

Tesis

Pemanfaatan Machine Learning untuk Memprediksi Dropout

Mahasiswa di STIKes Panti Rapih Yogyakarta



Asep Setiyono

235312541

MAGISTER INFORMATIKA

FAKULTAS TEKNOLOGI INDUSTRI

UNIVERSITAS ATMA JAYA YOGYAKARTA

2025

Pemanfaatan Machine Learning untuk Memprediksi Dropout

Mahasiswa di STIKes Panti Rapih Yogyakarta



Asep Setiyono

235312541

MAGISTER INFORMATIKA

FAKULTAS TEKNOLOGI INDUSTRI

UNIVERSITAS ATMA JAYA YOGYAKARTA

2025

HALAMAN PENGESAHAN

Tugas Akhir Berjudul

PEMANFAATAN MACHINE LEARNING UNTUK MEMPREDIKSI DROPOUT MAHASISWA DI STIKES PANTI
RAPIH YOGYAKARTA

yang disusun oleh

Asep Setiyono

235312541

dinyatakan telah memenuhi syarat pada tanggal 24 Januari 2025

		Keterangan
Dosen Pembimbing 1	: Prof. Dr. Ir. Alb. Joko Santoso, M.T.	Telah Menyetujui
Dosen Pembimbing 2	: Dra. Ernawati, M.T., Ph.D	Telah Menyetujui
Tim Penguji		
Penguji 1	: Prof. Dr. Ir. Alb. Joko Santoso, M.T.	Telah Menyetujui
Penguji 2	: Paulus Mudjihartono, S.T., M.T., Ph.D	Telah Menyetujui

Yogyakarta, 24 Januari 2025

Universitas Atma Jaya Yogyakarta

Teknologi Industri

Dekan

ttt.

Dr. Ir. Parama Kartika Dewa SP., S.T., M.T.

Dokumen ini merupakan dokumen resmi UAJY yang tidak memerlukan tanda tangan karena dihasilkan secara elektronik oleh Sistem Bimbingan UAJY. UAJY bertanggung jawab penuh atas informasi yang tertera di dalam dokumen ini

HALAMAN PERNYATAAN

Saya yang bertanda tangan dibawah ini:

Nama Lengkap : Asep Setiyono

NPM : 235312541

Program Studi : Informatika

Judul Penelitian : Pemanfaatan Machine Learning untuk Memprediksi Dropout
Mahasiswa di STIKes Panti Rapih Yogyakarta

Dengan ini menyatakan bahwa penelitian ini merupakan hasil karya penulis sendiri dan bukan merupakan hasil duplikasi dari penelitian sebelumnya. Penelitian terdahulu hanya menjadi referensi bagi penulis dan sudah disertakan acuan dan terlampir dalam daftar pustaka.

Yogyakarta, 22 Januari 2025

Yang menyatakan

Asep Setiyono

ABSTRAK

Penelitian ini bertujuan untuk menguji efektivitas berbagai algoritma *machine learning* dalam memprediksi kemungkinan *dropout* mahasiswa. Tingkat *dropout* yang tinggi merupakan masalah serius di banyak institusi pendidikan, mempengaruhi kinerja institusi dan masa depan mahasiswa. Empat algoritma utama *Random Forest*, KNN, *Naive Bayes*, dan SVM telah diimplementasikan untuk mengatasi masalah ini, menggunakan dataset yang mencakup berbagai fitur mahasiswa yang relevan.

Pendekatan-pendekatan *machine learning* diuji dalam tiga kondisi berbeda: tanpa modifikasi, dengan *hyperparameter tuning*, dan dengan penambahan *noise*, untuk menilai ketahanan dan efektivitas mereka dalam berbagai skenario. Evaluasi kinerja dilakukan menggunakan metrik seperti akurasi, *precision*, *recall*, *F1-score*, serta analisis ROC dan AUC. Analisis *feature importance* juga dilakukan untuk menentukan fitur-fitur yang memiliki pengaruh signifikan dalam prediksi.

Random Forest secara konsisten menunjukkan kinerja terbaik dalam semua kondisi, mencapai *Mean Accuracy* sekitar 0.99725 tanpa modifikasi, 0.99718 dengan *hyperparameter tuning*, dan 0.99127 dengan penambahan *noise*. Analisis distribusi fitur melalui *boxplot* memberikan pemahaman lebih lanjut tentang sebaran nilai dan keberadaan outlier, yang membantu dalam menyempurnakan algoritma prediksi. Penelitian ini memberikan kontribusi penting dalam bidang prediksi *dropout* mahasiswa dengan mengidentifikasi algoritma dan fitur yang paling efektif untuk prediksi. Saran yang diberikan berdasarkan hasil analisis dapat membantu institusi pendidikan dalam merancang intervensi yang lebih tepat guna mengurangi tingkat *dropout*.

Kata kunci: Machine learning, dropout mahasiswa, Feature importance

ABSTRACT

This study aims to test the effectiveness of various machine learning algorithms in predicting the probability of student dropout. The high dropout rate is a significant problem in many educational institutions, impacting both institutional performance and the future of students. Four major algorithms Random Forest, KNN, Naive Bayes, and SVM were implemented to address this issue, utilizing datasets that encompass various relevant student features.

The machine learning approaches were evaluated under three different conditions: without modification, with hyperparameter tuning, and with noise addition, to assess their robustness and effectiveness in diverse scenarios. Performance evaluation was conducted using metrics such as accuracy, precision, recall, F1-score, and ROC and AUC analysis. Additionally, feature importance analysis was performed to identify the features that significantly influence predictions.

Random Forest consistently demonstrated the best performance across all conditions, achieving a Mean Accuracy of approximately 0.99725 without modification, 0.99718 with hyperparameter tuning, and 0.99127 with noise addition. The analysis of feature distribution through boxplots provided further insight into value distributions and the presence of outliers, aiding in the refinement of the prediction model. This study contributes significantly to the field of student dropout prediction by identifying the most effective algorithms and features for accurate predictions. The recommendations based on the analysis results can assist educational institutions in designing more effective interventions to reduce dropout rates.

Keywords: Machine learning, student dropout, feature importance

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Tuhan Yesus Kristus yang telah memberikan bimbingan dan kekuatan selama proses penelitian dan penyusunan laporan tesis ini, sehingga penulis dapat menyelesaikannya dengan baik. Mahasiswa pascasarjana Universitas Atma Jaya Yogyakarta yang ingin menempuh program studi Magister Informatika di Fakultas Teknologi Industri diwajibkan untuk menyelesaikan tesis ini guna memperoleh gelar Magister Komputer. Penulis berharap penelitian ini dapat dikembangkan lebih lanjut dan memberikan manfaat bagi semua pihak yang terlibat.

Selama pelaksanaan penelitian ini, penulis menerima banyak bantuan dan saran dari berbagai pihak, baik secara langsung maupun tidak langsung. Dengan penuh rasa hormat, penulis mengucapkan terima kasih kepada:

1. Bapak Yonathan Dri Handarkho ST., M.Eng., Ph.D., selaku Ketua Program Studi Magister Informatika.
2. Prof. Dr. Ir. Alb. Joko Santoso, M.T., selaku dosen pembimbing satu yang telah meluangkan banyak waktu untuk membimbing penulis selama melaksanakan penelitian.
3. Ibu Dra. Ernawati, M.T., Ph.D yang telah memberikan arahan dan bimbingan sehingga dapat menyelesaikan penelitian ini.
4. Bapak Paulus Mudjihartono, S.T., M.T., Ph.D selaku dosen penguji yang telah memberikan saran dan masukan yang sangat berarti untuk tesis ini.
5. Dosen-dosen di Universitas Atma Jaya Yogyakarta yang telah memberikan arahan dan bimbingan kepada penulis selama menjalani program studi Magister Informatika.
6. Keluarga saya (Bapak, Alm. Ibu, Ibu Mertua, kakak dan adik saya yang selalu memberikan doa dan dukungan penuh dalam menyelesaikan tesis ini.
7. Istri saya Bernardia Prastiwi Haryuningsih dan juga anak saya Dorothea Adara Garvita yang selalu semangat, motivasi dan kasih sayang tiada hentinya.
8. Teman-teman Magister Informatika Angkatan 2023: Damba, Ian, Rio, Hendri, Laura, Elza dan Tesa yang selalu kompak dan selalu memberikan dukungan dalam menyelesaikan tesis ini.

Demikian saya ucapkan terima kasih, saya menyadari bahwa dalam penulisan Tesis ini masih terdapat banyak kekurangan karena itu segala kritik dan saran yang membangun akan menyempurnakan penulisan Tesis ini serta bermanfaat bagi penulis dan para pembaca.

Yogyakarta, Januari 2025

Asep Setiyono

DAFTAR ISI

HALAMAN PENGESAHAN	ii
HALAMAN PERNYATAAN TESIS	iii
ABSTRAK.....	iv
ABSTRACT	v
KATA PENGANTAR	vi
DAFTAR ISI	viii
DAFTAR GAMBAR.....	xi
DAFTAR TABEL	xii
BAB 1 PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah.....	3
1.3. Pertanyaan Penelitian.....	3
1.4. Batasan Masalah	4
1.5. Tujuan dan Manfaat Penelitian	4
BAB 2 TINJAUAN PUSTAKA	5
BAB 3 LANDASAN TEORI	15
3.1 Dropout.....	15
3.2 Machine Learning	16
3.3 Prediksi	19
3.4 Klasifikasi	19
3.5 Random Forest.....	19
3.6 Naive Bayes	20
3.7 K-Nearest Neighbor	20

3.8	Support Vector Machine	21
3.9	Hyperparameter tuning	22
3.10	K-fold cross validation	22
3.11	Confusion Matrix, Precision, Recall, dan F1-Score	22
3.12	ROC Curve dan AUC	23
3.13	Teknik Imputasi, Normalisasi, Standarisasi, dan Encoding	23
3.14	Oversampling dan Undersampling	24
BAB 4 METODOLOGI PENELITIAN		25
4.1.	Tahapan Penelitian	25
4.1.1	Perumusan Masalah	26
4.1.2	Studi Pustaka	26
4.1.3	Pengumpulan Data	26
4.1.4	Pra Processing Data	26
4.1.5	Pembagian Training dan Testing	28
4.1.6	Penerapan Algoritma	29
4.1.7	Analisis Hasil	30
4.1.8	Analisis Performa	30
4.1.9	Kesimpulan dan saran	30
4.1.10	Pelaporan	31
BAB 5 HASIL DAN PEMBAHASAN		32
5.1.	Penyeimbangan Menggunakan SMOTE	32
5.2.	Penerapan menggunakan Algoritma RF, KNN, NB dan SVM	33
5.3.	Perbandingan Kinerja Tiap Algoritma	38
5.4.	Hyperparameter Tuning dan Penambahan Noise	39
5.5.	Analisis Grafik ROC dan AUC	45
5.6.	Analisis Mendalam Algoritma Random Forest	47

5.7.	Analisis Feature Importance dan Boxplot dari Random Forest	49
5.8.	Penyajian Hasil Prediksi Dropout	55
BAB 6 KESIMPULAN DAN SARAN		56
6.1.	Kesimpulan	56
6.2.	Saran	57
DAFTAR PUSTAKA		58

DAFTAR GAMBAR

Gambar 3. 1 Skema Supervised Machine Learning	18
Gambar 4. 1 Diagram Alir Metodologi Penelitian	25
Gambar 5. 1 Distribusi Kelas Sebelum dan sesudah penyeimbangan.....	33
Gambar 5. 2 Kinerja Random Forest.....	34
Gambar 5. 3 Kinerja KNN.....	35
Gambar 5. 4 Kinerja Naïve Bayes	36
Gambar 5. 5 Grafik Kinerja SVM	37
Gambar 5. 6 Grafik Kinerja Berbagai Algoritma.....	38
Gambar 5. 7 Grafik Akurasi Tiap Algoritma	38
Gambar 5. 8 Grafik Kinerja Random Forest dengan Hyperparameter Tuning dan Penambahan Noise	40
Gambar 5. 9 Grafik Kinerja KNN dengan Hyperparameter Tuning dan Penambahan Noise	41
Gambar 5. 10 Grafik Kinerja NB dengan Hyperparameter Tuning dan Penambahan Noise	42
Gambar 5. 11 Grafik Kinerja SVM dengan Hyperparameter Tuning dan Penambahan Noise	43
Gambar 5. 12 Grafik Akurasi Tiap Algoritma dengan Hyperparameter Tuning dan Penambahan Noise	45
Gambar 5. 13 ROC-AUC Tiap Algoritma	46
Gambar 5. 14 Confusion Matrix.....	48
Gambar 5. 15 Feature Importance	49
Gambar 5. 16 Distribusi Feature Importance	50
Gambar 5. 17 Distribusi Bloxpot Feature Importance	51
Gambar 5. 18 Distribusi Propinsi Mahasiswa Berpotensi Dropout	55

DAFTAR TABEL

Tabel 2. 1 Daftar Penelitian Terdahulu.....	11
Tabel 4. 1 Data Akademik.....	26
Tabel 4. 2 Merubah Type Data	27
Tabel 4. 3 Transformasi type data numerik.....	27
Tabel 5. 1 Perbandingan Kinerja Tiap Algoritma	44