



**KLASIFIKASI KETEPATAN LAMA STUDI MAHASISWA
DENGAN ALGORITMA RANDOM FOREST DAN GRADIENT
BOOSTING**

**(Studi Kasus Fakultas Ilmu Komputer Universitas Pembangunan
Nasional Veteran Jakarta)**

SKRIPSI

MUHAMMAD LABIB MU'TASHIM

1810511067

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI INFORMATIKA

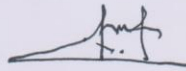
2022

LEMBAR PENGESAHAN

Tugas Akhir ini diajukan oleh:

Nama : Muhammad Labib Mu'tashim
NIM : 1810511067
Program Studi : S1 Informatika
Judul Tugas Akhir : **Klasifikasi Ketepatan Lama Studi Mahasiswa Dengan Algoritma Random Forest Dan Gradient Boosting (Studi Kasus Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta)**

Telah berhasil dipertahankan dihadapan Tim Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.



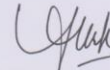
Dr. Ermatita, M.Kom

Penguji I



Bayu Hananto, S.Kom, M.Kom

Penguji II



Ati Zaidiah, S.Kom, M.TI

Pembimbing



Dr. Ermatita, M.Kom

Dekan



Dr. Widya Cholil, S.Kom, M.I.T.

Ketua Program Studi

Ditetapkan di : Jakarta

Tanggal Ujian : 10 Januari 2023



LEMBAR PERSETUJUAN

Skripsi dengan judul **KLASIFIKASI KETEPATAN LAMA STUDI MAHASISWA
DENGAN ALGORITMA RANDOM FOREST DAN GRADIENT BOOSTING**
yang ditulis oleh:

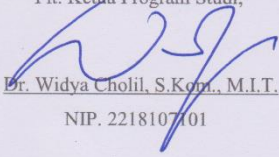
Muhammad Labib Mu'tashim

NIM. 1810511067

Telah diperiksa dan disetujui, serta layak untuk diujikan.

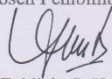
Jakarta, 9 Desember 2022

Mengetahui,
Plt. Ketua Program Studi,


Dr. Widya Cholil, S.Kom., M.I.T.

NIP. 2218107701

Mengetahui,
Dosen Pembimbing,


Ati Zaidiah, S. Kom., M.TI.

NIP. 197402212021212004

LEMBAR ORISINILITAS

Skripsi ini adalah hasil karya sendiri dan sumber yang dikutip maupun yang dirujuk telah saya nyatakan dengan benar.

Nama : Muhammad Labib Mu'tashim

NIM : 1810511067

Tanggal : 10 Januari 2023

Judul Skripsi : **Klasifikasi Ketepatan Lama Studi Mahasiswa Dengan Algoritma Random Forest Dan Gradient Boosting (Studi Kasus Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta)**

Bilamana dikemudian hari ditemukan ketidaksamaan dengan pernyataan ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Jakarta, 10 Januari 2023



Muhammad Labib Mu'tashim

**PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI
UNTUK KEPENTINGAN AKADEMIS**

Sebagai civitas akademik Universitas Pembangunan Nasional Veteran Jakarta yang bertanda tangan dibawah ini :

Nama : Muhammad Labib Mu'tashim

NIM : 1810511067

Program Studi : S1 Informatika

Fakultas : Ilmu Komputer

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Pembangunan Nasional Veteran Jakarta Hak Bebas Royalti Non Eksklusif (Non-exclusive Royalty Free Right) atas karya ilmiah saya yang berjudul

**KLASIFIKASI KETEPATAN LAMA STUDI MAHASISWA DENGAN
ALGORITMA RANDOM FOREST DAN GRADIENT BOOSTING**

**(Studi Kasus Fakultas Ilmu Komputer Universitas Pembangunan Nasional
Veteran Jakarta)**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional Veteran Jakarta berhak untuk menyimpan, mengalih media atau format, mengelola dalam bentuk pangkalan data (database), merawat, serta mempublikasikan skripsi saya selama tetap mencantumkan nama saya sebagai penulis atau pencipta dan pemilih Hak Cipta.

Demikian pertanyaan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta, 25 Januari 2023

Yang Menyatakan,



(Muhammad Labib Mu'tashim)

KLASIFIKASI KETEPATAN LAMA STUDI MAHASISWA DENGAN ALGORITMA RANDOM FOREST DAN GRADIENT BOOSTING (Studi Kasus Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta)

MUHAMMAD LABIB MU'TASHIM

ABSTRAK

Universitas umumnya pada setiap tahun menerima mahasiswa baru dan memiliki kuota yang berbeda pada setiap jurusannya, namun hal ini tidak diimbangi dengan jumlah mahasiswa yang lulus, sehingga mengakibatkan jumlah mahasiswa meningkat setiap tahunnya, begitu juga dengan Fakultas Ilmu Komputer (FIK) Universitas Pembangunan Nasional (UPN) Veteran Jakarta. Melimpahnya data akademik di FIK UPN Veteran Jakarta bisa diolah sesuai yang dibutuhkan dan berguna mencari informasi penting demi pengembangan fakultas menjadi lebih baik. Maka dari itu dilakukan penelitian untuk menganalisis mahasiswa yang lulus tepat waktu maupun tidak tepat waktu dengan data mining.

Penelitian ini menggunakan metode Random Forest dan Gradient Boosting untuk mengetahui tingkat akurasi dan menentukan mana model klasifikasi yang terbaik pada ketepatan lulus mahasiswa. Kedua algoritma ini memiliki akurasi yang sangat baik pada kebutuhan data yang sangat besar. Analisis menggunakan data mahasiswa S1 FIK UPN Veteran Jakarta angkatan 2015 - 2017. Hasil uji coba sampel pada 590 data, algoritma random forest 10 k-fold mendapatkan akurasi 82,64% dan pada gradient boosting 3 k-fold mendapatkan akurasi 79,66%. Hasil penelitian ini digunakan sebagai salah satu dasar pengambilan keputusan untuk menentukan kebijakan oleh pihak fakultas.

Kata kunci : Klasifikasi, Random Forest, Gradient Boosting, Kelulusan

CLASSIFICATION OF STUDENT STUDY DURATION USING RANDOM FOREST AND GRADIENT BOOSTING ALGORITHMS (Case Study of the Faculty of Computer Science, Pembangunan Nasional Veteran University Jakarta)

MUHAMMAD LABIB MU'TASHIM

ABSTRACT

Universities generally admit new students every year and have different quotas for each department, but this is not matched by the number of students graduating, resulting in the number of students increasing every year, as well as the Faculty of Computer Science (FIK) at the National Development University (UPN).) Jakarta Veterans. The abundance of academic data at FIK UPN Veteran Jakarta can be processed according to what is needed and is useful for finding important information for better faculty development. Therefore, research was conducted to analyze students who graduated on time or not on time with data mining.

This study uses the Random Forest and Gradient Boosting methods to determine the level of accuracy and determine which classification model is the best for the accuracy of student graduation. Both of these algorithms have very good accuracy on very large data requirements. The analysis uses data from S1 FIK UPN Veteran Jakarta students class 2015 - 2017. The results of the sample trial on 590 data, the 10 k-fold random forest algorithm obtains 82.64% accuracy and the 3 k-fold gradient boosting obtains 79.66% accuracy. The results of this study are used as a basis for decision making for determining policies by the faculty.

Keywords : *Classification, Random Forest, Gradient Boosting, Graduation*

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Allah SWT atas segala karunia-Nya sehingga Skripsi ini berhasil diselesaikan. Penulis juga ingin mengucapkan terima kasih kepada :

1. Ibu Ati Zaidiah selaku dosen pembimbing yang telah banyak membantu dan memberikan saran yang bermanfaat.
2. Bapak Noor Falih yang memberikan pengajaran yang membantu penulis menyelesaikan skripsi ini.
3. Orang tua, keluarga yang selalu memberikan dukungan kepada penulis agar dapat menyelesaikan skripsi ini.
4. Bapak Desta Sandya Prasvita selaku Kaprodi Informatika yang banyak membantu penulis memberikan saran terkait skripsi ini.
5. Ibu Iin Ernawati juga membantu saya dalam menyelesaikan skripsi ini.
6. Serta semua pihak yang membantu penulis.

Jakarta, 26 Desember 2022

Penulis

DAFTAR ISI

ABSTRAK	vi
ABSTRACT	vii
KATA PENGANTAR	viii
DAFTAR ISI	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiv
BAB 1	1
PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	3
BAB 2	5
TINJAUAN PUSTAKA	5
2.1 Kelulusan Mahasiswa	5
2.2 Data Mining	8
2.2.1 Ensemble Learning	9
2.2.1.1 Klasifikasi	10
2.2.1.2 Algoritma Random Forest	11
2.2.1.3 Algoritma Gradient Boosting	13
2.2 Gini Impurity	15
2.3 Cross Validation	16
2.4 Confusion Matrix	17
2.5 Penelitian Terkait	18
2.5.1 Skripsi Klasifikasi Ketepatan Lama Studi Mahasiswa Menggunakan SVM dan Random Forest	18
2.5.2 Jurnal Integrasi Gradient Boosted Trees dengan SMOTE dan Bagging untuk Deteksi Kelulusan Mahasiswa	19

2.5.3	Skripsi Implementasi Metode Random Forest Dan Xgboost Pada Klasifikasi Customer Churn.....	20
2.5.4	Skripsi Prediksi Kelulusan Mahasiswa Fakultas Sains dan Teknologi Universitas Sanata Dharma Menggunakan Metode Klasifikasi Naïve Bayes	20
2.5.5	Jurnal Implementation of Data Mining for Drop-Out Prediction using Random Forest Method	21
2.6	Perbandingan Penelitian Terkait	22
BAB 3	24
METODOLOGI PENELITIAN.....		24
3.1	Alur Penelitian.....	24
3.2	Identifikasi Masalah	25
3.3	Studi Pustaka	26
3.4	Pengumpulan Data	26
3.5	Pra-processing Data.....	27
3.6	Pembagian Data.....	29
3.7	Data Mining dan Pemodelan	30
3.8	Analisis Performa	33
3.9	Peralatan Penelitian	33
3.10	Jadwal Penelitian	33
BAB 4	35
HASIL DAN PEMBAHASAN.....		35
4.1	Tahapan Pre-processing	35
4.1.1	Data Cleaning.....	35
4.1.2	Data Transformation	36
4.1.3	Data Reduction.....	38
4.2	Persiapan Data Training dan Data Testing.....	39
4.3	Pengolahan Data.....	40
4.3.1	Analisis Perhitungan Random Forest.....	40
4.3.2	Analisis Perhitungan Gradient Boosting.....	46
4.3.3	Klasifikasi Algoritma Random Forest dan Gradient Boosting	56
4.4	Pengujian Data	66
4.4.1	Uji Variasi Atribut.....	66
4.5	Analisis Hasil	69

4.5.1	Perbandingan Akurasi	80
4.5.2	Pengujian Data Asli.....	81
BAB 5	72
PENUTUP	72
5.1	Kesimpulan.....	72
5.2	Saran.....	72
DAFTAR PUSTAKA	73
LAMPIRAN	76

DAFTAR TABEL

Tabel 1. Tabel Jumlah Mahasiswa Informatika S1 FIK UPNVJ	6
Tabel 2. Tabel Jumlah Mahasiswa Sistem Informasi S1 FIK UPNVJ.....	6
Tabel 3. Confusion Matrix	17
Tabel 4. Daftar penelitian terkait.....	22
Tabel 5. Variabel/ Atribut Penelitian	27
Tabel 6. Jadwal Penelitian.....	34
Tabel 7. Atribut kategorial yang dilakukan <i>data transformation</i>	37
Tabel 8. Data sesudah reduce variabel	38
Tabel 9. Balancing Data	39
Tabel 10. Pembagian data training dan data testing.....	39
Tabel 11. Bootstrap Dataset.....	40
Tabel 12. Label/ Atribut IPS 2	41
Tabel 13. Label/ Atribut IPS 5	41
Tabel 14. Hasil Perhitungan Gini Impurity IPS 2	42
Tabel 15. Hasil Perhitungan Gini Impurity IPS 5	42
Tabel 16. Hasil Perhitungan Gini Impurity Beasiswa.....	42
Tabel 17. Perbandingan Gini Impurity tiap atribut (root node).....	43
Tabel 18. Hasil Perhitungan Gini Impurity IPS 2 (node ke-2).....	43
Tabel 19. Hasil Perhitungan Gini Impurity Jalur Masuk (node ke-2).....	44
Tabel 20. Perbandingan Gini Impurity tiap atribut (node ke-2).....	44
Tabel 21. Hasil Klasifikasi Random Forest.....	45
Tabel 22. Pengujian hasil dengan Confussion Matrix.....	46
Tabel 23. Contoh Data Training.....	47
Tabel 24. Data Training dengan Residual	48
Tabel 25. Hasil Perhitungan SSR IPS 2	48
Tabel 26. Hasil Perhitungan SSR IPS 5	49
Tabel 27. Hasil Perhitungan SSR Jalur Masuk	49
Tabel 28. Perbandingan SSR tiap atribut (root node).....	49
Tabel 29. Hasil Perhitungan SSR IPS 2	50
Tabel 30. Hasil Perhitungan SSR Jalur Masuk	50

Tabel 31. Perbandingan SSR tiap atribut (node selanjutnya).....	50
Tabel 32. Tabel data dengan Residual baru (kedua)	52
Tabel 33. Tabel data dengan residual baru (ketiga)	53
Tabel 34. Tabel Log(Odds) Data Testing.....	55
Tabel 35. Tabel Klasifikasi Data Testing	55
Tabel 36. Pengujian hasil dengan Confussion Matrix.....	55
Tabel 37. Pemeringkatan atribut.....	61
Tabel 38. Pengujian hasil RF dengan Confussion Matrix.....	62
Tabel 39. Pengujian hasil GB dengan Confussion Matrix	63
Tabel 40. Pengecekan overfitting dengan akurasi	65
Tabel 41. Hasil Pengujian menggunakan 2 Atribut	67
Tabel 42. Hasil Pengujian menggunakan 5 Atribut	67
Tabel 43. Hasil Pengujian menggunakan 10 Atribut	68
Tabel 44. Hasil Pengujian menggunakan 15 Atribut	69
Tabel 45. Hasil Optimasi Semua Variasi	69
Tabel 46. Pengujian data angkatan 2018 dengan Confussion Matrix	70
Tabel 47. Perbandingan jumlah data asli dan klasifikasi angkatan 2018	70
Tabel 48. Perbandingan jumlah data asli dan klasifikasi per prodi.....	71

DAFTAR GAMBAR

Gambar 1. Perbandingan ketepatan lulus mahasiswa S1 FIK UPNVJ.	7
Gambar 2. Tahapan dalam Data Mining	8
Gambar 3. Ensemble learning dengan tipe bagging dan boosting.	10
Gambar 4. Tahapan Klasifikasi dalam Data Mining	11
Gambar 5. Alur kerja Random Forest	12
Gambar 6. Alur kerja Gradient Boosting	14
Gambar 7. Cross Validation	17
Gambar 8. Alur Penelitian	25
Gambar 9. Tabel dataset sebelum dilakukan cleaning data.....	36
Gambar 10. Tabel dataset setelah dilakukan cleaning data.....	36
Gambar 11. Dataset sebelum dilakukan <i>data transformation</i>	37
Gambar 12. Dataset setelah dilakukan <i>data transformation</i>	38
Gambar 13. Root node hasil perhitungan	43
Gambar 14. Pohon Decision Tree ke-1	44
Gambar 16. Root node hasil perhitungan	50
Gambar 17. Tree 1 hasil perhitungan	51
Gambar 18. Output Values pada setiap leaf node	52
Gambar 19. Regression Tree Kedua.....	53
Gambar 20. Regression Tree Ketiga	54
Gambar 21. Regression Tree dengan Program Python	54
Gambar 22. Kode program cleaning dan reduction dataset.....	56
Gambar 23. Kode program data transformation.....	57
Gambar 24. Kode program oversampling dataset	57
Gambar 25. Kode program split dataset	58
Gambar 26. Kode program klasifikasi random forest	58
Gambar 27. Kode program jumlah data testing random forest.	59
Gambar 28. Kode program klasifikasi gradient boosting.....	59
Gambar 29. Kode program jumlah data testing gradient boosting.....	60
Gambar 30. Grafik pemeringkatan atribut.....	61
Gambar 31. Kode program confussion matrix.	62

Gambar 32. Hasil Pengukuran Kinerja Algoritma Random Forest.....	63
Gambar 33. Hasil Pengukuran Kinerja Algoritma Gradient Boosting.....	64
Gambar 34. Kode Program perbandingan akurasi data training dan testing	65
Gambar 35. Kode Program Plot Tree Random Forest.....	66
Gambar 36. Kode Program Plot Tree Gradient Boosting.....	66