KLASIFIKASI WAKTU KELULUSAN MAHASISWA DENGAN KOMBINASI METODE *RANDOM FOREST* DAN *ADABOOST*

(Studi Kasus: Mahasiswa Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Tanjungpura)

HANA SALSABILA
NIM H1091211063

SKRIPSI

PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS TANJUNGPURA
PONTIANAK
2025

KLASIFIKASI WAKTU KELULUSAN MAHASISWA DENGAN KOMBINASI METODE *RANDOM FOREST* DAN *ADABOOST*

(Studi Kasus: Mahasiswa Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Tanjungpura)

HANA SALSABILA
NIM H1091211063

SKRIPSI

Sebagai salah satu syarat untuk memperoleh gelar
Sarjana Statistika pada Program Studi Statistika

PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS TANJUNGPURA
PONTIANAK
2025

LEMBAR PENGESAHAN

Judul Tugas Akhir	:	Klasifikasi Waktu Kelulusan Mahasiswa dengan Kombinasi Metode *Random Forest* dan *AdaBoost* (Studi Kasus: Mahasiswa Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Tanjungpura)
Nama Mahasiswa	:	Hana Salsabila
NIM	:	H1091211063
Jurusan/Program Studi	:	Matematika/Statistika
Tanggal Lulus	:	30 Januari 2025
SK Pembimbing	:	No. 2637/UN22.8/TD.06/2024/Tanggal 19 Desember 2024
SK Penguji	:	No. 94/UN22.8/TD.06/2025/Tanggal 23 Januari 2025

Dosen Pembimbing

Pembimbing I	Pembimbing II
Dr. Evy Sulistianingsih, M.Sc	Hendra Perdana, M.Sc
NIP 198502172008122006	NIP 198810102019031020

Dosen Penguji

Ketua Penguji	Anggota Penguji
Neva Satyahadewi, M.Sc., CRA., CRP., CRMP	Shantika Martha, M.Si
NIP 198212042005012001	NIP 198403082008122003
Pimpinan Sidang (merangkap anggota penguji)	Sekretaris Sidang (merangkap anggota penguji)
Dr. Evy Sulistianingsih, M.Sc	Hendra Perdana, M.Sc
NIP 198502172008122006	NIP 198810102019031020

Mengesahkan
Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Tanjungpura

Prof. Dr. Gusrizal, M.Si.
NIP 197108022000031001

PERNYATAAN INTEGRITAS AKADEMIK

Saya yang bertanda tangan di bawah ini:

Nama : Hana Salsabila

NIM : H1091211063

Program Studi/ Jurusan : Statistika/Matematika

Fakultas : Matematika dan Ilmu Pengetahuan Alam

dengan ini menyatakan bahwa dokumen ilmiah Tugas Akhir yang disajikan ini tidak mengandung unsur pelanggaran integritas akademik sesuai Peraturan Menteri Pendidikan, Kebudayaan, Riset, Dan Teknologi Republik Indonesia Nomor 39 Tahun 2021. Apabila di kemudian hari dokumen ilmiah Tugas Akhir ini mengandung unsur pelanggaran integritas akademik sesuai ketentuan perundangan tersebut, maka saya bersedia menerima sanksi akademik dan/atau sanksi hukum yang berlaku.

Demikian pernyataan ini untuk digunakan sebagaimana mestinya.

Pontianak, 30 Januari 2025

Hana Salsabila

NIM H1091211063

Klasifikasi Waktu Kelulusan Mahasiswa dengan Kombinasi Metode *Random Forest* dan *AdaBoost* (Studi Kasus: Mahasiswa Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Tanjungpura)

Abstrak

Institusi perguruan tinggi berperan sebagai penyelenggara pendidikan akademik yang ditujukan bagi mahasiswa. Mahasiswa menjadi tolak ukur dalam menilai kualitas perguruan tinggi, salah satunya melalui tingkat kelulusan tepat waktu. Proporsi mahasiswa yang menyelesaikan studi tepat waktu merupakan salah satu indikator krusial dalam evaluasi akreditasi perguruan tinggi. Namun, masih banyak mahasiswa Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) Universitas Tanjungpura (Untan) yang tidak lulus tepat waktu, sehingga mempengaruhi kualitas dan akreditasi institusi. Penelitian ini bertujuan untuk menerapkan kombinasi metode Random Forest dan AdaBoost dalam klasifikasi kelulusan mahasiswa serta menentukan akurasinya. Penelitian ini menggunakan data sekunder terkait kelulusan mahasiswa FMIPA Untan dari Periode I Tahun Akademik 2018/2019 hingga Periode II Tahun Akademik 2023. Analisis dimulai dari mendeskripsikan data kelulusan mahasiswa. Kemudian dilanjutkan ke tahap analisis Random Forest, lalu dilakukan evaluasi dengan menghitung nilai akurasinya. Selanjutnya algoritma AdaBoost diterapkan pada algoritma Random Forest, kemudian dilakukan evaluasi kembali dengan menghitung nilai akurasinya untuk melihat kinerja model. Hasil penelitian menunjukkan bahwa Algoritma Random Forest mampu mengklasifikasikan data sebanyak 53 data yang benar diklasifikasikan tepat waktu sedangkan 439 lainnya salah diklasifikasikan sedangkan pada data kelulusan tidak tepat waktu diketahui bahwa 1323 data yang benar diklasifikasikan tidak tepat waktu sedangkan 78 diantaranya salah diklasifikasikan dengan akurasi sebesar 72,7%. Penerapan tahap boosting dengan Algoritma AdaBoost menghasilkan peningkatan akurasi yang signifikan mencapai 96,8%. Atribut yang memiliki kontribusi terbesar dalam klasifikasi berdasarkan Random Forest adalah IPS Semester 3. Berdasarkan hasil penelitian, Algoritma AdaBoost terbukti efektif untuk meningkatkan akurasi Algoritma Random Forest dalam klasifikasi kelulusan mahasiswa.

Kata kunci: *Data Mining*, *Iterasi*, *Bobot*

Classification of Student Graduation Time with a Combination of Random Forest and AdaBoost Methods (Case Study: Students of the Faculty of Mathematics and Natural Sciences, Tanjungpura University)

Abstract

Higher education institutions act as providers of academic education aimed at students. Students are the benchmark in assessing the quality of higher education, one of which is through the rate of on-time graduation. The proportion of students who complete their studies on time is one of the crucial indicators in evaluating higher education accreditation. However, there are still many students of the Faculty of Mathematics and Natural Sciences (FMIPA) Tanjungpura University (Untan) who do not graduate on time, thus affecting the quality and accreditation of the institution. This study aims to apply a combination of Random Forest and AdaBoost methods in classifying student graduation and determining its accuracy. This study uses secondary data related to the graduation of FMIPA Untan students from Period I of the 2018/2019 Academic Year to Period II of the 2023 Academic Year. The analysis begins by describing student graduation data. Then proceed to the Random Forest analysis stage, then an evaluation is carried out by calculating the accuracy value. Furthermore, the AdaBoost algorithm is applied to the Random Forest algorithm, then a re-evaluation is carried out by calculating the accuracy value to see the performance of the model. The results of the study showed that the Random Forest Algorithm was able to classify 53 data that were correctly classified on time while 439 others were misclassified while in the data on graduation not on time it was known that 1323 data were correctly classified not on time while 78 of them were misclassified with an accuracy of 72.7%. The application of the boosting stage with the AdaBoost Algorithm resulted in a significant increase in accuracy reaching 96.8%. The attribute that had the largest contribution in the classification based on Random Forest was IPS Semester 3. Based on the results of the study, the AdaBoost Algorithm proved effective in increasing the accuracy of the Random Forest Algorithm in classifying student graduation.

Keywords: *Data Mining, Iteration, Weight*

PRAKATA

Puji syukur ke hadirat Allah Azza wa Jalla karena atas limpahan rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan penulisan skripsi ini. Skripsi ini berjudul **"Klasifikasi Waktu Kelulusan Mahasiswa Menggunakan Kombinasi Metode *Random Forest* dan *AdaBoost* (Studi Kasus: Mahasiswa Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Tanjungpura)"**. Skripsi ini dibuat dengan tujuan untuk memenuhi salah satu syarat dalam memperoleh gelar Sarjana Statistika pada Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Tanjungpura Pontianak.

Dengan selesainya penulisan skripsi ini, penulis menyadari bahwa selama proses penulisan skripsi ini tidak lepas dari bimbingan, arahan, dan dukungan dari berbagai pihak. Maka dari itu penulis ingin mengucapkan terima kasih kepada:

1. Allah Azza wa Jalla atas segala nikmat dan rahmat serta karunia pertolongan-Nya selama penulis menyelesaikan masa perkuliahan dan penulisan skripsi ini.
2. Keluarga tercinta terkhusus cinta pertama dan panutanku, ayahanda tercinta (Alm) Ayah Rasyid, beliau memang tidak sempat menemani penulis dalam perjalanan selama menempuh pendidikan. Namun semasa hidupnya beliau selalu memberikan kasih sayang, nasihat, motivasi, semangat dan doa tiada henti kepada penulis, dan Ibu Nurmawati serta saudara penulis yaitu kakak Siti Muthiah Zulfa dan adik Khanza Maulidya yang senantiasa memberikan doa, semangat, dan motivasi serta dukungan moril dan material kepada penulis selama proses perkuliahan dan dalam menyelesaikan skripsi ini.
3. Ibu Dr. Evy Sulistianingsih, M.Sc. selaku dosen pembimbing pertama yang telah banyak memberikan bimbingan serta arahan selama penulisan skripsi ini.
4. Bapak Hendra Perdana, M.Sc. selaku dosen pembimbing kedua yang telah banyak memberikan bimbingan serta arahan selama penulisan skripsi ini.

5. Bapak Yudhi, M.Si. selaku Dosen Pembimbing Akademik yang telah meluangkan waktu, memberikan saran, masukan, dan motivasi kepada penulis selama proses perkuliahan.
6. Ibu Neva Satyahadewi, M.Sc., CRA., CRP., CRMP. selaku dosen penguji pertama yang telah memberikan masukan dan saran kepada penulis.
7. Ibu Shantika Martha, M.Si. selaku dosen penguji kedua yang telah memberikan masukan dan saran kepada penulis.
8. Sahabat Queenitha Edenia Danovan yang telah memberikan dukungan dan doa kepada penulis, serta telah membersamai sejak masa SMP.
9. Sahabat Aulia Hayuningtyas, Angela Nadya Putri Ditya, Diva Rahma Kamila, Hafizah Ridha, serta Nindy Aisah yang telah memberikan dukungan dan doa kepada penulis, serta telah membersamai sejak masa SMA.
10. Sahabat Diva Rahma Kamila, Fadhela Trifaiza, Mira Asmara, dan Wina Priani yang telah memotivasi dan mendukung penulis, serta berjuang bersama selama menempuh perkuliahan.
11. Teman-teman Statistika 2021 (Expos) selaku teman seperjuangan yang telah memberikan dukungan dan motivasi kepada penulis.
12. Berbagai pihak yang telah membantu dalam penyusunan skripsi ini dan belum bisa disebutkan satu per satu.

Semoga Allah membalas semua kebaikan atas segala bantuan yang diberikan dalam penulisan skripsi ini. Penulis berharap skripsi ini dapat bermanfaat bagi semua pihak. Kritik dan saran yang bersifat membangun sangat diharapkan oleh penulis untuk menjadikan skripsi ini lebih baik.

Pontianak, 30 Januari 2025

Hana Salsabila

DAFTAR ISI

LEMBAR PENGESAHAN ii
PERNYATAAN INTEGRITAS AKADEMIK iii
Abstrak iv
Abstract v
PRAKATA vi
DAFTAR ISI viii
DAFTAR TABEL x
DAFTAR SIMBOL xi
BAB I PENDAHULUAN 1
1.1 Latar Belakang 1
1.2 Rumusan Masalah 4
1.3 Tujuan Penelitian 4
1.4 Batasan Penelitian 4
1.5 Tinjauan Pustaka 5
1.6 Metodologi Penelitian 6
BAB II LANDASAN TEORI 9
2.1 *Data Mining* 9
2.2 Klasifikasi 10
2.3 *Decision Tree* 10
2.4 *Classification and Regression Tree* (CART) 11
BAB III ENSEMBLE *RANDOM FOREST* DENGAN *ADABOOST* 14
3.1 Algoritma *Random Forest* 14
3.2 Algoritma *Adaptive Boosting* (*AdaBoost*) 16
3.3 Evaluasi Model Klasifikasi 17
BAB IV HASIL DAN PEMBAHASAN 19
4.1 Data Penelitian 19
4.2 Statistik Deskriptif 20
4.3 Pembentukan Algoritma *Random Forest* 22

4.4 Penerapan *Adaptive Boosting* 25
BAB V PENUTUP 29
DAFTAR PUSTAKA 31

DAFTAR TABEL

Tabel 3.1 *Confusion Matrix*..... 18
Tabel 4.1 Pengklasifikasian Atribut Independen..... 19
Tabel 4.2 Statistik Deskriptif Mahasiswa FMIPA 21
Tabel 4.3 Nilai Indeks Gini dan Gini Split Pada Setiap Atribut 23
Tabel 4.4 *Confusion Matrix Random Forest*..... 25
Tabel 4.5 Hasil *Update* Bobot..... 27
Tabel 4.6 *Confusion Matrix AdaBoost*..... 28

DAFTAR SIMBOL

n	:	Jumlah data
k	:	Jumlah kelas
P_R	:	Proporsi banyaknya data yang masuk terhadap simpul kanan
P_L	:	Proporsi banyaknya data yang masuk terhadap simpul kiri
$i(t)$	:	Indeks gini ke-*t*
$i(t_L)$	:	Indeks gini ke-*t* pada simpul kiri
$i(t_R)$	:	Indeks gini ke-*t* pada simpul kanan
$n(t)$	:	Jumlah data pada *node t*
$n_j(t)$	:	Jumlah data pada $node\ t$ di kelas ke-*j*, di mana $j = 1,2,3,4, \dots, k$
$n(t_L)$	:	Jumlah data pada $node\ t$ di simpul kiri
$n(t_R)$	:	Jumlah data pada $node\ t$ di simpul kanan
$P(j\|t)$	:	Proporsi kelas j pada $node\ t$ di mana $j = 1,2,3,4, \dots, k$
$P(j\|t_L)$	:	Proporsi kelas j pada $node\ t$ di mana $j = 1,2,3,4, \dots, k$ pada simpul kiri
$P(j\|t_R)$	:	Proporsi kelas j pada $node\ t$ di mana $j = 1,2,3,4, \dots, k$ pada simpul kanan
$error^s$	:	*Weighted errors* iterasi ke-s
$error_i^s$	:	*Weighted errors* objek ke-i iterasi ke-*s*
a^s	:	Indeks kepercayaan iterasi ke-s
w_i^s	:	Bobot objek ke-i iterasi ke-s
S	:	Jumlah iterasi maksimum
$f^s(x)$	:	Model prediksi iterasi ke-s
y_j	:	Prediksi akhir untuk kelas ke-j

BAB I
PENDAHULUAN

1.1 Latar Belakang

Perguruan tinggi merupakan satu diantara penyelenggara pendidikan bagi mahasiswa untuk menempuh pendidikan akademik (Kadafi, 2020). Menurut data yang diperoleh dari Pangkalan Data Pendidikan Tinggi yang dimiliki oleh Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi, terdapat sebanyak 4.654 lembaga yang menyelenggarakan perguruan tinggi di Indonesia hingga tahun 2018. Persentase kemampuan mahasiswa yang berhasil menyelesaikan studi tepat waktu menjadi salah satu faktor penentu kualitas suatu perguruan tinggi (Agwil, Fransiska, & Hidayati, 2020).

Penilaian kinerja akademik mahasiswa diarahkan pada kemampuan mahasiswa dalam menyelesaikan studi tepat waktu (Hasana & Rahmati, 2022). Sesuai dengan ketentuan lama studi yang telah ditetapkan oleh Kementerian Pendidikan dan Kebudayaan Direktorat Jendral Perguruan Tinggi tentang kompetensi lulus bagi mahasiswa jenjang S1 yaitu dapat menyelesaikan masa studi selama 8 semester atau 4 tahun dengan minimal 144 SKS sebagai beban wajib. Selain itu, pemberian akreditasi pada program studi perguruan tinggi oleh Badan Akreditasi Nasional – Perguruan Tinggi dipengaruhi kelulusan tepat waktu, penyelesaian studi dan proporsi mahasiswa yang menyelesaikan studi dalam batas masa studi yang telah ditentukan sebagai salah satu kriteria penilaian (Ban-PT, 2008).

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. Pada Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) terdapat 10 program studi. 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 di FMIPA masih banyak ditemukan mahasiswa yang tidak lulus tepat waktu sehingga akan berpengaruh pada penilaian akreditasi. Untuk itu, setiap pemangku kepentingan di perguruan tinggi memiliki tanggung jawab dalam menjalankan fungsi dan peran masing-masing terutama pihak fakultas. Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Tanjungpura (FMIPA Untan) juga memiliki peran dalam menentukan kebijakan pada sistem pendidikan di masa-masa yang akan datang agar menghasilkan lulusan yang bermutu (Hidayat, Sulistianingsih, & Debataraja, 2015). Oleh karena itu, karakteristik mahasiswa yang lulus tepat waktu maupun tidak tepat waktu dalam penentuan kelulusan mahasiswa dapat dianalisis menggunakan teknik dalam *data mining* (Crismayella, Satyahadewi, & Perdana, 2023).

Data mining merupakan proses untuk menemukan informasi atau pola menggunakan teknik statistik yang diperoleh dengan cara mengekstrak dan mengidentifikasi pola yang berguna serta menarik dari berbagai *database* yang besar. Salah satu metode dalam *data mining* yang sering digunakan adalah klasifikasi. 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, & Maruddani, 2021). *Decision Tree* dapat menangani *overfitting*, 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*). Selain itu algoritma dari *decision tree* yang dapat digunakan adalah *Classification and Regression Trees* (CART). CART adalah metode eksplorasi data berdasarkan teknik pohon keputusan yang dapat menghasilkan pohon klasifikasi untuk data kategorik atau

pohon regresi untuk data numerik. Ada sebuah metode kembangan dari sebuah metode CART yaitu *Random Forest* (Dewi, Mulyadi, & Syafitri, 2012).

Random Forest pertama kali diperkenalkan oleh Breiman pada tahun 2001 merupakan pengembangan dari metode CART yang menerapkan metode *bootstrap aggregating* (*bagging*) dan seleksi fitur random (*random feature selection*). Metode ini membangun beberapa pohon keputusan yang bekerja secara independen untuk menghasilkan model klasifikasi yang lebih akurat. Keunggulan *Random Forest* terletak pada kemampuannya untuk menghindari *overfitting* serta mengurangi korelasi antar pohon keputusan. *Random Forest* juga merupakan gabungan dari masing-masing teknik pohon keputusan yang ada, yang kemudian dikombinasikan ke dalam suatu model klasifikasi. Setiap keputusan pada *Random Forest* menghasilkan prediksi semakin banyak prediksi, semakin banyak pula pohon sehingga bisa disebut *Random Forest* (Hendapratama, Hamzah, & Astuti, 2022).

Ketidakseimbangan kelas (*class imbalance*) merupakan salah satu permasalahan dalam klasifikasi, dimana satu class diwakili oleh sampel yang besar, sedangkan yang lainnya hanya diwakili oleh beberapa sampel (Bisri & Wahono, 2015). Kondisi ini dapat mempengaruhi kinerja klasifikasi. Oleh karena itu, diperlukan pendekatan tambahan seperti *boosting* untuk meningkatkan kinerja klasifikasi *Decision Tree* yang lebih baik. *Boosting* merupakan metode *ensemble* yang paling popular 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.

Random Forest dan *AdaBoost* merupakan dua algoritma klasifikasi yang memiliki karakteristik serta keunggulan tersendiri. *Random Forest* bekerja dengan membangun sejumlah pohon keputusan (*decision tree*) dan menggabungkan hasil prediksi-nya guna meningkatkan tingkat akurasi serta meminimalkan risiko *overfitting*. Di sisi lain, *AdaBoost* berfokus pada peningkatan akurasi model dengan memberikan bobot lebih besar pada data yang salah diprediksi oleh model sebelumnya, sehingga model selanjutnya lebih terarah dalam menangani data yang

sulit diprediksi. Dengan mengombinasikan kedua metode tersebut, diharapkan dapat diperoleh model klasifikasi yang lebih kuat (*robust*) dan akurat dalam memprediksi waktu kelulusan mahasiswa.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah dari penelitian ini yaitu:

1. Bagaimana penerapan kombinasi metode *Random Forest* dan *AdaBoost* dalam klasifikasi kelulusan mahasiswa?
2. Bagaimana akurasi dari kombinasi metode *Random Forest* dan *AdaBoost* dalam klasifikasi kelulusan mahasiswa?

1.3 Tujuan Penelitian

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

1. Untuk menerapkan kombinasi metode *Random Forest* dan *AdaBoost* dalam klasifikasi kelulusan mahasiswa.
2. Untuk menentukan akurasi dari kombinasi metode *Random Forest* dan *AdaBoost* dalam klasifikasi kelulusan mahasiswa.

1.4 Batasan Penelitian

Diperlukan Batasan masalah agar penelitian ini lebih terarah. Adapun Batasan masalah dalam penelitian ini antara lain:

1. Data yang digunakan dalam penelitian adalah data kelulusan mahasiswa Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Tanjungpura Periode I Tahun Ajaran 2018/2019 sampai Periode II Tahun Ajaran 2023/2024
2. Atribut dependen yang digunakan adalah Status kelulusan mahasiswa (Y), sedangkan atribut independennya yaitu IPS semester 1 (X_1), IPS semester 2 (X_2), IPS semester 3 (X_3), IPS semester 4 (X_4), jalur masuk (X_5), daerah asal domisili (X_6), kota asal domisili (X_7), jurusan SMA (X_8), pekerjaan ayah

(X_9), pekerjaan ibu (X_{10}), pendidikan ayah (X_{11}), pendidikan ibu (X_{12}), pendapatan ayah (X_{13}), dan pendapatan ibu (X_{14}).

1.5 Tinjauan Pustaka

Kamagi dan Hansun (2014) melakukan penelitian tentang penerapan Algoritma C4.5 untuk memprediksi tingkat kelulusan mahasiswa Universitas Multimedia Nusantara. Penelitian dilakukan dengan membagi 4 kategori tingkat kelulusan mahasiswa, yaitu lulus tepat waktu, lulus cepat, lulus terlambat dan tidak lulus atau *drop out*. Adapun variabel-variabel yang digunakan meliputi jumlah SKS, Indeks Prestasi Sementara (IPS) dari semester 1 sampai 6, asal SMA, dan jenis kelamin. Berdasarkan hasil penelitian, diperoleh nilai akurasi dari 60 data *training* dan 40 data *testing* sebesar 87,5%, dengan Indeks Prestasi Sementara (IPS) 6 sebagai atribut yang paling berpengaruh dalam hasil prediksi.

Dhawangkhara dan Riksakomara (2017) melakukan analisis perbandingan antara dua algoritma dalam klasifikasi yaitu Algoritma *Classification and Regression Tree* (CART) dengan Algoritma *Random Forest*. Data yang digunakan merupakan data cuaca yang bersumber dari *website National Climate Data Center* (NCDC) berasal dari stasiun cuaca Juanda, Surabaya selama 17 tahun (2000-2016). Atribut-atribut yang digunakan adalah kecepatan angin, tekanan udara, titik embun, suhu udara, visibilitas dan curah hujan. Diperoleh akurasi menggunakan Algoritma CART sebesar 72%. Sedangkan dengan Algoritma *Random Forest* diperoleh akurasi sebesar 78%. Sehingga dapat disimpulkan bahwa untuk memprediksi potensi curah hujan di Kota Surabaya lebih baik menggunakan Algoritma *Random Forest*.

Penelitian menggunakan *Classification and Regression Tree* (CART) dan *Random Forest* untuk mengklasifikasikan ketidaktepatan waktu lulus mahasiswa juga dilakukan oleh Suwardika dan Suniantara tahun 2019. Data yang digunakan merupakan data wisuda Universitas Terbuka Denpasar dari 133 wisudawan tahun 2018. Atribut-atribut yang digunakan adalah jurusan/prodi, status mahasiswa, jenis kelamin, penghasilan orang tua, pendidikan orang tua, dan pekerjaan orang tua. Berdasarkan hasil penelitian, diperoleh bahwa akurasi dengan menggunakan

Algoritma CART sebesar 65,41%. Sedangkan dengan Algoritma *Random Forest* pada CART diperoleh akurasi sebesar 93,23%.

Dalam penelitian Mauludiyah tahun 2020 dilakukan penerapan Algoritma *Random Forest* untuk mengklasifikasikan Indeks Pengetahuan Manusia (IPM) di setiap kabupaten/kota di Indonesia yang mencapai 514 kabupaten/kota. Atribut yang digunakan penelitian ini mencakup harapan lama sekolah, rata-rata lama sekolah, angka harapan hidup, dan pendapatan per kapita. Berdasarkan hasil analisis menggunakan Algoritma *Random Forest* diperoleh akurasi sebesar 93,69% dengan sebanyak 100 pohon yang dipilih.

Rohan, Rahmah, dan Siddik (2019) melakukan penelitian tentang klasifikasi deteksi kanker payudara menggunakan Algoritma *Random Forest* dengan *AdaBoost*. Penelitian dilakukan dengan membagi 2 kategori kanker payudara, yaitu tumor jinak dan tumor ganas. Adapun variabel-variabel yang digunakan meliputi ketebalan gumpalan, keseragaman ukuran sel, keseragaman bentuk sel, adhesi marjinal, ukuran sel epitel Tunggal, nucleus, kromatin hambar, nukleus normal dan mitosis. Berdasarkan hasil penelitian, diperoleh nilai akurasi *Random Forest* sebesar 97,1429% sedangkan kombinasi *Random Forest* dan *AdaBoost* sebesar 98,5714%.

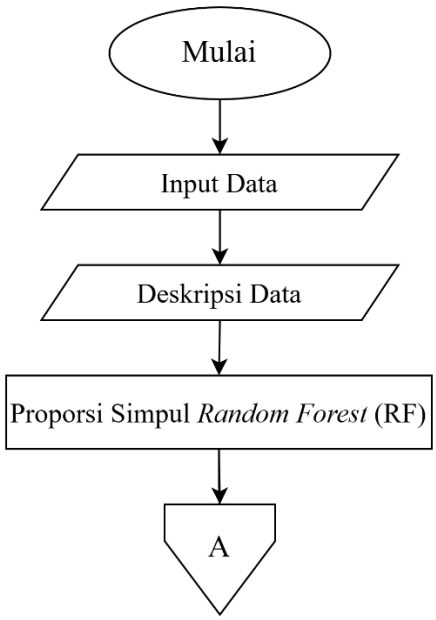
1.6 Metodologi Penelitian

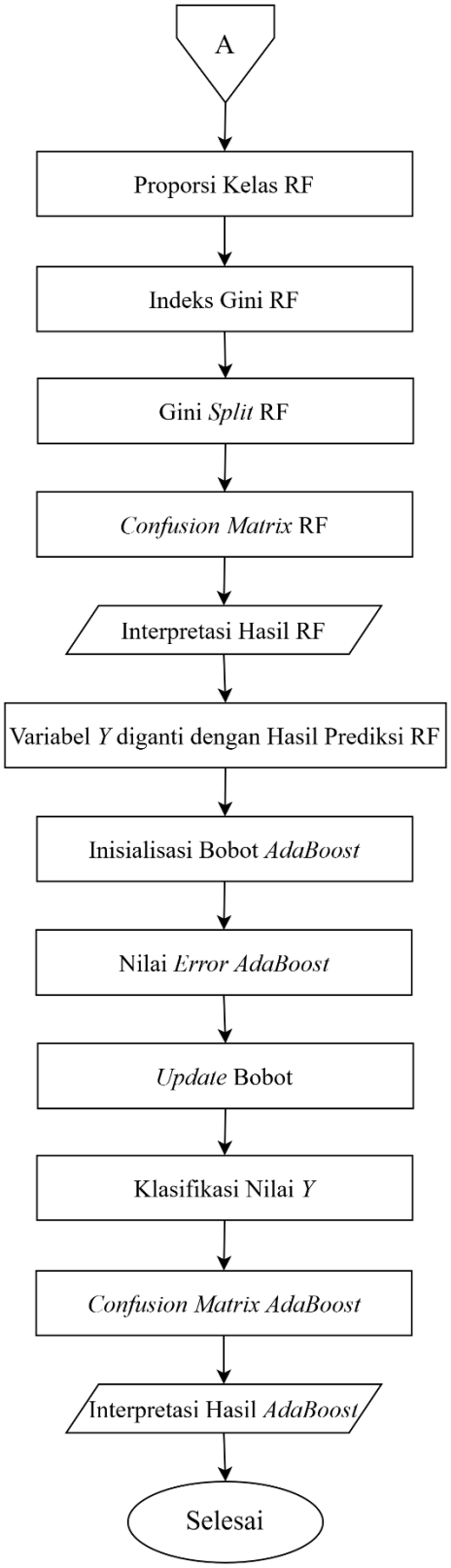
Penelitian ini menggunakan data sekunder yang diperoleh dari Biro Akademik dan Kemahasiswaan (BAK) Untan, yaitu berupa data kelulusan mahasiswa Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) Untan. Atribut independen yang digunakan adalah sebanyak 14 atribut, yaitu IPS semester 1, IPS semester 2, IPS semester 3, IPS semester 4, jalur masuk, daerah asal domisili, kota asal domisili, jurusan SMA, pekerjaan ayah, pekerjaan ibu, pendidikan ayah, pendidikan ibu, pendapatan ayah, dan pendapatan ibu. Atribut dependen yang digunakan adalah status kelulusan mahasiswa.

Tahapan-tahapan dalam penelitian ini dimulai dengan menginput data ke dalam *software* R Studio. Data yang telah di input kemudian dideskripsikan untuk memberi informasi awal mengenai karakteristik dataset. Selanjutnya, Algoritma

Random Forest digunakan untuk memproses data, dimulai dari pembagian data ke dalam simpul-simpul (*node*), perhitungan proporsi kelas di setiap simpul, hingga perhitungan indeks gini setiap simpul. Nilai gini digunakan untuk mengukur tingkat kesenjangan data, sementara proses gini split dilakukan untuk menentukan pembagian terbaik dalam pembuatan pohon-pohon *Random Forest*. Evaluasi awal dilakukan dengan *confusion matrix Random Forest* untuk mengukur kinerja prediksi.

Hasil prediksi dari Algoritma *Random Forest* selanjutnya digunakan untuk menggantikan nilai variabel target (Y) pada data. Pada tahap berikutnya, Algoritma *AdaBoost* diterapkan dengan memulai inisialisasi bobot awal yang merata pada setiap sampel. Proses ini diikuti dengan perhitungan nilai kesalahan prediksi (*error rate*) yang digunakan untuk memperbarui bobot pada iterasi berikutnya. Sampel yang sulit diprediksi diberikan bobot yang lebih tinggi untuk meningkatkan akurasi prediksi pada iterasi berikutnya. Setelah proses pembelajaran selesai, data kembali diklasifikasikan menggunakan hasil akhir prediksi dari *AdaBoost*, yang kemudian dievaluasi dengan *confusion matrix* untuk menilai kinerja model secara keseluruhan. Pada tahap akhir, dilakukan interpretasi hasil dari kombinasi kedua algoritma, yaitu *Random Forest* dan *AdaBoost* berupa nilai akurasi. Langkah-langkah analisis dalam penelitian ini dapat dilihat pada *flowchart* berikut ini:





Gambar 1.1 *Flowchart* Kombinasi Algoritma *Random Forest* dan *AdaBoost*