



**IDENTIFIKASI DINI MAHASISWA BERPOTENSI *DROP-OUT* DENGAN METODE
MACHINE LEARNING (STUDI KASUS: UNIVERSITAS PEMBANGUNAN
NASIONAL "VETERAN" JAKARTA)**

SKRIPSI

**SALSABILA FAIHA PUTERI
NIM. 2010512138**

**UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN" JAKARTA
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI S1 SISTEM INFORMASI
2024**



**IDENTIFIKASI DINI MAHASISWA BERPOTENSI *DROP-OUT* DENGAN
METODE *MACHINE LEARNING* (STUDI KASUS: UNIVERSITAS
PEMBANGUNAN NASIONAL "VETERAN" JAKARTA)**

SKRIPSI

**Diajukan Sebagai Salah Satu Syarat untuk Memperoleh Gelar
Sarjana Komputer**

SALSABILA FAIHA PUTERI

NIM. 2010512138

**UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN" JAKARTA
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI S1 SISTEM INFORMASI**

2024

PERNYATAAN ORISINALITAS

PERNYATAAN ORISINALITAS

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

Nama : Salsabila Faiha Puteri
NIM : 2010512138
Tanggal : 26 Juli 2024

Bilamana di kemudian hari ditemukan ketidaksesuaian dengan pernyataan ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Jakarta, 26 Juli 2024

Yang Menyatakan,



Salsabila Faiha Puteri

PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS

PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas akademik Universitas Pembangunan Nasional “Veteran” Jakarta, saya yang bertanda tangan di bawah ini:

Nama : Salsabila Faiha Puteri
NIM : 2010512138
Fakultas : Ilmu Komputer
Program Studi : S1 Sistem Informasi

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:

IDENTIFIKASI DINI MAHASISWA BERPOTENSI *DRO-OUT* DENGAN METODE *MACHINE LEARNING* (STUDI KASUS: UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN” JAKARTA)

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional “Veteran” Jakarta berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan Skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sesungguhnya.

Dibuat di : Jakarta
Pada Tanggal : 26 Juli 2024
Yang Menyatakan,



Salsabila Faiha Puteri

LEMBAR PENGESAHAN

LEMBAR PENGESAHAN

Skripsi ini diajukan oleh:

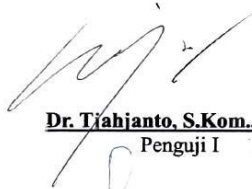
Nama : Salsabila Faiha Puteri

NIM : 2010512138

Program Studi : S1 Sistem Informasi

Judul Skripsi : Identifikasi Dini Mahasiswa Berpotensi *Drop-Out* dengan Metode *Machine Learning* (Studi Kasus: Universitas Pembangunan Nasional “Veteran” Jakarta)


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



Dr. Tjahjanto, S.Kom., M.M.
Penguji I





Sarika, S.Kom., M.Kom.
Penguji II



Dr. Bambang Saras Yulistiawan, S.T., M.Kom.
Pembimbing I



M. Octaviano, S.Kom., M.Kom.
Pembimbing II



Prof. Dr. Ir. Supriyanto, ST., M.Sc., IPM
Dekan Fakultas Ilmu Komputer



Anita Muliawati, S.Kom., M.TI.
Ketua Program Studi S1 Sistem Informasi

Ditetapkan di : Jakarta

Tanggal Ujian : 1 Juli 2024

**IDENTIFIKASI DINI MAHASISWA BERPOTENSI *DROP-OUT* DENGAN
METODE *MACHINE LEARNING* (STUDI KASUS: UNIVERSITAS
PEMBANGUNAN NASIONAL “VETERAN” JAKARTA)**

ABSTRAK

Salah satu kriteria dalam menentukan kesuksesan dari sebuah instansi pendidikan khususnya perguruan tinggi adalah performa mahasiswa. Semakin baik performa mahasiswa maka akan semakin baik kualitas dari sebuah perguruan tinggi. Jumlah mahasiswa *drop out* merupakan salah satu indikator penting terhadap penilaian kualitas sebuah perguruan tinggi. Semakin dini mahasiswa berpotensi *drop out* terdeteksi, maka akan semakin cepat dilakukan intervensi sehingga akan meminimalisir terjadinya mahasiswa *drop out*. Oleh sebab itu, perlu adanya sistem untuk memprediksi mahasiswa berpotensi *drop out* pada suatu perguruan tinggi, dengan tujuan meningkatkan kualitas mahasiswa dan perguruan tinggi. Upaya yang dilakukan dalam identifikasi dini mahasiswa berpotensi *drop out* juga bermanfaat untuk mencegah lonjakan mahasiswa *drop out* yang terjadi pada tahun-tahun akademik berikutnya. Berdasarkan observasi yang dilakukan di Universitas Pembangunan Nasional “Veteran” Jakarta, terdapat sebanyak 565 mahasiswa *drop out* pada tahun 2015 hingga tahun 2022. Dengan demikian, pada penelitian ini dilakukan pembuatan sistem yang berfungsi untuk melakukan identifikasi dini mahasiswa berpotensi *drop out* pada Universitas Pembangunan Nasional “Veteran” Jakarta. Penelitian dilakukan dengan menggunakan metode *machine learning* yaitu dengan menggunakan algoritma KNN dan Naïve Bayes. Pada penelitian ini juga dibangun sebuah sistem *dashboard* sebagai wadah untuk menyajikan visualisasi data terhadap data identifikasi dini mahasiswa berpotensi *drop out* yang informatif.

Kata Kunci: *drop out, machine learning, KNN, Naïve Bayes*

**EARLY IDENTIFICATION OF POTENTIAL DROP-OUT STUDENTS
USING MACHINE LEARNING METHODS (CASE STUDY: UNIVERSITAS
PEMBANGUNAN NASIONAL “VETERAN” JAKARTA)**

ABSTRACT

One of the criteria for determining the success of an educational institution, particularly a university, is student performance. The better the student performance, the better the quality of the university. The number of drop-out students is a significant indicator in assessing the quality of a university. The earlier potential drop-out students are detected, the quicker interventions can be implemented, thus minimizing the occurrence of student drop-outs. Therefore, there is a need for a system to predict potential drop-out students at a university to improve both student quality and university quality. Efforts in early identification of potential drop-out students also help prevent a surge in drop-outs in subsequent academic years. Based on observations conducted at Universitas Pembangunan Nasional “Veteran” Jakarta, there were 565 drop-out students from 2015 to 2022. Consequently, this research aims to develop a system that can perform early identification of potential drop-out students at Universitas Pembangunan Nasional “Veteran” Jakarta. The study employs machine learning methods using the KNN and Naïve Bayes algorithms. Additionally, this research develops a dashboard system to provide informative visualizations of the early identification data of potential drop-out students.

Keywords: drop out, machine learning, KNN, Naïve Bayes

KATA PENGANTAR

Puji syukur kehadiran Allah SWT yang telah melimpahkan rahmat dan hidayahnya kepada penulis, sehingga penulis dapat menyusun dan menyelesaikan skripsi dengan baik. Selama kegiatan dan penyelesaian skripsi, penulis banyak menerima dukungan dan bantuan oleh banyak pihak. Oleh karena itu, pada kesempatan ini penulis ingin mengucapkan terimakasih yang sebesar-besarnya pada pihak-pihak terkait yaitu kepada:

1. Allah SWT yang telah melimpahkan rahmat-Nya sehingga penulis dapat mengerjakan skripsi dengan lancar, semangat, dan pantang menyerah.
2. Kedua orang tua penulis yang senantiasa memberi dukungan kepada penulis agar selalu semangat dalam menyusun penelitian skripsi ini.
3. Bapak Prof. Dr. Ir. Supriyanto, S.T., M.Sc., IPM selaku Dekan Fakultas Ilmu Komputer UPN “Veteran” Jakarta.
4. Ibu Anita Muliawati, S.Kom., MTI selaku Ketua Program Studi Sarjana Jurusan Sistem Informasi.
5. Bapak Dr. Bambang Saras Yulistiawan, S.T., M.Kom selaku Dosen Pembimbing I dan Bapak M. Octaviano, S.Kom., M.Kom selaku Dosen Pembimbing II yang telah memberikan bimbingan kepada penulis.
6. Teman-teman penulis yang selalu memberikan dukungan penuh, bantuan dan kebahagiaan kepada penulis.

Adapun kekurangan yang terdapat pada penelitian skripsi ini, penulis sangat berharap atas kritik dan saran agar menjadi lebih baik lagi kedepannya. Selain itu, penulis juga memohon maaf sebesar-besarnya kepada semua pihak apabila terdapat kesalahan pada penelitian skripsi ini. Harapan penulis, semoga skripsi ini dapat bermanfaat bagi semua.

Jakarta, 20 Juni 2024

Penulis

DAFTAR ISI

PERNYATAAN ORISINALITAS	ii
PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS	iii
LEMBAR PENGESAHAN.....	v
ABSTRAK.....	vi
KATA PENGANTAR	viii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xii
DAFTAR GAMBAR	xiii
DAFTAR SIMBOL	xv
DAFTAR LAMPIRAN	xviii
BAB I.....	1
PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Tujuan Penelitian	3
1.4 Manfaat Penelitian	3
1.5 Batasan Masalah	3
1.6 Luaran yang Diharapkan.....	4
1.7 Sistematika Penulisan	4
BAB II.....	6
TINJAUAN PUSTAKA	6
2.1 Drop Out	6
2.2 Artificial Intelligence	6
2.3 Machine Learning	6
2.4 Database.....	7
2.5 Data Mining	7
2.5.1 Data Collecting.....	7
2.5.2 Data Preprocessing	8
2.5.3 Data Splitting	9
2.6 Modeling.....	9
2.6.1 Algoritma <i>Naïve Bayes</i>	9

2.6.2 Algoritma K-Nearest Neighbour (KNN).....	10
2.7 Uji Performa Model	13
2.7.1 Confusion Matrix	13
2.8 UML	15
2.9 Bahasa Pemrograman <i>Python</i>	15
2.10 <i>Website</i>	15
2.10.1 <i>HTML & CSS</i>	16
2.10.2 <i>Javascript</i>	16
2.10.4 <i>Flask</i>	16
2.10.5 <i>Black Box Testing</i>	17
2.11 Review Penelitian Terdahulu.....	17
BAB III.....	17
METODOLOGI PENELITIAN.....	17
3.1 Alur Penelitian	17
3.1.1 Identifikasi Masalah	18
3.1.2 Studi Literatur	19
3.1.2 Pengumpulan Data	19
3.1.3 <i>Preprocessing Data</i>	19
3.1.4 Splitting Data	20
3.1.5 Membangun Model	20
3.1.6 Evaluasi Model	21
3.1.7 Pemilihan Model Terbaik.....	21
3.1.8 Perancangan dan Pengujian Sistem.....	21
3.2 Perangkat penelitian.....	22
3.2.1 Perangkat Keras	22
3.2.2 Perangkat Lunak.....	22
3.3 Waktu dan Tempat Penelitian.....	23
3.4 Jadwal Penelitian	23
BAB IV	18
HASIL DAN PEMBAHASAN.....	18
4.1 Profil UPN “Veteran” Jakarta.....	18
4.1.1 Struktur Organisasi UPN “Veteran” Jakarta	18
4.2 Dataset	19
4.3 Preprocessing Data	22

4.3.1 <i>Data Cleaning</i>	23
4.3.2 <i>Data Transformation (Label Encoding)</i>	26
4.3.3 <i>Feature Engineering</i>	27
4.3.4 <i>Data Scalling</i>	28
4.4 <i>Modeling</i>	29
4.4.1 <i>Data Preparation</i>	29
4.4.2 <i>Modeling K-Nearest Neighbor</i>	31
4.4.3 <i>Evaluasi Model K-Nearest Neighbor</i>	31
4.4.4 <i>Modeling Naïve Bayes</i>	33
4.4.5 <i>Evaluasi Model Naïve Bayes</i>	33
4.5 <i>Pemilihan Model Terbaik</i>	35
4.7 <i>Perancangan dan Pengujian Sistem</i>	36
4.7.1 <i>Analisis Kebutuhan</i>	36
4.7.2 <i>Sistem Desain</i>	37
4.7.3 <i>Implementasi</i>	43
4.7.4 <i>Pengujian Sistem</i>	45
4.7.5 <i>Penerapan</i>	46
4.7.6 <i>Pemeliharaan</i>	46
BAB V	47
PENUTUP	47
5.1 <i>Kesimpulan</i>	47
5.2 <i>Saran</i>	47
DAFTAR PUSTAKA	49
LAMPIRAN	52

DAFTAR TABEL

Tabel 2.1 Confusion Matrix	13
Tabel 2.2 Review Penelitian Terdahulu	17
Tabel 3.1 Jadwal Penelitian	23
Tabel 4.1 Menentukan Parameter k.....	29
Tabel 4.2 Perbandingan Rasio.....	30
Tabel 4.3 Perbandingan Metric	30
Tabel 4.4 Confusion Matrix Data Training KNN	31
Tabel 4.5 Confusion Matrix Data Testing KNN	32
Tabel 4.6 Confusion Matrix Data Training Naive Bayes.....	33
Tabel 4.7 Confusion Matrix Data Testing Naive Bayes.....	34
Tabel 4.8 Pemilihan Model Terbaik.....	35
Tabel 4.9 Use Case Beranda.....	38
Tabel 4.10 Use Case Menu Dropout Mahasiswa	38
Tabel 4.11 Use Case Filter Tahun dan Program Studi	39
Tabel 4.12 Use case Chart dan Tabel Prediksi Dropout.....	39
Tabel 4.13 Activity Diagram Dropout Mahasiswa.....	40
Tabel 4.14 Black Box Testing	45

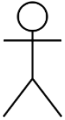


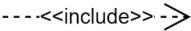
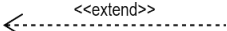


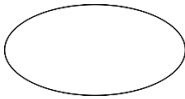

DAFTAR GAMBAR

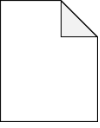
Gambar 2.1 Naive Bayes Classifier.....	10
Gambar 2.2 K Nearest Neighbor	11
Gambar 4.1 Struktur Organisasi UPN "Veteran" Jakarta.....	19
Gambar 4.2 Inisialisasi API Key	20
Gambar 4.3 Get Data Histori Mahasiswa.....	20
Gambar 4.4 Get API Data List Mahasiswa	20
Gambar 4.5 Field Data Histori Mahasiswa	21
Gambar 4.6 Field Data List Mahasiswa	21
Gambar 4.7 Inisialisasi API Key Prediksi	21
Gambar 4.8 Get Data Prediksi (Histori Mahasiswa)	22
Gambar 4.9 Get Data Prediksi (List Mahasiswa)	22
Gambar 4.10 Memilih Atribut Mahasiswa Tingkat S1	23
Gambar 4.11 Drop Column IPS	23
Gambar 4.12 Membuat DataFrame SKS Tempuh	23
Gambar 4.13 Merge DataFrame SKS dan Data List Mahasiswa	23
Gambar 4.14 Penghitungan Masa Studi	24
Gambar 4.15 Merge Data IPK, SKS, Masa Studi	24
Gambar 4.16 Menghapus Duplikasi Data	25
Gambar 4.17 Membuat Fungsi Preprocess Data Prediksi	25
Gambar 4.18 Penghitungan Masa Studi Data Prediksi	25
Gambar 4.19 Drop Column Data Prediksi	26
Gambar 4.20 Label Encoding	26
Gambar 4.21 Hasil Label Encoding	26
Gambar 4.22 Inisialisasi Status Mahasiswa	27
Gambar 4.23 Membuat Kriteria Masa Studi dan SKS	27
Gambar 4.24 Membuat Atribut Status Mahasiswa.....	28
Gambar 4.25 Scalling Data Model	28
Gambar 4.26 Scalling Data Prediksi	29
Gambar 4.27 Membuat Variable Fitur dan Label.....	29
Gambar 4.28 Informasi Field DataFrame df_mahasiswa.....	29
Gambar 4.29 Splitting Data Rasio (80:20).....	30

Gambar 4.30 Splitting Data Rasio (70:30).....	30
Gambar 4.31 Training Model KNN	31
Gambar 4.32 Training Model Naive Bayes.....	33
Gambar 4.33 Use Case Dropout Mahasiswa.....	38
Gambar 4.34 Class Diagram Dropout Mahasiswa	42
Gambar 4.35 Rancangan Wireframe Menu Beranda	43
Gambar 4.36 Rancangan Wireframe Menu Dropout Mahasiswa.....	43
Gambar 4.37 User Interface Menu Beranda.....	44
Gambar 4.38 User Interface Menu Dropout Mahasiswa (1)	44
Gambar 4.39 User Interface Menu Dropout Mahasiswa (2)	44
Gambar 4.40 User Interface Menu Dropout Mahasiswa (3)	45
Gambar 4.41 User Interface Menu Dropout Mahasiswa (4)	45

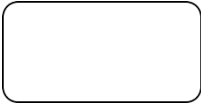



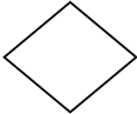

DAFTAR SIMBOL

Use Case Diagram

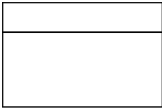

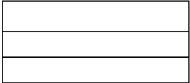
NO	GAMBAR	NAMA	KETERANGAN
1		<i>Actor</i>	Peran, sistem, atau alat ketika berinteraksi dengan <i>Use Case</i> .
2		<i>Dependency</i>	Hubungan dimana perubahan yang terjadi pada suatu elemen mandiri mempengaruhi elemen yang bergantung.
3		<i>Generalization</i>	Hubungan objek anak berbagi perilaku dan struktur data dari objek di atas objek induk.
4		<i>Include</i>	Menspesifikasikan <i>Use Case</i> secara eksplisit.
5		<i>Extend</i>	Menspesifikasikan <i>Use Case target</i> memperluas perilaku dari <i>Use Case</i> .
6		<i>Association</i>	Menghubungkan objek satu dengan lainnya.
7		<i>System</i>	Menspesifikasikan bahwa tampilan sistem terbatas.
8		<i>Use Case</i>	Deskripsi dari aksi-aksi pada sistem yang menghasilkan hasil yang terukur.
9		<i>Collaboration</i>	Interaksi aturan-aturan dan elemen-elemen yang bekerja sama untuk menyediakan perilaku yang lebih besar dari jumlah

			elemen-elemennya.
10		<i>Note</i>	Elemen fisik yang eksis saat aplikasi dijalankan dan mencerminkan suatu sumber daya komputasi

Activity Diagram

NO	GAMBAR	NAMA	KETERANGAN
1		<i>Activity</i>	Menunjukkan bahwa bagaimana masing-masing kelas antarmuka saling berinteraksi.
2		<i>Action</i>	State dari sistem yang mencerminkan eksekusi dari suatu aksi.
3		<i>Initial Node</i>	Awal dari bagaimana objek dibentuk.
4		<i>Activity Final Node</i>	Bagaimana objek diakhiri.
5		<i>Decision</i>	Digunakan sebagai suatu keputusan pada kondisi tertentu.
6		<i>Line Connector</i>	Menghubungkan satu simbol dengan simbol lainnya.

Class Diagram

NO	GAMBAR	NAMA	KETERANGAN
1		<i>Class Diagram</i>	Menggambarkan kelas yang terdiri dari dari atribut dan <i>method</i> .
2		<i>Relation</i>	Menggambarkan hubungan komponen-komponen didalam <i>diagram statis</i> .
3		<i>Association Class</i>	<i>Class</i> yang terbentuk dari hubungan antara dua buah <i>class</i>

DAFTAR LAMPIRAN

Lampiran 1. Wawancara Bersama Pihak AKPK.....	52
Lampiran 2. Wawancara Bersama Pihak LP3M	53
Lampiran 3. Field Data Histori Mahasiswa.....	54
Lampiran 4. Field Data List Mahasiswa.....	55
Lampiran 5. Atribut Data Model dan Prediksi	56
Lampiran 6. Atribut Data Visualisasi Prediksi Mahasiswa.....	57
Lampiran 7. Source Code Model	58
Lampiran 8. Source Code Website dan Prediksi Model	66
Lampiran 9. Hasil Pemodelan	114
Lampiran 10. Sistem Dashboard Analytic.....	115