

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perguruan tinggi merupakan penyelenggara pendidikan akademik bagi mahasiswa (Kadafi, 2020). Kualitas dari perguruan tinggi dapat ditentukan dari persentase tingkat ketuntasan belajar mahasiswa selama masa studinya. Mahasiswa menjadi tolak ukur yang digunakan untuk menilai kualitas dan mengevaluasi rencana pembelajaran perguruan tinggi. Jumlah mahasiswa yang terus meningkat disetiap tahunnya, menyebabkan tingkat kelulusan dan daya tampung menjadi bagian yang penting dalam penilaian kelulusan dan bahan pengambilan keputusan. Peraturan Menteri Ristek Dikti Nomor 44 tahun 2015 menyebutkan bahwa standar proses pembelajaran maksimum dalam masa studi adalah 7 tahun, IPK di atas 2.0, minimum SKS adalah 144 untuk program Sarjana (Rahman dan Wartulas, 2020). Kriteria ketepatan waktu lulus setiap program pada jenjang perguruan tinggi berbeda-beda.

Universitas Tanjungpura (Untan) adalah salah satu perguruan tinggi negeri yang berada di Kota Pontianak, Kalimantan Barat. Untan terdiri dari sembilan fakultas dengan 96 program studi. Salah satu program studi yang ada di Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) adalah Program Studi Statistika. Berdasarkan buku pedoman akademik FMIPA Untan tahun 2021, evaluasi keberhasilan studi mahasiswa dilakukan pada setiap akhir semester, pada akhir empat semester pertama, pada akhir delapan semester, dan pada akhir program pendidikan. Mahasiswa program S1 (Sarjana) FMIPA Untan dikatakan lulus tepat waktu jika dapat menyelesaikan studi kurang dari atau sama dengan empat tahun dengan beban belajar minimal 144 sks. Mahasiswa yang berpotensi lulus tidak tepat waktu dapat dianalisis pada evaluasi keberhasilan mahasiswa pada akhir empat semester pertama.

Pada Program Studi Statistika masih banyak ditemukan mahasiswa yang tidak lulus tepat waktu sehingga akan berpengaruh pada penilaian akreditasi. Ada beberapa faktor yang mempengaruhi masa studi mahasiswa yaitu IPK, jenis

kelamin, asal daerah, jurusan kuliah, asal sekolah, dan pekerjaan orang tua (Chair, Nasution, dan Rizki, 2017). Oleh karena itu, karakteristik mahasiswa yang lulus tepat waktu maupun tidak tepat waktu dalam penentuan kelulusan mahasiswa dapat dianalisis menggunakan teknik dalam *data mining*.

Data mining adalah proses menemukan informasi atau pola menggunakan teknik statistik yang diperoleh dengan cara mengekstrak dan mengidentifikasi pola yang berguna serta menarik dari berbagai *database* yang besar. Klasifikasi merupakan metode yang sering digunakan dalam *data mining*. Metode klasifikasi yang banyak digunakan oleh peneliti adalah *Decision Tree*. Algoritma C5.0 adalah salah satu algoritma klasifikasi dalam *data mining* yang khususnya diterapkan pada *Decision Tree* (Umma, Warsito, dan Maruddani, 2021). Algoritma C5.0 lebih baik daripada C4.5 pada kecepatan, memori, dan efisiensi (Pratiwi, Hayati, dan Prangga, 2020). *Decision Tree* dapat menangani *over-fitting*, menangani atribut yang kontinu, memilih yang tepat untuk *attribute selection*, menangani *training data* dengan nilai atribut yang hilang, dan meningkatkan efisiensi komputasi (Quinlan, 1993). Tetapi Metode *Decision Tree* memiliki kelemahan dalam derajat yang tinggi dari ketidakseimbangan kelas (*class imbalance*).

Distribusi *class imbalance* dapat ditandai sebagai sesuatu yang memiliki lebih banyak kasus dari beberapa kelas yang lain, masalah keseimbangan adalah salah satu dimana satu *class* diwakili oleh sampel yang besar, sedangkan yang lainnya hanya diwakili oleh beberapa sampel (Bisri dan Wahono, 2015). Ketidakseimbangan kelas mempengaruhi kinerja klasifikasi, sehingga diperlukan metode yang dapat mengatasi masalah *Decision Tree* dengan kasus *class imbalance* agar dapat mencapai dan meningkatkan kinerja klasifikasi *Decision Tree* yang lebih baik. *Boosting* merupakan metode *ensemble* yang paling populer dalam menangani ketidakseimbangan kelas. Algoritma *Boosting* memberikan bobot pada distribusi *training data* dalam setiap iterasi dengan nilai yang berbeda. Metode *Adaptive Boosting (Adaboost)* adalah salah satu algoritma *supervised* dalam *data mining* yang digunakan untuk membangun model klasifikasi.

Oleh karena itu, dilakukan penelitian untuk mengatasi masalah ketidakseimbangan kelas pada penentuan kelulusan mahasiswa Program Studi Statistika Untan, sehingga dapat meningkatkan kinerja klasifikasi *Decision Tree* dan menghasilkan kinerja yang lebih baik. Pada penelitian untuk klasifikasi menggunakan *Decision Tree*, sedangkan untuk mengatasi masalah ketidakseimbangan kelas menggunakan *Adaboost*.

1.2 Rumusan Masalah

Berdasarkan uraian pada latar belakang, maka rumusan masalah dalam penelitian ini yaitu:

1. Bagaimana penerapan Algoritma *Adaboost* pada Metode *Decision Tree* dalam klasifikasi kelulusan mahasiswa.
2. Bagaimana akurasi Algoritma *Adaboost* pada Metode *Decision Tree* dalam klasifikasi kelulusan mahasiswa.

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dibuat maka tujuan yang ingin dicapai melalui penelitian ini yaitu:

1. Untuk menerapkan Algoritma *Adaboost* pada Metode *Decision Tree* dalam klasifikasi kelulusan mahasiswa.
2. Untuk menentukan akurasi Algoritma *Adaboost* pada Metode *Decision Tree* dalam klasifikasi kelulusan mahasiswa.

1.4 Batasan Masalah

Adapun batasan masalah dalam penelitian ini yaitu:

1. Data yang digunakan dalam penelitian adalah data kelulusan mahasiswa Program Studi Statistika Universitas Tanjungpura Periode I Tahun Ajaran 2017/2018 sampai Periode II Tahun Ajaran 2022/2023.
2. Atribut dependen yang digunakan adalah Status kelulusan mahasiswa (Y), sedangkan atribut independennya yaitu jenis kelamin (X_1), IPK semester 1 (X_2), IPK semester 2 (X_3), IPK semester 3 (X_4), IPK

- semester 4 (X_5), daerah asal domisili (X_6), akreditasi sekolah (X_7), jalur masuk (X_8), beasiswa (X_9), dan status kelulusan tes TUTEP pertama (X_{10}).
3. Ketidakseimbangan kelas dikategorikan menjadi tiga yaitu tingkat ketidakseimbangan ringan dengan proporsi kelas minoritas (dalam kasus ini adalah status kelulusan “Tepat Waktu”) adalah 20 - 40% dari *dataset*, tingkat ketidakseimbangan sedang dengan proporsi kelas minoritas adalah 1 - 20% dari *dataset*, dan tingkat ketidakseimbangan ekstrem dengan proporsi kelas minoritas adalah $< 1\%$ dari *dataset*.
 4. Metode *Decision Tree* yang digunakan adalah Algoritma C5.0.
 5. Jumlah iterasi maksimum adalah 100.

1.5 Tinjauan Pustaka

Bisri dan Wahono (2015) melakukan penelitian tentang penerapan *Adaboost* untuk penyelesaian ketidakseimbangan kelas pada penentuan kelulusan mahasiswa dengan Metode *Decision tree*. Penelitian dilakukan dengan menerapkan Metode *Adaboost* pada *Decision Tree* dengan atribut yang digunakan adalah Nomor Induk Mahasiswa (NIM), Prodi, Jenis Kelamin, Shift waktu perkuliahan, IPS dari semester satu hingga semester delapan, IPK, dan Status Kelulusan. Berdasarkan hasil penelitian, diperoleh nilai akurasi sebesar 87,18%, RMSE sebesar 0,320, dan AUC sebesar 0,864 dengan menggunakan Metode *Decision Tree*. Sedangkan hasil dari *Decision Tree* dengan *Adaboost* (DTboost) diperoleh nilai akurasi sebesar 90,45%, RMSE sebesar 0,273, dan AUC sebesar 0,951. Pada kasus ini dapat disimpulkan bahwa Metode *Decision Tree* dengan *Adaboost* terbukti mampu menyelesaikan masalah ketidakseimbangan kelas dalam penentuan kelulusan mahasiswa dengan meningkatkan akurasi dan dapat menurunkan tingkat *error* klasifikasi.

Prianti, Santoso, dan Hakim (2020) melakukan penelitian dengan membandingkan metode *K-Nearest Neighbor* (KNN) dan *Adaptive Boosting* (*Adaboost*) pada kasus klasifikasi multi kelas. Data yang digunakan adalah laporan keuangan dari 575 perusahaan yang tercatat di Bursa Efek Indonesia (BEI) tahun 2019. Variabel-variabel yang digunakan adalah kinerja perusahaan,

Return on Assets (ROA), Return on Equity (ROE), Working Capital on Total Assets (WCTA), Total Assets Turn Over (TATO), Debt to Equity Ratio (DER), Longterm Debt to Assets Ratio (LDAR), dan Return on Investment (ROI). Diperoleh akurasi menggunakan metode KNN dengan $k = 5$ sebesar 0,82087 atau 82,087%. Sedangkan dengan menggunakan *Adaboost* diperoleh akurasi sebesar 0,84522 atau 84,522%. Sehingga dapat disimpulkan bahwa untuk memprediksi kinerja perusahaan di Indonesia tahun 2019 sebaiknya menggunakan Metode *Adaboost*.

Penelitian menggunakan *Decision Tree* dan *Adaboost* untuk mengklasifikasikan penerima program bantuan sosial dilakukan oleh Qadrini, Seppewali, dan Aina tahun 2021. Data yang digunakan adalah data penduduk desa di Kelurahan Banggae yang mendapat Bantuan Langsung Tunai (BLT). Variabel yang digunakan yaitu pendidikan, luas lantai, jenis lantai, jenis dinding, jenis penerangan, bahan bakar masak, sumber air minum, jenis jamban, pendapatan, dan pekerjaan. Berdasarkan hasil penelitian, diperoleh bahwa akurasi dengan Metode *Decision Tree* sebesar 94,17%. Sedangkan setelah digabungkan dengan *Adaboost* akurasinya menjadi sebesar 95%.

Dalam penelitian Novianti, Zarlis, dan Sihombing tahun 2022 dilakukan penerapan Algoritma *Adaboost* untuk peningkatan kinerja klasifikasi data *mining* pada *imbalance dataset* diabetes. Terdapat sepuluh atribut dan dua target kelas pada penyakit diabetes. Klasifikasi dilakukan menggunakan algoritma *K-Nearest Neighbor* (KNN) dan memiliki *dataset* sebanyak 1.001 kasus dengan tiga hasil klasifikasi tidak seimbang. Berdasarkan penelitian terbukti bahwa *Adaboost* membantu meningkatkan kinerja Algoritma KNN untuk klasifikasi pada penyakit diabetes. Terbukti dari lima pengujian untuk nilai $K = 7, 13, 19, 25, \text{ dan } 31$ terdapat peningkatan nilai akurasi setelah menggunakan *Adaboost*.

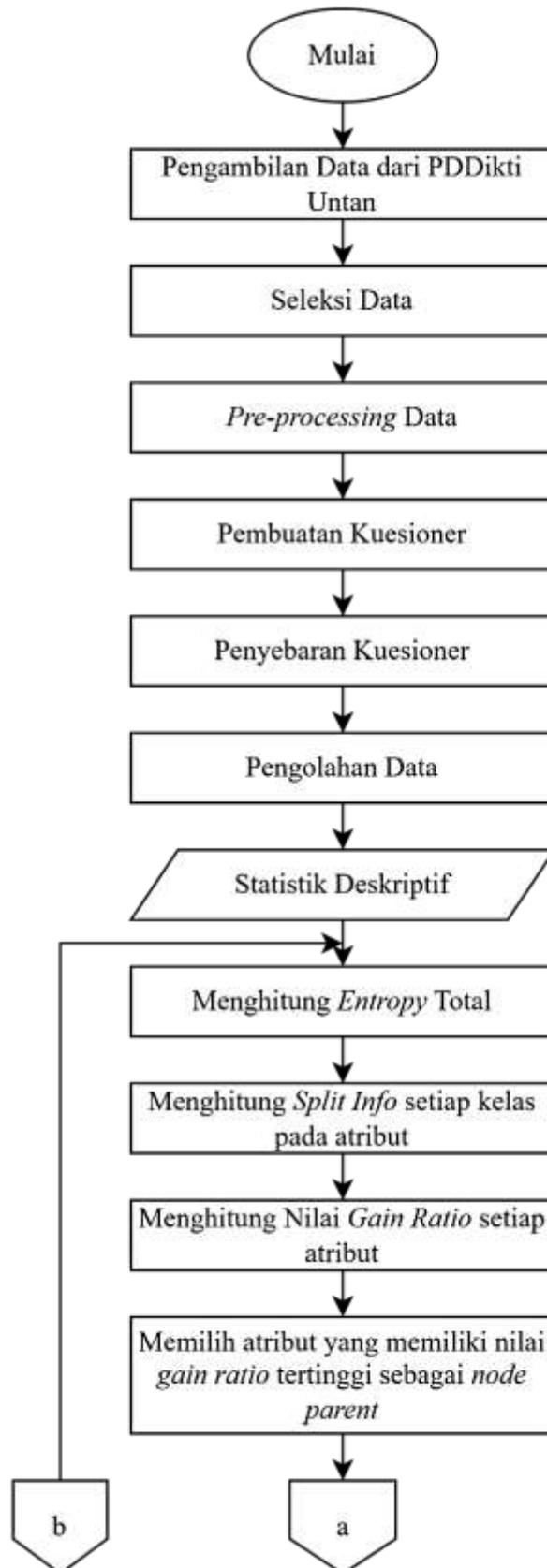
1.6 Metodologi Penelitian

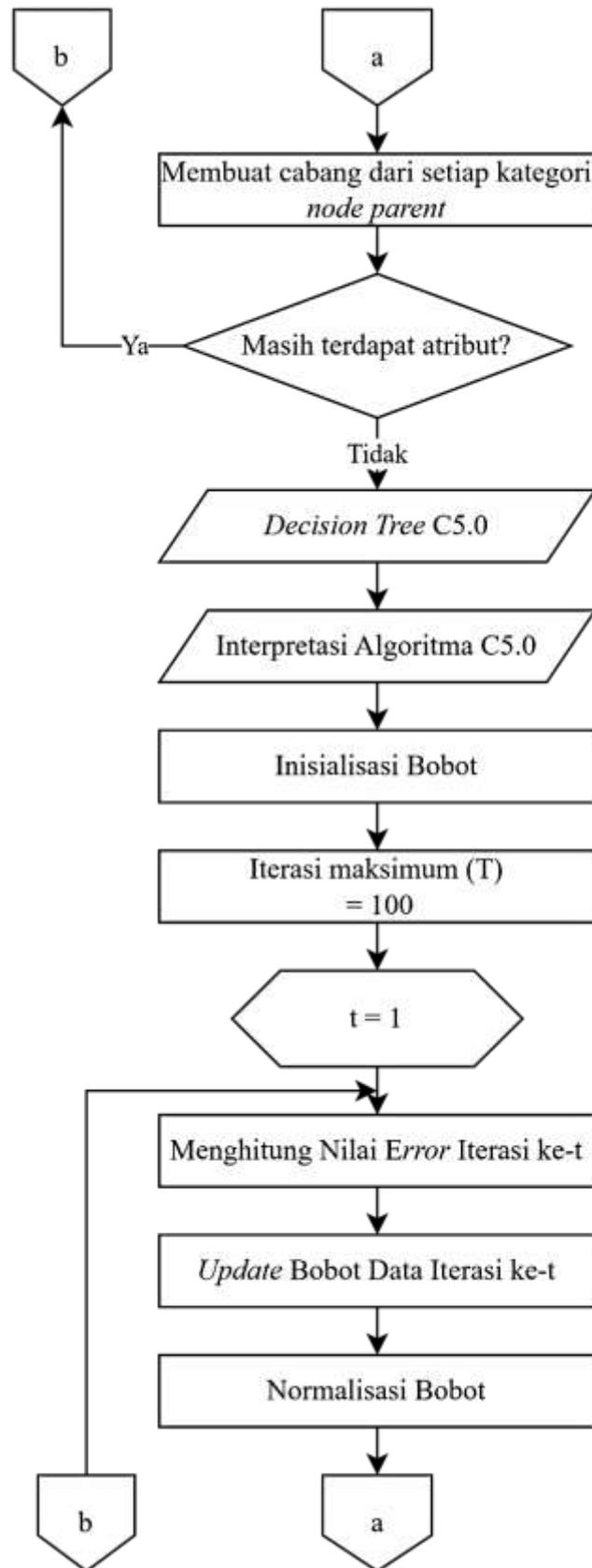
Tahapan-tahapan dalam penelitian ini yaitu melakukan pengambilan data dari *website* PDDikti Untan. Kemudian dilakukan seleksi data, dimana data yang diambil adalah data alumni mahasiswa Program Studi Statistika Untan. Setelah

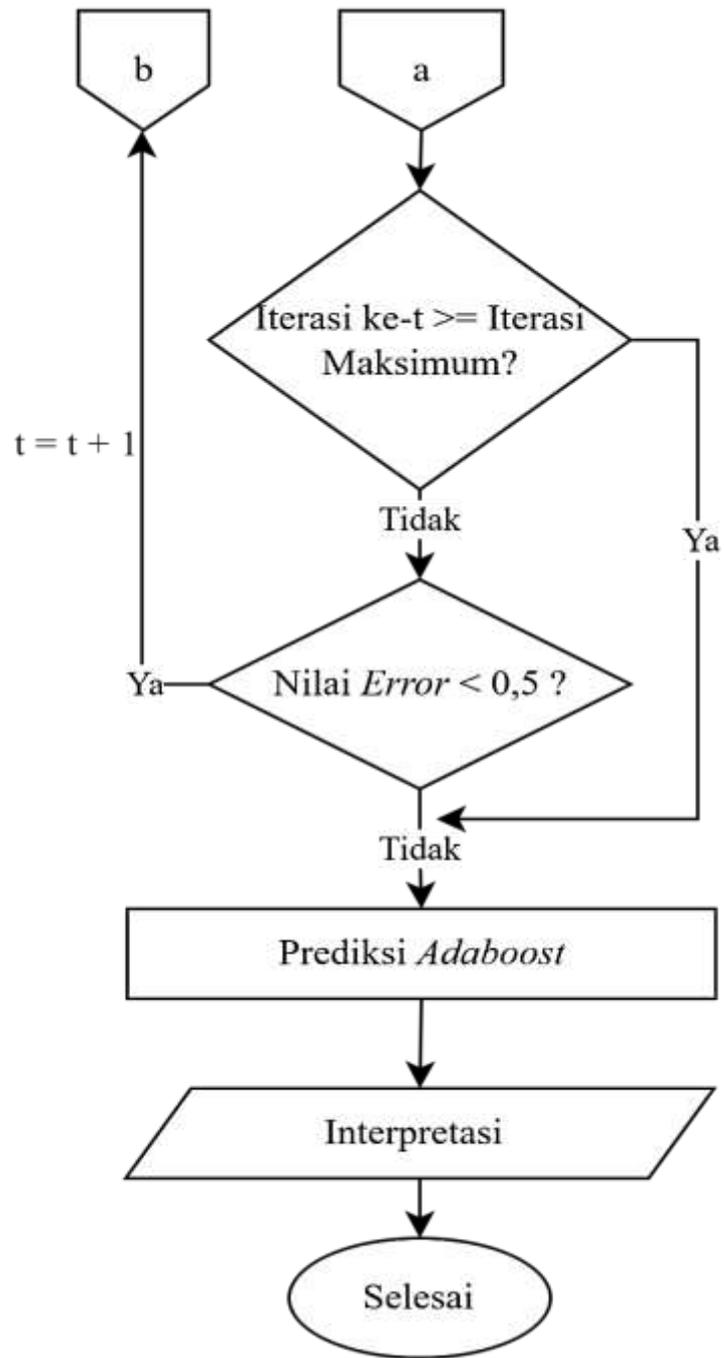
itu dilakukan *pre-processing* data dengan memilah data menjadi sebuah *dataset* baru yang berisi data-data yang diperlukan saja. Data yang diambil dari *website* PDDikti Untan adalah IPK semester 1 hingga IPK semester 4.

Proses selanjutnya adalah pembuatan kuesioner yang tersaji pada Lampiran III. Selanjutnya kuesioner yang telah dibuat disebarakan kepada alumni mahasiswa Program Studi Statistika Untan. Setelah itu data-data yang telah terkumpul dari *website* PDDikti Untan dan kuesioner diolah dengan membuat statistik deskriptif dari data, menghitung nilai *entropy* total, menghitung *split info* setiap kelas pada variabel, menghitung nilai *gain ratio* setiap atribut, memilih atribut yang memiliki nilai *gain ratio* tertinggi sebagai *node parent*, dan membuat cabang dari setiap kategori *node parent*. Lalu diperiksa apakah masih terdapat atribut yang tersisa. Jika kondisi Ya, maka ulangi proses dari menghitung nilai *entropy*. Jika kondisi Tidak, maka dilanjutkan dengan membangun *Decision Tree* Algoritma C5.0 dan melakukan interpretasi hasil dari *Decision Tree* yang telah dibangun serta menghitung nilai akurasi menggunakan *confusion matrix*.

Adapun langkah-langkah dalam Algoritma *Adaboost* adalah melakukan inisialisasi bobot dan menetapkan iterasi maksimum sebanyak 100 kali. Selanjutnya adalah melakukan proses iterasi dan menghitung nilai *error* pada iterasi ke- t . Lalu, melakukan *update* bobot data iterasi ke- t dan melakukan normalisasi bobot. Cek kondisi apakah iterasi ke- t yang dilakukan lebih besar sama dengan jumlah iterasi maksimum. Jika kondisi Ya, maka lanjutkan pada proses prediksi *Adaboost*. Jika kondisi Tidak, maka cek kondisi apakah nilai $error < 0,5$. Jika Ya maka ulangi proses menghitung nilai *error* iterasi ke- t dengan $t = t + 1$ dan jika kondisi Tidak maka lanjutkan pada proses prediksi *Adaboost*. Lalu, lakukan interpretasi hasil dan pengambilan keputusan.







Gambar 1.1 Flowchart Penerapan Algoritma Adaboost pada Algoritma C5.0