untuk melakukan seleksi pada mahasiswa tersebut. Penelitian ini menggunakan metode *data mining Decision Tree* yaitu *C 4.5* dengan tambahan *Adaptive Boosting (Adaboost)*. Tujuan lain yaitu mencari dan membandingkan

akurasi kerja algoritma *C 4.5* tanpa dan dengan *Adaboost*. Dari penelitian yang telah dilakukan diperoleh nilai akurasi yang sama antara Algoritma *C 4.5* dan Algoritma *C 4.5* berbasis *Adaboost* yaitu sebesar *Precision* 77.33%, *Accuracy* 90.28%, *Recall* 45.54% akan tetapi terjadi perbedaan pada nilai AUC untuk Algoritma *C 4.5* sebesar 0.683 sedangkan untuk Algoritma *C 4.5* berbasis *Adaboost* sebesar 0,717. Pola tersebut dapat membantu untuk mengambil keputusan penerimaan mahasiswa baru yang dapat lulus tepat waktu dan mahasiswa yang lulus terlambat dapat terprediksi lebih awal.

Penelitian selanjutnya adalah penelitian yang dilakukan oleh (Prianto & Harani, 2019). Penelitian ini dilakukan karena tidak seimbangnya antara mahasiswa yang lulus sesuai kriteria dengan mahasiswa yang melakukan pendaftaran ulang di Universitas Semarang. Proses penerimaan mahasiswa baru sebelumnya sudah menggunakan proses *Decision Tree* namun masih memiliki akurasi hanya 61,4%. Penelitian ini dilakukan dengan cara menambahkan *Adaptive Boosting* pada proses pembuatan *Decision Tree* menggunakan *C 4.5*. Hasil dari penelitian ini yaitu pohon keputusan berbasis *Adaboost* menunjukkan tingkat akurasi yang meningkat sebesar 20%. Penyajian hasil sebagai berikut, 61,4% (*Decision Tree*); 91,35% (*Decision Tree* dengan *Adaboost*).

Penelitian selanjutnya yaitu penelitian yang dilakukan oleh (Wong dkk. 2019). Penelitian ini bermaksud untuk membandingkan akurasi dari algoritma *C 4.5* dan *Classification and Regression Tree* (CART) dalam proses penerimaan calon karyawan pada perusahaan. Penelitian ini menggunakan dataset dengan kriteria umur, lama pengalaman kerja, pendidikan terakhir, status pernikahan, jumlah

kemampuan yang dimiliki, serta nilai tes seleksi masuk. Pengujian menggunakan 200 data seleksi calon karyawan secara manual dari suatu perusahaan. Pengujian dilakukan dengan menggunakan *K-Fold Cross Validation* dan menghitung keakuratan algoritma dengan *Confusion matrix*. Algoritma *C 4.5* memiliki tingkat akurasi, tingkat keberhasilan sistem, dan tingkat ketepatan hasil keputusan sebesar 52,83%, 41,48%, dan 43,98%, sedangkan algoritma CART sebesar 53,33%, 44,06%, dan 42,81%.

Penelitian selanjutnya adalah penelitian yang dilakukan oleh (Rabbani dkk. 2020). Penelitian ini dilakukan untuk memprediksi calon mahasiswa yang akan melakukan proses daftar kembali. Penelitian ini bermaksud membandingkan metode datamining jenis *Decision Tree* menggunakan metode *Adaboost* dengan metode *Decision Tree* saja. Untuk memaksimalkan performa model, maka dilakukan analisa variabel dengan menggunakan *chi square* dalam proses *feature selection*-nya. Hasil akhir menunjukkan bahwa model prediksi *Adaboost* memiliki performa yang lebih baik daripada model *Decision Tree* dengan skor *f-measure* 90.9%, *precision* 83.7% dan *recall* 99.5%. Selain itu didapatkan juga ciri dari kandidat yang cenderung melanjutkan pendaftaran hingga akhir. Sehingga dengan hasil tersebut bisa membantu pihak perguruan tinggi dalam mengambil keputusan dalam proses seleksi calon mahasiswa baru.

Penelitian selanjutnya adalah penelitian yang dilakukan oleh (Debiyanti dkk. 2020). Penelitian ini bertujuan untuk memprediksi kondisi perusahaan terhadap kemampuan dalam pengelolaan dan mempertahankan stabilitas kinerja keuangan. Penelitian ini menggunakan *Decision Tree* yaitu dengan *C 4.5* dengan

menggunakan *Adaptive Boosting*. Jumlah data awal yang digunakan adalah 755 *records* namun setelah dilakukan proses KDD data menjadi 746 *records*. Berdasarkan pengujian yang telah dilakukan dengan rasio data latih dan data uji 90%:10% hasil akurasi dari Algoritma *C 4.5* adalah 72,97% dan setelah ditambah dengan *Adaboost* akurasi meningkat menjadi 86,49%. Dengan demikian dapat disimpulkan bahwa Algoritma *C 4.5* dan *Adaboost* baik digunakan dalam prediksi *financial distress* perusahaan.

Berdasarkan latar belakang tersebut, dilakukan penelitian yang berjudul Analisis Perbandingan Algoritma *C 4.5* berbasis *Adaboost* Dalam Memprediksi Daftar Kembali Mahasiswa Baru (Studi Kasus: Universitas Pendidikan Ganesha). Penelitian ini mengukur kinerja dari algoritma *C 4.5* dan *C 4.5* berbasis *Adaboost*, dalam memprediksi calon mahasiswa baru yang akan melakukan pendaftaran kembali.

1.2 Identifikasi Masalah

Identifikasi masalah untuk penelitian yang berjudul Analisis Perbandingan Algoritma *C 4.5* dengan *C 4.5* Berbasis *Adaboost* dalam Memprediksi Daftar Kembali Mahasiswa Baru (Studi Kasus: Universitas Pendidikan Ganesha) adalah sebagai berikut.

1. Prediksi calon mahasiswa baru yang melakukan daftar ulang, masih sulit.
2. Penelitian untuk tujuan di atas sudah dilakukan namun hasil akurasinya masih cukup rendah yaitu 71%.
3. Terdapat beberapa atribut pada proses prediksi yang tidak digunakan pada penelitian sebelumnya.

4. Penelitian sebelumnya menyarankan untuk menerapkan metode lain dan menggunakan atribut yang lebih relevan untuk mendapatkan nilai akurasi yang lebih tinggi.

1.3 Batasan Masalah

Batasan masalah untuk penelitian yang berjudul Analisis Perbandingan Algoritma $C 4.5$ dengan $C 4.5$ Berbasis *Adaboost* dalam Memprediksi Daftar Kembali Mahasiswa Baru (Studi Kasus: Universitas Pendidikan Ganesha) adalah sebagai berikut.

1. Penelitian ini hanya menerapkan teknik *data mining* menggunakan perhitungan saja tanpa mengubah sistem cerdas yang telah dibuat sebelumnya.
2. Proses perhitungan menggunakan *tools Rapid Miner*.
3. Data berasal dari jalur SNMPTN dan SBMPTN tahun 2018 hingga 2020. Jalur SMBJM tidak digunakan. Hal ini disebabkan oleh proses seleksi SMBJM memiliki proses, kriteria dan strategi yang berbeda dengan jalur SNMPTN dan SBMPTN.

1.4 Rumusan Masalah

Berdasarkan latar belakang yang dipaparkan sebelumnya, dapat dirumuskan beberapa permasalahan yang akan dibahas dalam penelitian ini, adalah sebagai berikut.

1. Bagaimana implementasi algoritma $C 4.5$ dan $C 4.5$ berbasis *Adaboost* dalam memprediksi mahasiswa yang melakukan pendaftaran kembali?

2. Bagaimana perbandingan akurasi dari teknik *data mining C 4.5* dengan *C 4.5* berbasis *Adaboost* dalam memprediksi calon mahasiswa baru yang melakukan pendaftaran kembali?

1.5 Tujuan Penelitian

Adapun tujuan yang diharapkan dapat tercapai dari penelitian Analisis Perbandingan Algoritma *C 4.5* dengan *C 4.5* Berbasis *Adaboost* dalam Memprediksi Daftar Kembali Mahasiswa Baru (Studi Kasus: Universitas Pendidikan Ganesha) adalah sebagai berikut.

1. Untuk mengetahui implementasi algoritma *C 4.5* berbasis *Adaboost* dalam memprediksi mahasiswa yang melakukan pendaftaran kembali.
2. Untuk mengetahui perbandingan akurasi dari teknik *data mining C 4.5* dengan atau tanpa *Adaboost* dalam memprediksi mahasiswa yang melakukan pendaftaran kembali.

1.6 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian Analisis Perbandingan Algoritma *C 4.5* dengan *C 4.5* Berbasis *Adaboost* dalam Memprediksi Daftar Kembali Mahasiswa Baru (Studi Kasus: Universitas Pendidikan Ganesha) adalah sebagai berikut.

a. Manfaat Teoretis

1. Bagi peneliti, penelitian ini diharapkan akan mampu menambah pengetahuan, dapat memahami serta dapat menerapkannya di dalam proses perkuliahan.
2. Bagi penelitian sejenis, penelitian ini diharapkan dapat menjadi kajian dalam penelitian sejenis ini.

b. Manfaat Praktis

1. Bagi Peneliti, dapat mengimplementasikan ilmu yang telah didapat selama perkuliahan dalam Analisis Perbandingan Algoritma $C 4.5$ dengan $C 4.5$ Berbasis *Adaboost* Dalam Memprediksi Daftar Kembali Mahasiswa Baru (Studi Kasus: Universitas Pendidikan Ganesha) dan dapat menambah wawasan peneliti tentang perbandingan metode data mining dalam kasus ini.
2. Bagi Undiksha, penelitian ini memiliki manfaat sebagai evaluasi untuk sistem penerimaan mahasiswa baru.

