

**PENERAPAN SELEKSI FITUR *ANALYSIS OF VARIANCE* PADA
ALGORITMA *RANDOM FOREST CLASSIFIER* DALAM
KLASIFIKASI NILAI MAHASISWA**

**STUDI KASUS: UNIVERSITAS MUHAMMADIYAH
KALIMANTAN TIMUR**

SKRIPSI

Diajukan untuk memenuhi persyaratan dalam mencapai gelar Sarjana
Komputer

DISUSUN OLEH

MUHAMMAD FATH THORIQ

1911102441166



**PROGRAM STUDI S1 TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS MUHAMMADIYAH KALIMANTAN TIMUR
KOTA SAMARINDA**

2023

**Penerapan Seleksi Fitur *Analysis of Variance* pada
Algoritma *Random Forest Classifier* dalam Klasifikasi Nilai
Mahasiswa**

Studi Kasus: Universitas Muhammadiyah Kalimantan Timur

SKRIPSI

Diajukan untuk memenuhi persyaratan dalam mencapai gelar Sarjana
Komputer

Disusun oleh

Muhammad Fath Thoriq

1911102441166



**PROGRAM STUDI S1 TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS MUHAMMADIYAH KALIMANTAN TIMUR
KOTA SAMARINDA
2023**

HALAMAN PENGESAHAN

PENERAPAN SELEKSI FITUR *ANALYSIS OF VARIANCE* PADA ALGORITMA
RANDOM FOREST CLASSIFIER DALAM KLASIFIKASI NILAI MAHASISWA
STUDI KASUS: UNIVERSITAS MUHAMMADIYAH KALIMANTAN TIMUR

DISUSUN OLEH:

MUHAMMAD FATH THORIQ

1911102441166

Telah melaksanakan ujian skripsi dan dinyatakan lulus,
Pada tanggal 12 Juli 2023

Dosen Pembimbing



Wawan Joko Pranoto, S.Kom, M.TI

NIDN. 1102057701

Penguji



Faldi, S.Kom, M.TI

NIDN. 1121079101

Dekan



Prof. Ir. Sarjito, MT., Ph.D

NIDN. 0610116204

Ketua Program Studi



Asslia Johar Latipah, S.Kom., M.Cs

NIDN. 1124098902

PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Muhammad Fath Thoriq

NIM : 1911102441166

Konsentrasi : Sistem Informasi

Menyatakan bahwa Skripsi dengan judul berikut:

Penerapan Seleksi Fitur *Analysis of Variance* pada Algoritma *Random Forest Classifier* dalam Klasifikasi Nilai Mahasiswa Studi Kasus: Universitas Muhammadiyah Kalimantan Timur

Dosen Pembimbing : Wawan Joko Pranoto, S.Kom, M.Ti

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar Sarjana Komputer, baik di Universitas Muhammadiyah Kalimantan Timur (UMKT) maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab UMKT.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi Muhammadiyah.

Samarinda, 16 Juli 2023

Menyatakan


Muhammad Fath Thoriq

1911102441166

PRAKATA



Alhamdulillah puji syukur kepada Allah SWT berkat Rahmat dan Karunia-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul, “Penerapan Seleksi Fitur *Analysis of Variance* pada Algoritma *Random Forest Classifier* dalam Klasifikasi Nilai Mahasiswa Studi Kasus: Universitas Muhammadiyah Kalimantan Timur”. Adapun tujuan dari penulisan skripsi ini adalah untuk memenuhi salah satu persyaratan penyelesaian studi di jurusan Teknik Informatika Fakultas Sains & Teknologi Universitas Muhammadiyah Kalimantan Timur.

Penulis menyadari bahwa proses penulisan skripsi ini tidak akan berjalan dengan sebagaimana mestinya tanpa adanya dukungan dari berbagai pihak. Oleh karena itu, penulis menyampaikan ucapan terima kasih dan rasa hormat kepada:

1. Bapak Achmad Thoyieb dan Ibu Lilik Herawati selaku orang tua penulis yang sudah mendukung hingga pada titik ini.
2. Rafli Achmad Baihaqi, Renatha Dyah Syazintha, dan Keenan Zuhdi Ahmad Rayshi selaku saudara yang sudah mendukung.
3. Bapak Wawan Joko Pranoto, S.Kom, M.Kom selaku dosen pembimbing yang telah meluangkan waktu, tenaga, bimbingan dan sarannya kepada penulis.
4. Ibu Asslia Johar Latipah, M.Cs selaku Ketua Program Studi Teknik Informatika di Universitas Muhammadiyah Kalimantan Timur.
5. Bapak Prof. Ir. Sarjito, M.T., Ph.D., selaku Dekan Fakultas Sains dan Teknologi Universitas Muhammadiyah Kalimantan Timur.
6. Bapak Prof. Dr. H. Bambang Setiaji, selaku Rektor Universitas Muhammadiyah Kalimantan Timur.
7. Seluruh Dosen baik dari Program Studi Teknik Informatika maupun dosen dari Program Studi lain yang telah memberikan ilmunya kepada penulis.

8. Kepada teman-teman seperjuangan satu bimbingan yaitu Putra Saymen, Rahmat Ramadhani, Suci Mawaddah, Kamirul Hakim, Septia Intan Permata Sari, dan Yuliana Dilla Evtasari yang telah banyak membantu dan memberikan semangat dalam menyelesaikan skripsi ini.
9. Teman-teman Program Studi Teknik Informatika angkatan 2019 dan dari Program studi lain yang namanya tidak bisa disebutkan satu persatu.

Penulis menyadari skripsi ini masih jauh dari kata sempurna, oleh karena itu dengan segala hormat dan kerendahan hati, penulis berharap adanya saran dan kritik yang membangun untuk menyempurnakan hasil dari skripsi ini sehingga dapat bermanfaat untuk pihak-pihak yang memerlukan.

Samarinda, 24 Maret 2023



Muhammad Fath Thoriq

ABSTRAK

Dalam lembaga pendidikan seperti universitas, penting untuk memperhatikan kinerja mahasiswa agar mereka dapat menyelesaikan studi mereka tepat waktu. Namun, masih terdapat masalah di mana sebagian mahasiswa tidak mampu menyelesaikan studi mereka dengan tepat waktu, bahkan ada yang memutuskan untuk berhenti atau tidak lagi aktif sebagai mahasiswa. Hal ini diperlihatkan oleh penurunan nilai mahasiswa pada mata kuliah Bahasa Indonesia dari angkatan 2020 hingga 2021 di UMKT. Untuk mengatasi masalah ini, diperlukan metode yang dapat mengukur kinerja mahasiswa dalam menyelesaikan studi mereka. Penelitian ini bertujuan untuk mengetahui atribut-atribut yang berpengaruh terhadap penurunan nilai mahasiswa dalam mata kuliah Bahasa Indonesia, serta meningkatkan akurasi algoritma *Random Forest Classifier* dengan menggunakan seleksi fitur *ANOVA*. Data yang digunakan dalam penelitian ini adalah data mahasiswa UMKT yang mengambil mata kuliah Bahasa Indonesia pada periode 2020/2021 hingga 2021/2022. Data diperoleh dari Bagian Administrasi Akademik (BAA) UMKT dan unit Mata Kuliah Dasar Umum (MKDU) UMKT dengan jumlah data sebanyak 1028. Proses analisis data dilakukan menggunakan metode *5-Fold Cross Validation*. Hasil penelitian menunjukkan bahwa atribut-atribut seperti *Progress*, *% Course completed*, Tugas 1, dan Tugas 2 memiliki pengaruh signifikan terhadap penurunan nilai mahasiswa dalam mata kuliah Bahasa Indonesia. Selain itu, penggunaan seleksi fitur *ANOVA* pada algoritma *Random Forest Classifier* mampu meningkatkan kinerja algoritma tersebut dengan akurasi meningkat dari 85.65% menjadi 87.31%.

Kata Kunci: Nilai mahasiswa, *Data Mining*, *Random Forest*, *5-Fold Cross Validation*, *ANOVA*

ABSTRACT

In educational institutions, such as universities, it is important to pay attention to the performance of students so that they can complete their studies on time. However, there are still issues where some students are unable to finish their studies on time, and some even decide to drop out or become inactive as students. This is evidenced by the decline in student grades in the Indonesian Language course from the 2020 to 2021 cohorts at UMKT. To address this problem, a method is needed to measure students' performance in completing their studies. This study aims to identify the attributes that influence the decline in student grades in the Indonesian Language course, as well as improve the accuracy of the Random Forest Classifier algorithm using ANOVA feature selection. The data used in this study consists of UMKT student data who took the Indonesian Language course from the 2020/2021 to 2021/2022 academic years. The data was obtained from the academic administration department (BAA) of UMKT and the General Basic Course Unit (MKDU) of UMKT, with a total of 1028 data points. The data analysis process was conducted using the 5-Fold Cross Validation method. The results of the study indicate that attributes such as Progress, % Course completed, Assignment 1, and Assignment 2 have a significant influence on the decline in student grades in the Indonesian Language course. Furthermore, the use of ANOVA feature selection in the Random Forest Classifier algorithm improved its performance, with the accuracy increasing from 85.65% to 87.31%.

Keywords: *Student Grades, Data Mining, Random Forest, 5-Fold Cross Validation, ANOVA*

DAFTAR ISI

COVER	i
HALAMAN PENGESAHAN	iii
SURAT PERNYATAAN KEASLIAN SKRIPSI	iv
PRAKATA	v
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR	xiii
DAFTAR LAMPIRAN	xiv
BAB 1	1
PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	2
1.3. Tujuan Penelitian	2
1.4. Batasan Masalah	2
BAB 2	3
TINJAUAN PUSTAKA	3
2.1. Kinerja Mahasiswa	3
2.2. <i>Data Mining</i>	3
2.2.1. Algoritma Klasifikasi	4
2.2.2. <i>Random Forest</i>	4
2.2.3. <i>Data Preprocessing</i>	5
2.2.4. <i>K-Fold Cross Validation</i>	5
2.2.5. <i>Confusion Matrix</i>	6
2.2.6. CRISP-DM	7
2.2.7. <i>Analysis of Variance (ANOVA)</i>	8
2.3. Penelitian Terdahulu	9
BAB 3	12

METODOLOGI PENELITIAN	12
3.1. Objek Penelitian	12
3.2. Metode Pengumpulan Data	12
3.3. Teknik Analisis Data	13
3.4. Peralatan yang Digunakan	18
3.5. Jadwal Penelitian	18
BAB 4	20
HASIL DAN PEMBAHASAN.....	20
4.1. <i>Business Understanding</i> (Pemahaman Bisnis)	20
4.2. <i>Data Understanding</i> (Pemahaman Data)	20
4.3. <i>Data Preparation</i> (Persiapan Data)	25
4.3.1. Seleksi dan Integrasi Data	25
4.3.2. Pembersihan Data	25
4.3.3. Transformasi Data	26
4.3.4. Reduksi Data.....	29
4.4. <i>Modeling</i> (Pemodelan)	30
4.4.1. Perhitungan Algoritma <i>Random Forest</i>	30
4.4.2. Implementasi <i>Random Forest</i> pada <i>Python</i>	54
4.4.3. Implementasi <i>Random Forest</i> dengan <i>ANOVA</i> pada <i>Python</i>	58
4.5. <i>Evaluation</i> (Evaluasi)	61
BAB 5	63
PENUTUP.....	63
5.1. Kesimpulan	63
5.2. Saran	63
DAFTAR PUSTAKA.....	64
LAMPIRAN	68

DAFTAR TABEL

Tabel 2. 1 Confusion Matrix	6
Tabel 2. 2 Penelitian Terdahulu	9
Tabel 3. 1 Atribut Data MKDU.....	14
Tabel 3. 2 Atribut Data BAA	14
Tabel 3. 3 Atribut Seleksi dan Integrasi Data	15
Tabel 3. 4 Jadwal Penelitian	18
Tabel 4. 1 Data yang diperoleh dari BAA UMKT	20
Tabel 4. 2 Data yang diperoleh dari MKDU.....	24
Tabel 4. 3 Hasil Seleksi dan Integrasi Data.....	25
Tabel 4. 4 Transformasi Atribut Jenis Kelamin.....	27
Tabel 4. 5 Transformasi Atribut Time spent on course.....	27
Tabel 4. 6 Transformasi Atribut Simbol Tahap 1.....	28
Tabel 4. 7 Transformasi Atribut Simbol Tahap 2.....	28
Tabel 4. 8 Hasil dari Transformasi pada integrasi data.....	29
Tabel 4. 9 Hasil akhir dari Data Preparation	30
Tabel 4. 10 10 Sampel acak dari dataset.....	31
Tabel 4. 11 Hasil Perhitungan Gini pada atribut Time spent on course	33
Tabel 4. 12 Hasil Perhitungan Gini pada atribut Progress	35
Tabel 4. 13 Hasil Perhitungan Gini pada atribut % Course Completed	38
Tabel 4. 14 Hasil Perhitungan Gini pada atribut Comments.....	40
Tabel 4. 15 Hasil Perhitungan Gini pada atribut Tugas 1.....	42
Tabel 4. 16 Hasil Perhitungan Gini pada atribut Tugas 2.....	45
Tabel 4. 17 Hasil Perhitungan Gini pada atribut Tugas 3.....	47
Tabel 4. 18 Hasil Perhitungan Gini pada atribut UTS.....	49
Tabel 4. 19 Hasil Perhitungan Gini Impurity untuk Root Node.....	50
Tabel 4. 20 Hasil Perhitungan Gini Impurity untuk Cabang Kiri.....	52
Tabel 4. 21 Hasil Prediksi Dataset Sampel	53
Tabel 4. 22 Hasil Evaluasi Confusion Matrix Dataset Sampel	54
Tabel 4. 23 Hasil Akurasi tiap fold	61

Tabel 4. 24 Hasil Akurasi rata-rata dari fold..... 61

DAFTAR GAMBAR

Gambar 2. 1 K-Fold Cross Validation.....	6
Gambar 2. 2 CRISP-DM.....	7
Gambar 3. 1 Alur Analisis Data.....	13
Gambar 4. 1 Nilai kosong pada Tugas 3	26
Gambar 4. 2 Mengecek kembali nilai kosong	26
Gambar 4. 3 Hasil Reduksi Data	29
Gambar 4. 4 Hasil Penentuan Root Node	51
Gambar 4. 5 Hasil Penentuan Cabang Kanan	51
Gambar 4. 6 Hasil Penentuan Cabang Kiri	52
Gambar 4. 7 Hasil Penentuan Pohon Keputusan CART	53
Gambar 4. 8 Hasil Import Dataset.....	55
Gambar 4. 9 Import Dataset	55
Gambar 4. 10 Pemodelan Random Forest.....	56
Gambar 4. 11 Pemodelan Random Forest Lanjutan.....	57
Gambar 4. 12 Hasil Pemodelan Random Forest	57
Gambar 4. 13 Hasil Pohon.....	58
Gambar 4. 14 Hasil Seleksi Fitur.....	59
Gambar 4. 15 Menghapus atribut yang kurang berpengaruh.....	59
Gambar 4. 16 Hasil Pemodelan Random Forest setelah seleksi fitur	60
Gambar 4. 17 Hasil Pohon setelah seleksi fitur	60

DAFTAR LAMPIRAN

Lampiran 1 Data Mata Kuliah Bahasa Indonesia yang diperoleh dari BAA	69
Lampiran 2 Data Mata Kuliah Bahasa Indonesia yang diperoleh dari MKDU.....	71
Lampiran 3 Data BAA setelah melalui Tahapan Data Selection.....	73
Lampiran 4 Data MKDU setelah melalui Tahapan Data Selection.....	74
Lampiran 5 Data setelah melalui Tahap Integrasi dan Transformasi Data.....	75
Lampiran 6 Seluruh Data setelah melalui Tahap Data Reduction	76
Lampiran 7 Source Code	77
Lampiran 8 Surat Izin Pengambilan Data MKDU.....	87
Lampiran 9 Surat Izin Pengambilan Data BAA	88
Lampiran 10 Surat Keterangan Tidak Uji Validitas.....	89
Lampiran 11 Lembar Bimbingan	90
Lampiran 12 Hasil Uji Plagiasi	92
Lampiran 13 Riwayat Hidup	94