

TUGAS AKHIR

PERBANDINGAN AKURASI ALGORITMA K-NEAREST NEIGHBOR (KNN) DAN NAIVE BAYES DALAM MEMPREDIKSI WAKTU KELULUSAN MAHASISWA



Disusun Oleh :

MUHAMMAD FAJAR ANDRIANSYAH

NBI : 1461700085

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS 17 AGUSTUS 1945 SURABAYA**

2023

TUGAS AKHIR

PERBANDINGAN AKURASI ALGORITMA K-NEAREST NEIGHBOR (KNN) DAN NAIVE BAYES DALAM MEMPREDIKSI WAKTU KELULUSAN MAHASISWA



Disusun Oleh :

MUHAMMAD FAJAR ANDRIANSYAH

NBI : 1461700085

PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS 17 AGUSTUS 1945 SURABAYA

2023

TUGAS AKHIR

PERBANDINGAN AKURASI ALGORITMA
K-NEAREST NEIGHBOR (KNN) DAN NAIVE BAYES
DALAM MEMPREDIKSI WAKTU KELULUSAN
MAHASISWA

Diajukan sebagai salah satu syarat untuk memperoleh gelar
Sarjana Komputer di Program Studi Informatika



Oleh:

Muhammad Fajar Andriansyah

1461700085

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS 17 AGUSTUS 1945 SURABAYA
2023

FINAL PROJECT

COMPARISON OF ACCURACY ALGORITHMS K-NEAREST NEIGHBOR (KNN) AND NAIVE BAYES IN PREDICTING STUDENT GRADUATION TIME

Prepared as partial fulfilment of the requirement for the degree of
Sarjana Komputer at informatics department



By :

Muhammad Fajar Andriansyah

1461700085

INFORMATICS DEPARTMENT

FACULTY OF ENGINEERING

UNIVERSITAS 17 AGUSTUS 1945 SURABAYA

2023

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS 17 AGUSTUS 1945 SURABAYA

LEMBAR PENGESAHAN TUGAS AKHIR

Nama : Muhammad Fajar Andriansyah
NBI : 1461700085
Prodi : Teknik Informatika
Fakultas : Teknik
Judul : Perbandingan Akurasi Algoritma *K-Nearest Neighbor* (KNN)
dan *Naive Bayes* Dalam Memprediksi Waktu Kelulusan
Mahasiswa

Mengetahui / Menyetujui

Dosen Pembimbing 1



Fridy Mandita, S.Kom., M.Sc
NPP. 20460.15.0648

Dekan Fakultas Teknik
Universitas 17 Agustus 1945
Surabaya



Dr. Ir. Sajjyo, M.Kes., IPU., ASEAN Eng.
NPP. 20410.90.0197

Ketua Program Studi Informatika
Universitas 17 Agustus 1945
Surabaya



Aidil Primasetya, S.ST., MT.
NPP. 20460.16.0700

PERNYATAAN KEASLIAN DAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

Saya yang bertanda tangan di bawah ini

Nama : Muhammad Fajar Andriansyah
NBI : 1461700085
Fakultas/Program Studi : Teknik/Informatika
Judul Tugas Akhir : Perbandingan Akurasi Algoritma *K-Nearest Neighbor* (KNN) dan *Naive Bayes* Dalam Memprediksi Waktu Kelulusan Mahasiswa

Menyatakan dengan sesungguhnya bahwa :

1. Tugas Akhir dengan judul diatas bukan merupakan tiruan atau duplikasi dari Tugas Akhir yang sudah dipublikasikan dan atau pernah dipakai untuk mendapatkan gelar Sarjana Teknik di lingkungan Universitas 17 Agustus 1945 Surabaya maupun di Perguruan Tinggi atau instansi manapun, kecuali bagian yang sumber informasinya dicantumkan sebagaimana mestinya.
2. Tugas Akhir dengan judul diatas bukan merupakan plagiarisme, pencurian hasil karya milik orang lain, hasil kerja orang lain untuk kepentingan saya karena hubungan material maupun non – material, ataupun segala kemungkinan lain yang pada hakekatnya bukan merupakan karya tulis tugas akhir saya secara orisinal dan otentik.
3. Demi pengembangan ilmu pengetahuan, saya memberikan hak atas Tugas Akhir ini kepada Universitas 17 Agustus 1945 Surabaya untuk menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (database), merawat dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.
4. Pernyataan ini saya buat dengan kesadaran sendiri dan tidak atas tekanan ataupun paksaan dari pihak maupun demi menegakan integritas akademik di institusi ini dan bila kemudian hari diduga kuat ada ketidaksesuaian antara fakta dengan kenyataan ini, saya bersedia diproses oleh tim Fakultas yang dibentuk untuk melakukan verifikasi, dengan sanksi terberat berupa pembatalan kelulusan/kesarjanaan.



Muhammad Fajar Andriansyah

1461700085



UNIVERSITAS
17 AGUSTUS 1945
SURABAYA

BADAN PERPUSTAKAAN
JL. SEMOLOWARU 45 SURABAYA
TLP. 031 593 1800 (EX 311)
EMAIL: PERPUS@UNTAG-SBY.AC.ID.

**LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI KARYA
ILMIAH UNTUK KEPENTINGAN
AKADEMIS**

Sebagai sivitas akademik Universitas 17 Agustus 1945 Surabaya,
Saya yang bertanda tangan di bawah ini:

Nama : Muhammad Fajar Andriansyah
NIM : 1461700085
Fakultas : Teknik
Program Studi : Teknik Informatika
Jenis Karya : Under Graduate / S1

Demi perkembangan ilmu pengetahuan, saya meyetujui untuk memberikan kepada Badan Perpustakaan Universitas 17 Agustus 1945 Surabaya **Hak Bebas Royalti Noneklusif (Nonexclusive Royalty-Free Right)**, atas karya saya yang berjudul:

PERBANDINGAN AKURASI ALGORITMA K-NEAREST NEIGHBOR (KNN) DAN NAIVE BAYES DALAM MEMPREDIKSI WAKTU KELULUSAN MAHASISWA

Dengan **Hak Bebas Royalti Noneklusif (Nonexclusive Royalty-Free Right)**, Badan Perpustakaan Universitas 17 Agustus 1945 Surabaya berhak menyimpan, mengalihkan media atau memformatkan, mengolah dalam bentuk pangkalan data (database), merawat, mempublikasikan karya ilmiah saya selama tetap tercantum.

Dibuat di : Universitas 17 Agustus 1945 Surabaya
Pada Tanggal : 01 Juli 2023



(Muhammad Fajar Andriansyah)

KATA PENGANTAR

Puji syukur kepada Allah Yang Maha Esa dan Yang Maha Kuasa yang senantiasa melimpahkan Rahmat dan Hidayahnya Sehingga dapat menyelesaikan Tugas Akhir yang berjudul “Perbandingan Akurasi Algoritma *K-Nearest Neighbor* (KNN) dan *Naive Bayes* Dalam Memprediksi Waktu Kelulusan Mahasiswa” sebagai salah satu persyaratan menyelesaikan studi di Universitas 17 Agustus 1945 Surabaya dan mendapatkan gelar sarjana komputer. Penulis ingin menyampaikan terima kasih kepada:

1. Kedua orang tua, yang telah memberikan dukungan beserta doa, memotivasi, memperhatikan, dan melengkapkan segala keperluan sehingga terselesaikannya Tugas Akhir ini.
2. Bapak Fridy Mandita, S.Kom., M.Sc selaku dosen pembimbing, yang telah memberikan petunjuk, pengarahan, semangat serta bimbingan dari awal pembuatan sistem sampai terselesaikan Tugas Akhir penulis.
3. Bapak Aidil Primasetya Armin, S.ST., M.T, selaku dosen wali yang telah membimbing dan mengarahkan saya selama studi di Universitas 17 Agustus 1945 Surabaya.
4. Seluruh teman-teman yang telah membantu dan memberikan motivasi penuh sehingga Laporan Tugas Akhir ini terselesaikan.

Dalam keterbatasan kami dalam menyampaikan materi pada Laporan Tugas Akhir ini, kami sangat mengharapkan kritik dan saran yang bersifat membangun guna menyempurnakan laporan-laporan yang akan kami buat selanjutnya. Atas kerja sama dari semua pihak yang telah membantu penulisan laporan ini, penulis mengucapkan terima kasih.

ABSTRAK

Nama : Muhammad Fajar Andriansyah
Program Studi : Informatika
Judul : Perbandingan Akurasi Algoritma *K-Nearest Neighbor* (KNN) dan *Naive Bayes* Dalam Memprediksi Waktu Kelulusan Mahasiswa

Dalam sistem pendidikan, mahasiswa merupakan aset penting bagi lembaga pendidikan tinggi, dan untuk itu perlu diperhatikan waktu kelulusan mahasiswa. Oleh sebab itu, penting mengetahui prediksi lama waktu penyelesaian studi mahasiswa. Penelitian ini mengusulkan pembuatan sistem prediksi tingkat kelulusan mahasiswa agar dapat menjadi langkah preventif agar mahasiswa dapat memperbaiki proses belajar. Penelitian ini melakukan teknik *Machine Learning* yang dilakukan dengan tujuan untuk membandingkan algoritma *K-Nearest Neighbor* (KNN) dan *Naive Bayes*. Percobaan penelitian dilakukan untuk menentukan model terbaik seperti, jumlah pengambilan data, jumlah kelas klasifikasi, dan *handling imbalance class*. Berdasarkan keseluruhan uji coba, metode KNN lebih unggul dibandingkan metode *Naive Bayes*. Penerapan teknik *oversampling* SMOTE menghasilkan peningkatan yang signifikan dengan selisih nilai evaluasi (presisi, *recall*, *F1 Score*, dan akurasi) antara 12% sampai 41% pada metode *Naive Bayes* maupun KNN. Hasil model prediksi 4 kelas menggunakan metode KNN dengan SMOTE mendapatkan nilai presisi sebesar 79%, nilai *recall* sebesar 78%, nilai *F1 Score* sebesar 78%, dan akurasi sebesar 78%. Sedangkan hasil prediksi 8 kelas menggunakan metode KNN dengan SMOTE mendapatkan nilai presisi, *recall*, *F1 Score*, dan akurasi sama-sama sebesar 93%.

Kata Kunci : Prediksi Waktu Lulus, *Naive Bayes*, KNN, SMOTE

ABSTRACT

Name : Muhammad Fajar Andriansyah
Department : Informatics
Title : Comparison of the Accuracy of K-Nearest Neighbor (KNN) and Naïve Bayes Algorithms in Predicting Student Graduation Time

In the education system, students are an essential asset for higher education institutions, and for this reason, it is necessary to pay attention to the time of graduation of students. Therefore, it is important to know the prediction of the time length for completing studies. This study proposes the creation of a prediction system for student graduation rates, hence it could be a preventive measure for students to improve their learning process. This research used machine learning techniques to compare the K-Nearest Neighbor (KNN) and Naïve Bayes algorithms. The experiment aimed to determine the best model, such as the amount of data collection, the number of classification classes, and the handling of imbalanced classes. Based on all experiments, the KNN method achieved higher results compared to the Naïve Bayes method. The application of the SMOTE oversampling technique resulted in a significant increase in the difference in evaluation scores (precision, recall, F1 score, and accuracy) between 12% and 41% in the Naïve Bayes and KNN methods. The results of the 4-class prediction model using the KNN method with SMOTE get a precision value of 79%, a recall value of 78%, an F1 score of 78%, and an accuracy of 78%. In comparison, the prediction results for 8 classes using the KNN method with SMOTE get a precision, recall, F1 Score, and accuracy values of 93%.

Keywords : *Graduation Time Prediction, Naïve Bayes, KNN, SMOTE*

DAFTAR ISI

LEMBAR PENGESAHAN TUGAS AKHIR	i
PERNYATAAN KEASLIAN DAN PERSETUJUAN PUBLIKASI TUGAS AKHIR	iii
KATA PENGANTAR.....	iv
ABSTRAK	ix
ABSTRACT	xi
DAFTAR ISI.....	xiii
DAFTAR GAMBAR.....	xv
DAFTAR TABEL.....	xvii
BAB 1 PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah.....	4
1.4 Tujuan	4
1.5 Manfaat	5
BAB 2 LANDASAN TEORI.....	6
2.1. Prediksi.....	7
2.2. Klasifikasi	7
2.3. Kecerdasan Buatan.....	8
2.4. <i>Machine Learning</i>	8
2.5. Algoritma <i>K-Nearest Neighbor</i> (KNN).....	10
2.6. Algoritma <i>Naïve Bayes</i>	13
2.7. Seleksi Fitur	15
2.8. <i>Synthetic Minority Oversampling Technique</i> (SMOTE).....	16
2.9. <i>K-Fold Cross Validation</i>	18
2.10. <i>Confusion Matrix</i>	19
2.11. Python	22
BAB 3 METODOLOGI PENELITIAN	24
3.1. Bahan dan Perangkat Penelitian	25
3.2. Objek Penelitian	25
3.3. Tahapan Penelitian	26

3.3.1.	Tahap Studi Literatur dan Perancangan.....	27
3.3.2.	Tahapan Persiapan.....	27
3.3.3.	Tahapan Implementasi Sistem.....	28
3.3.4.	Tahapan Analisis dan Pelaporan.....	29
BAB 4 HASIL DAN PEMBAHASAN		30
4.1.	Hasil <i>Exploratory Data Analysis</i> (EDA).....	31
4.2.	Hasil Data <i>Synthetic</i> Pada Teknik <i>Oversampling</i> SMOTE.....	36
4.3.	Hasil Keseluruhan Uji Coba Penelitian.....	53
4.4.	Hasil Komparasi Penerapan <i>Oversampling</i> SMOTE.....	56
4.5.	Hasil Komparasi Metode KNN dan <i>Naïve Bayes</i>	57
4.6.	Hasil Komparasi Beda Kelas dan Pengambilan Data.....	58
4.7.	Hasil Komparasi Beda Nilai K-Fold.....	59
4.8.	Hasil Pengujian <i>White Box</i>	60
4.9.	Hasil Implementasi <i>Testing Model</i>	68
BAB 5 PENUTUP		71
DAFTAR PUSTAKA		73

DAFTAR GAMBAR

Gambar 2.1 Contoh Penerapan Model Klasifikasi Menggunakan Metode K-Nearest Neighbor (KNN)	10
Gambar 2.2 Ilustrasi Perhitungan Algoritma SMOTE	18
Gambar 2.3 Visualisasi Pembagian Data Menggunakan K-Fold Cross Validation. .	19
Gambar 3.1 Flowchart Tahapan Penelitian	26
Gambar 3.2 Flowchart Pembuatan Model Klasifikasi Waktu Kelulusan Mahasiswa	28
Gambar 4.1 Analisis Perbandingan Mahasiswa yang Lulus Mata Kuliah Tugas Akhir dengan Waktu Lulus atau Lama Studi Keseluruhan	33
Gambar 4.2 Analisis Perbandingan Mahasiswa yang Lulus Mata Kuliah Tugas Akhir dengan Waktu Lulus atau Lama Studi 4 Kelas	34
Gambar 4.3 Uji Korelasi Seluruh Variabel Terhadap Lama Studi Mahasiswa 8 Kelas	35
Gambar 4.4 Uji Korelasi Seluruh Variabel Terhadap Lama Studi Mahasiswa 4 Kelas	366
Gambar 4. 5 Grafik Perbandingan Jumlah Label 8 Kelas Setelah Proses Cleaning .	37
Gambar 4. 6 Grafik Perbandingan Jumlah Label 4 Kelas	38
Gambar 4.7 Plot Perbandingan Sebaran (a) Data Original dan (b) Data Synthetic Pada Uji Coba 8 Kelas Klasifikasi dengan Pengambilan 100 Data	40
Gambar 4.8 Plot Perbandingan Sebaran (a) Data Original dan (b) Data Synthetic Pada Uji Coba 8 Kelas Klasifikasi dengan Pengambilan 257 Data	43
Gambar 4.9 Plot Perbandingan Sebaran (a) Data Original dan (b) Data Synthetic Pada Uji Coba 8 Kelas Klasifikasi dengan Pengambilan Semua Data	45
Gambar 4.10 Plot Perbandingan Sebaran (a) Data Original dan (b) Data Synthetic Pada Uji Coba 4 Kelas Klasifikasi dengan Pengambilan 100 Data	477
Gambar 4.11 Plot Perbandingan Sebaran (a) Data Original dan (b) Data Synthetic Pada Uji Coba 4 Kelas Klasifikasi dengan Pengambilan 257 Data	500
Gambar 4.12 Plot Perbandingan Sebaran (a) Data Original dan (b) Data Synthetic Pada Uji Coba 4 Kelas Klasifikasi dengan Pengambilan Semua Data	52

Gambar 4.13 Hasil Confusion Matrix dari Model Klasifikasi 8 Kelas dengan Pengambilan 257 Data Menggunakan Metode Naive Bayes dan (a) Tanpa SMOTE (b) dengan SMOTE	55
Gambar 4.14 Hasil Komparasi Nilai Evaluasi Metode KNN dan Naive Bayes Pada Uji Coba Teknik Oversampling SMOTE	566
Gambar 4.15 Hasil Komparasi Nilai Evaluasi Metode KNN dan Naive Bayes Pada Uji Coba Beda Kelas Klasifikasi	577
Gambar 4.16 Hasil Komparasi Nilai Akurasi Pada Tiap Uji Coba Jumlah Kelas dan Jumlah Pengambilan Data	58
Gambar 4.17 Listing Program function generateListColor	61
Gambar 4.18 Listing Program function calc Menggunakan Metode KNN	622
Gambar 4.19 Listing Program function calc Menggunakan Metode Naive Bayes ...	63
Gambar 4.20 Listing Program function printCount	633
Gambar 4.21 Listing Main Program	64
Gambar 4.22 Diagram Alir dari function generateListColor	644
Gambar 4.23 Diagram Alir dari function calc Menggunakan Metode KNN	65
Gambar 4.24 Diagram Alir dari function calc Menggunakan Metode Naive Bayes ..	65
Gambar 4.25 Diagram Alir dari function printCount	655
Gambar 4.26 Diagram Alir dari Main Program	66
Gambar 4.27 Hasil Uji Testing Dalam Prediksi Waktu Kelulusan Mahasiswa dengan 4 Kelas Klasifikasi	68
Gambar 4.28 Hasil Uji Testing Dalam Prediksi Waktu Kelulusan Mahasiswa dengan 8 Kelas Klasifikasi	69

Pseudocode 2.1 M
Pseudocode 2.2 M
Pseudocode 2.3 M

DAFTAR PSEUDOCODE

Pseudocode 2.1 Metode Algoritma KNN	12
Pseudocode 2.2 Metode Algoritma <i>Naïve Bayes</i>	15
Pseudocode 2.3 Metode Algoritma SMOTE.....	17

DAFTAR TABEL

Tabel 2.1 Confusion Matrix dengan Binary Class	20
Tabel 2.2 Confusion Matrix dengan Multi-class	21
Tabel 3.1 Perangkat Keras yang Digunakan	25
Tabel 3.2 Perangkat Lunak yang Digunakan	25
Tabel 4.1 Sampel Data Penelitian	31
Tabel 4.2 Sampel Data Original Pada Uji Coba 8 Kelas Klasifikasi dengan Pengambilan 100 Data	38
Tabel 4.3 Sampel Data Synthetic Pada Uji Coba 8 Kelas Klasifikasi dengan Pengambilan 100 Data	39
Tabel 4.4 Sampel Data Original Pada Uji Coba 8 Kelas Klasifikasi dengan Pengambilan 257 Data	41
Tabel 4.5 Sampel Data Synthetic Pada Uji Coba 8 Kelas Klasifikasi dengan Pengambilan 257 Data	42
Tabel 4.6 Sampel Data Original Pada Uji Coba 8 Kelas Klasifikasi dengan Pengambilan Semua Data.....	43
Tabel 4.7 Sampel Data Synthetic Pada Uji Coba 8 Kelas Klasifikasi dengan Pengambilan Semua Data.....	44
Tabel 4.8 Sampel Data Original Pada Uji Coba 4 Kelas Klasifikasi dengan Pengambilan 100 Data	46
Tabel 4.9 Sampel Data Synthetic Pada Uji Coba 4 Kelas Klasifikasi dengan Pengambilan 100 Data	47
Tabel 4.10 Sampel Data Original Pada Uji Coba 4 Kelas Klasifikasi dengan Pengambilan 257 Data	48
Tabel 4.11 Sampel Data Synthetic Pada Uji Coba 4 Kelas Klasifikasi dengan Pengambilan 257 Data	49
Tabel 4.12 Sampel Data Original Pada Uji Coba 4 Kelas Klasifikasi dengan Pengambilan Semua Data.....	51
Tabel 4.13 Sampel Data Synthetic Pada Uji Coba 4 Kelas Klasifikasi dengan Pengambilan Semua Data.....	52
Tabel 4.14 Hasil Keseluruhan Uji Coba Prediksi Waktu Kelulusan Mahasiswa.....	53

Tabel 4.15 Perbandingan Nilai K-Fold Pada Uji Coba 8 Kelas Klasifikasi	59
Tabel 4.16 Test Case function generateListColor	67
Tabel 4.17 Test Case function calc Menggunakan Metode KNN	67
Tabel 4.18 Test Case function calc Menggunakan Metode Naive Bayes	67
Tabel 4.19 Test Case function printCount	68
Tabel 4.20 Test Case Main Program	68