

**PREDIKSI PEMINATAN MAHASISWA DENGAN
ALGORITMA BERBASIS *MACHINE LEARNING* (STUDI
KASUS: SISTEM INFORMASI UNIVERSITAS MULTIMEDIA
NUSANTARA)**



SKRIPSI

**Rendy Wirawan Tamrin
00000053580**

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2025**

**PREDIKSI PEMINATAN MAHASISWA DENGAN
ALGORITMA BERBASIS *MACHINE LEARNING* (STUDI
KASUS: SISTEM INFORMASI UNIVERSITAS MULTIMEDIA
NUSANTARA)**



SKRIPSI

Diajukan sebagai Salah Satu Syarat untuk Memperoleh

Gelar Sarjana Komputer

Rendy Wirawan Tamrin

00000053580

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG**

2025

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Rendy Wirawan Tamrin

Nomor Induk Mahasiswa : 00000053580

Program Studi : Sistem Informasi

Skripsi dengan judul:

“Prediksi Peminatan Mahasiswa dengan Algoritma Berbasis *Machine Learning* (Studi Kasus: Sistem Informasi Universitas Multimedia Nusantara)”

Merupakan hasil karya saya sendiri bukan plagiat dari laporan karya tulis ilmiah yang ditulis oleh orang lain, dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan maupun dalam penulisan laporan karya tulis ilmiah, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk mata kuliah yang telah saya tempuh.

Tangerang, 4 Juni 2025



Rendy Wirawan Tamrin

HALAMAN PENGESAHAN

Skripsi dengan judul

Prediksi Peminatan Mahasiswa dengan Algoritma Berbasis *Machine Learning* (Studi Kasus: Sistem Informasi Universitas Multimedia Nusantara)

Oleh

Nama : Rendy Wirawan Tamrin

NIM : 00000053580

Program Studi : Sistem Informasi

Fakultas : Teknik dan Informatika

Telah diujikan pada hari Senin, 16 Juni 2025

Pukul 15.00 s.d 17.00 dan dinyatakan

LULUS

Dengan susunan penguji sebagai berikut.

Ketua Sidang

Penguji



30/06/2025

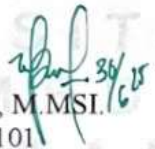
Ahmad Faza, S.Kom., M.T.I.
0312019501



26/6'25

Ririn Ikana Desanti, S.Kom., M.Kom
0313058001

Dosen Pembimbing



Wella, S.Kom., M.MSI.
0305119101

Ketua Program Studi Sistem Informasi



30/6'25

Ririn Ikana Desanti, S.Kom., M.Kom
0313058001

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Yang bertanda tangan di bawah ini:

Nama : Rendy Wirawan Tamrin

NIM : 0000053580

Program Studi : Sistem Informasi

Jenjang : S1

Judul Karya Ilmiah :

Prediksi Peminatan Mahasiswa dengan Algoritma Berbasis *Machine Learning* (Studi Kasus: Sistem Informasi Universitas Multimedia Nusantara)

Menyatakan dengan sesungguhnya bahwa saya bersedia* (pilih salah satu):

- ☒ Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya ke dalam repositori Knowledge Center sehingga dapat diakses oleh Sivitas Akademika UMN/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial.
- ☐ Saya tidak bersedia mempublikasikan hasil karya ilmiah ini ke dalam repositori Knowledge Center, dikarenakan: dalam proses pengajuan publikasi ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*) **.
- ☐ Lainnya, pilih salah satu:
 - ☐ Hanya dapat diakses secara internal Universitas Multimedia Nusantara
 - ☐ Embargo publikasi karya ilmiah dalam kurun waktu 3 tahun.

Tangerang, 4 Juni 2025



Rendy Wirawan Tamrin

* Pilih salah satu

** Jika tidak bisa membuktikan LoA jurnal/HKI, saya bersedia mengizinkan penuh karya ilmiah saya untuk dipublikasikan ke KC UMN dan menjadi hak institusi UMN.

KATA PENGANTAR

Puji syukur kepada Tuhan Yang Maha Esa atas rahmat dan karunianya yang tiada henti sehingga skripsi yang berjudul “Prediksi Peminatan Mahasiswa dengan Algoritma Berbasis *Machine Learning* (Studi Kasus: Sistem Informasi Universitas Multimedia Nusantara)” dapat terselesaikan dengan baik dan tepat waktu. Skripsi ini ditulis sebagai bentuk pemenuhan syarat kelulusan studi Strata 1 (S1) dan perolehan gelar sarjana pada Program Studi Sistem Informasi, Fakultas Teknik dan Informatika, Universitas Multimedia Nusantara

Penulisan skripsi ini tidak akan terselesaikan dengan baik dan tepat waktu tanpa adanya bimbingan, dorongan, dan bantuan dari berbagai pihak. Untuk itu penulis mengucapkan terima kasih kepada:

1. Bapak Dr. Andrey Andoko, selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika, Universitas Multimedia Nusantara.
3. Ibu Ririn Ikana Desanti, S.Kom., M.Kom., selaku Ketua Program Studi Sistem Informasi, Universitas Multimedia Nusantara.
4. Ibu Wella, S.Kom., M.MSI., selaku Dosen Pembimbing yang telah memberikan bimbingan, arahan, dan motivasi sehingga terselesainya tugas akhir ini.
5. Keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tugas akhir ini.

Semoga hasil penulisan karya ilmiah ini dapat memberikan manfaat dan dampak positif bagi para pembaca.

Tangerang, 4 Juni 2025



Rendy Wirawan Tamrin

**PREDIKSI PEMINATAN MAHASISWA DENGAN
ALGORITMA BERBASIS *MACHINE LEARNING* (STUDI
KASUS: SISTEM INFORMASI UNIVERSITAS MULTIMEDIA
NUSANTARA)**

Rendy Wirawan Tamrin

ABSTRAK

Pemilihan mata kuliah peminatan merupakan tahapan penting dalam proses akademik mahasiswa Program Studi Sistem Informasi Universitas Multimedia Nusantara (UMN). Berdasarkan survei terhadap 125 mahasiswa Sistem Informasi UMN angkatan 2024 menunjukkan bahwa sebanyak 90% responden merasa kesulitan dalam menentukan peminatan yang tepat. Kesulitan ini dipicu oleh minimnya informasi dan panduan terkait mata kuliah peminatan yang tersedia, kurangnya pemahaman terhadap hubungan antara mata kuliah dasar dan peminatan, serta ketidakjelasan prospek kerja pasca kelulusan. Untuk mengatasi permasalahan tersebut, penelitian ini bertujuan mengembangkan model prediksi yang mampu merekomendasikan peminatan mahasiswa berdasarkan performa akademik.

Model dikembangkan menggunakan metode CRISP-ML dengan menerapkan algoritma Decision Tree, Random Forest, dan XGBoost. Untuk meningkatkan performa model, digunakan teknik SMOTE untuk mengatasi ketidakseimbangan data, *hyperparameter tuning* untuk optimasi, dan *feature selection* untuk menyaring fitur yang paling relevan. Selain itu, dilakukan eksperimen terhadap segmentasi data berdasarkan angkatan guna mengetahui pengaruhnya terhadap performa model. Penelitian ini juga mengidentifikasi variabel akademik yang paling berpengaruh dalam proses prediksi, seperti nilai mata kuliah yang relevan dengan kompetensi tiap peminatan.

Hasil eksperimen menunjukkan bahwa algoritma Random Forest menghasilkan performa terbaik dengan akurasi sebesar 78,08% pada data mahasiswa angkatan 2021-2022 setelah penerapan SMOTE dan *hyperparameter tuning*. Hasil ini membuktikan bahwa segmentasi data berdasarkan angkatan terbukti meningkatkan performa model secara signifikan. Dengan demikian, sistem ini diharapkan dapat menjadi solusi alternatif dalam mendukung proses pengambilan keputusan akademik yang lebih terarah bagi mahasiswa Sistem Informasi Universitas Multimedia Nusantara.

Kata kunci: Decision Tree, Prediksi Peminatan, Random Forest, XGBoost

**STUDENT SPECIALIZATION PREDICTION USING
MACHINE LEARNING-BASED ALGORITHMS
(CASE STUDY: INFORMATION SYSTEMS DEPARTMENT
AT UNIVERSITAS MULTIMEDIA NUSANTARA)**

Rendy Wirawan Tamrin

ABSTRACT (English)

Selecting academic specializations is a crucial step in the academic journey of Information Systems students at Universitas Multimedia Nusantara (UMN). A survey of 125 UMN Information Systems students from the 2024 cohort showed that 90% of respondents experienced difficulties in choosing the right concentration. These challenges arise from limited information and guidance regarding concentration courses, a lack of understanding of the relationship between core subjects and specializations, and unclear post-graduation career prospects. To address this issue, this study aims to develop a predictive model that can recommend suitable concentrations for students based on their academic performance.

The model was developed using the CRISP-ML methodology and implemented using classification algorithms including Decision Tree, Random Forest, and XGBoost. To enhance model performance, several techniques were applied: SMOTE to address class imbalance, hyperparameter tuning for optimization, and feature selection to identify the most relevant academic features. Additionally, data segmentation by student cohort was conducted to examine its impact on model performance. This study also identified the most influential academic variables in the prediction process, such as course grades related to the competencies required in each concentration.

Experimental results show that the Random Forest algorithm achieved the best performance with an accuracy of 78.08% on data from the 2021–2022 student cohorts after applying SMOTE and hyperparameter tuning. These findings demonstrate that data segmentation by cohort significantly improves model performance. Therefore, the proposed system is expected to serve as an alternative solution to support more targeted academic decision-making for Information Systems students at Universitas Multimedia Nusantara.

Keywords: *Decision Tree, Random Forest, Specialization Prediction, XGBoost*

DAFTAR ISI

HALAMAN PERNYATAAN TIDAK PLAGIAT	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	iv
KATA PENGANTAR.....	v
ABSTRAK	vi
<i>ABSTRACT (English)</i>	vii
DAFTAR ISI.....	viii
DAFTAR TABEL	x
DAFTAR GAMBAR.....	xi
DAFTAR RUMUS	xii
DAFTAR LAMPIRAN	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	5
1.3 Batasan Masalah	5
1.4 Tujuan dan Manfaat Penelitian.....	6
1.4.1 Tujuan Penelitian.....	6
1.4.2 Manfaat Penelitian.....	6
1.5 Sistematika Penulisan	6
BAB II LANDASAN TEORI	8
2.1 Penelitian Terdahulu	8
2.2 Tinjauan Teori.....	12
2.2.1 Sistem Rekomendasi	12
2.2.2 Machine Learning.....	13
2.2.2.1 Decision Tree	14
2.2.2.2 Random Forest	15
2.2.2.3 XGBoost	16
2.3 Framework yang digunakan	16
2.3.1 <i>Cross-Industry Standard Process for Machine Learning (CRISP-ML)</i>	16

2.3.2	<i>Cross-Industry Standard Process for Data Mining (CRISP-DM)</i>	18
2.4	Teori tentang tools/software yang digunakan	20
2.4.1	Visual Studio Code	21
2.4.2	Python	21
2.4.3	Streamlit	21
BAB III	METODOLOGI PENELITIAN	23
3.1	Gambaran Umum Objek Penelitian	23
3.2	Metode Penelitian	23
3.3	Teknik Pengumpulan Data	28
3.4	Variabel Penelitian	28
3.5	Teknik Analisis Data	29
BAB IV	ANALISIS DAN HASIL PENELITIAN	30
4.1	<i>Business & Data Understanding</i>	30
4.2	Data Engineering (Data Preparation)	32
4.3	Machine Learning Model Engineering	35
4.4	Evaluation	42
4.5	Deployment	44
4.6	Monitoring and Maintenance	46
4.7	Hasil dan Diskusi	49
4.7.1	Teknik SMOTE	50
4.7.2	Teknik SMOTE + <i>Hyperparameter Tuning</i>	51
4.7.3	Teknik SMOTE + <i>Hyperparameter Tuning + Feature Selection</i>	52
4.7.4	Performa Model Berdasarkan Segmentasi Angkatan	52
4.7.5	Hasil dan Temuan Penelitian	54
4.7.6	Perbandingan dengan Penelitian Terdahulu	56
BAB V	SIMPULAN DAN SARAN	61
5.1	Simpulan	61
5.2	Saran	62
DAFTAR PUSTAKA		65
LAMPIRAN		68

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu	8
Tabel 3.1 Perbandingan <i>framework</i> CRISP-DM dan CRISP-ML	23
Tabel 4.1 Data Angkatan 2023 Semester 1	30
Tabel 4.2 Data Angkatan 2023 Semester 1	31
Tabel 4.3 Data Angkatan 2023 Semester 1	31
Tabel 4.4 Data Angkatan 2023 Semester 1	31
Tabel 4.5 Hasil Modeling dengan Data Angkatan 2021-2023	35
Tabel 4.6 Hasil Modeling dengan Teknik SMOTE	36
Tabel 4.7 Hasil Modeling dengan Teknik SMOTE & <i>Hyperparameter Tuning</i> ..	37
Tabel 4.8 Hasil Modeling dengan Teknik SMOTE & <i>Hyperparameter Tuning</i> dan <i>Feature Selection</i> berdasarkan <i>Feature Importance</i>	38
Tabel 4.9 Hasil Modeling dengan Data Angkatan 2021-2022	40
Tabel 4.10 Hasil Modeling dengan Data Angkatan 2023	41
Tabel 4.11 Rencana Proses <i>Monitoring & Maintenance</i>	47
Tabel 4.12 Hasil Evaluasi Model Sebelum SMOTE	50
Tabel 4.13 Hasil Evaluasi Model Sesudah SMOTE	50
Tabel 4.14 Akurasi Model Setelah SMOTE dan <i>Hyperparameter Tuning</i>	51
Tabel 4.15 Akurasi Model Setelah Penambahan <i>Feature Selection</i>	52
Tabel 4.16 Akurasi Model pada Angkatan 2021-2022	53
Tabel 4.17 Akurasi Model pada Angkatan 2023	53
Tabel 4.18 Tabel <i>Summary</i> Eksperimen	54
Tabel 4.19 Perbandingan Akurasi Model dengan Penelitian Terdahulu	56

DAFTAR GAMBAR

Gambar 1.1. Hasil Survey Tingkat Kesulitan Mahasiswa dalam Pemilihan Peminatan.....	2
Gambar 1.2 Hasil Survey Faktor yang Memengaruhi Kesulitan Mahasiswa dalam Memilih Peminatan	3
Gambar 2.1 Siklus Tahapan Framework CRISP-ML [24]	17
Gambar 2.2 Siklus Tahapan <i>Framework</i> CRISP-DM [25].....	19
Gambar 4.1 Penggabungan Data Mahasiswa Angkatan 2023 Semester 1 & 2	32
Gambar 4.2 Penggabungan Data Mahasiswa Angkatan 2023 Semester 1 & 2 dengan Data Peminatan Mahasiswa	33
Gambar 4.3 Pivot Tabel dan Transformasi Data.....	34
Gambar 4.4 Persiapan Data Angkatan 2021-2022	34
Gambar 4.5 Penggabungan Data Angkatan 2021-2023	35
Gambar 4.6 <i>Feature Importance</i>	38
Gambar 4.7 Data Mahasiswa Angkatan 2021-2022	40
Gambar 4.8 Feature Importance Data Angkatan 2021-2022	41
Gambar 4.9 Aplikasi Streamlit untuk Prediksi Peminatan Mahasiswa SI UMN..	45



DAFTAR RUMUS

Rumus 2.1 Formula Perhitungan Entropy.....	15
Rumus 2.2 Formula Perhitungan Gain.....	15
Rumus 2.3 Formula Perhitungan Gini	15
Rumus 2.4 Formula Perhitungan Algoritma XGBoost.....	16



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR LAMPIRAN

Lampiran A Turnitin Similarity Report	68
Lampiran B Form Konsultasi Bimbingan	78
Lampiran C Uji Validitas dan Realibilitas Kuesioner.....	79



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA