



**INTEGRASI NAIVE BAYES DENGAN SMOTE  
PADA KLASIFIKASI PENYAKIT DIABETES**

**SKRIPSI**

**Untuk Memenuhi Persyaratan  
Dalam Menyelesaikan Program Strata-1 Statistika**

**Oleh:**

**GUSTI RAHIMATUL HUSNA**

**NIM. 1811017320028**

**KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI  
UNIVERSITAS LAMBUNG MANGKURAT  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
PROGRAM STUDI STATISTIKA  
BANJARBARU  
APRIL 2023**



**INTEGRASI NAIVE BAYES DENGAN SMOTE  
PADA KLASIFIKASI PENYAKIT DIABETES**

**SKRIPSI**

**Untuk Memenuhi Persyaratan  
Dalam Menyelesaikan Program Strata-1 Statistika**

**Oleh:**

**GUSTI RAHIMATUL HUSNA  
NIM. 1811017320028**

**KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI  
UNIVERSITAS LAMBUNG MANGKURAT  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
PROGRAM STUDI STATISTIKA  
BANJARBARU  
APRIL 2023**

# SKRIPSI

## INTEGRASI NAIVE BAYES DENGAN SMOTE PADA KLASIFIKASI PENYAKIT DIABETES

Oleh:

**Gusti Rahmatul Husna**

**NIM. 1811017320028**

Telah dipertahankan di depan Dosen Penguji tanggal 10 April 2023. Susunan  
Dosen Penguji:

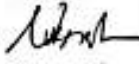

### Pembimbing I



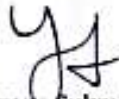
Oni Soesanto, S.Si., M.Si

NIP. 19730126 200501 1 003

### Dosen Penguji

1. Nur salam, S.Si., M.Sc 
2. Dr. Meitria Syahadatina Noor, dr., M.Kes 

### Pembimbing II



Yuana Sukmawaty, S.Si., M.Si


NIP. 19880420 201404 2 001

Banjarbaru, 17 Mei 2023

Mengetahui,

Koordinator Program Studi Statistika



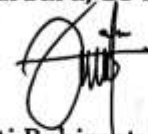
  
Dewi Anggraini, S.Si., M.App.Sci., Ph.D

NIP. 19830328 200501 2 001

## **PERNYATAAN**

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam Daftar Pustaka.

Banjarbaru, 16 Mei 2023



Gusti Rahimatul Husna  
NIM. 1811017320028

## ABSTRAK

**INTEGRASI NAIVE BAYES DENGAN SMOTE PADA KLASIFIKASI PENYAKIT DIABETES** (Oleh: Gusti Rahimatul Husna; Pembimbing: Oni Soesanto, S.Si., M.Si dan Yuana Sukmawaty, S.Si., M.Si; 69 halaman)

Klasifikasi data dengan distribusi kelas yang tidak seimbang (*imbalance class*) menjadi kelemahan bagi beberapa algoritma klasifikasi. Klasifikasi dengan distribusi kelas yang tidak seimbang (*imbalance class*) sering ditemui pada diagnosis penyakit misalnya pada penyakit diabetes. Tujuan dari penelitian ini adalah untuk mengetahui proses klasifikasi penyakit diabetes menggunakan *Naive Bayes* dengan *Synthetic Minority Oversampling Technique* (SMOTE) dan menganalisa performa kinerja dari *Naive Bayes* dengan SMOTE. SMOTE merupakan metode yang dapat menangani ketidakseimbangan kelas (*imbalance class*) pada sekelompok data. Model yang digunakan dalam penelitian ini ada dua yaitu klasifikasi *Naive Bayes* (model pertama) dan klasifikasi *Naive Bayes* dengan SMOTE (model kedua). Proses klasifikasi *Naive Bayes* diawali dengan membaca data *training* kemudian menghitung nilai *mean* dan standar deviasi, lalu menghitung nilai *Likelihood Naive Bayes* nya menggunakan fungsi *densitas gauss*. Sedangkan proses klasifikasi *Naive Bayes* dengan SMOTE diawali dengan sintesis data terlebih dahulu kemudian dilanjutkan dengan proses *Naive Bayes* sebelumnya. Dari hasil pengujian didapatkan model pertama menghasilkan nilai AUC sebesar 0,871 sedangkan model kedua menghasilkan nilai AUC sebesar 0,980. Kedua nilai AUC tersebut diuji dengan *Paired Sample t-Test* untuk melihat perbedaan kinerja model sebelum dan sesudah diterapkan metode SMOTE, menghasilkan *p-value* sebesar 0,003 artinya terdapat perbedaan antara model pertama dan model kedua dikarenakan nilai *p-value* < 0,05 (95% kepercayaan). Sehingga model terbaik untuk klasifikasi data Pima Indians Diabetes menggunakan model kedua yaitu klasifikasi *Naive Bayes* dengan SMOTE.

Kata kunci: Diabetes, *Naive Bayes*, SMOTE, *Area Under Curve* (AUC).

## ABSTRACT

### ***INTEGRATION OF NAIVE BAYES WITH SMOTE IN DIABETES CLASSIFICATION***

*(By: Gusti Rahimatul Husna; Supervisor: Oni Soesanto, S.Si., M.Si and Yuana Sukmawaty, S.Si., M.Si; 69 pages)*

Classification of data with an unbalanced class distribution is a weakness for several classification algorithms. Classification with an unequal class distribution is often found in the diagnosis of diseases such as diabetes. The purpose of this study was to determine the process of classification of diabetes using *Naive Bayes* with *Synthetic Minority Oversampling Technique* (SMOTE) and analyze the performance of the performance *Naive Bayes* with SMOTE. SMOTE is a method that can handle class imbalance in a group of data. There are two models used in this study, namely classification *Naive Bayes* (first models) and classification *Naive Bayes* with SMOTE (second model). Classification process *Naive Bayes* Start by reading the data *training* then calculate the value *mean* and the standard deviation, then calculates the value *Likelihood Naive Bayes* her use function *gauss density*. While the classification process *Naive Bayes* with SMOTE starting with data synthesis first then proceed with the process *Naive Bayes* previously. From the test results, it was found that the first model produced an AUC value of 0.871 while the second model produced an AUC value of 0.980. The two AUC values were tested with *Paired Sample t-Test* to see the difference in the performance of the model before and after the SMOTE method was applied, yield *p-value* of 0.003 means that there is a difference between the first model and the second model because of the value *p-value* < 0.05 (95% confidence). So the best model for classification of Pima Indians Diabetes data uses the second model, namely classification *Naive Bayes* with SMOTE.

*Keywords: Diabetes, Naive Bayes, SMOTE, Area Under Curve (AUC).*

## PRAKATA

Puji syukur penulis panjatkan atas kehadiran Allah *Subhanahu wata'ala*, karena atas rahmat-Nya penulis dapat menyelesaikan skripsi yang berjudul **"Integrasi Naive Bayes Dengan SMOTE Pada Klasifikasi Penyakit Diabetes"**. Penyusunan skripsi ini bertujuan untuk memenuhi salah satu persyaratan dalam rangka menyelesaikan program sarjana di Program Studi Statistika Fakultas MIPA Universitas Lambung Mangkurat.

Penulis menyampaikan ucapan terima kasih kepada pihak-pihak yang telah membantu dalam menyelesaikan ataupun membimbing penulisan skripsi ini, yaitu:

1. Bapak Drs. Abdul Gafur, M.Si., M.Sc., Ph.D. selaku Dekan Fakultas MIPA ULM.
2. Ibu Dewi Anggraini, S.Si., M.App.Sci., Ph.D. selaku Koordinator Program Studi Statistika, seluruh jajaran dosen, dan staff Program Studi Statistika Fakultas MIPA ULM.
3. Bapak Oni Soesanto, S.Si., M.Si dan Ibu Yuana Sukmawaty, S.Si., M.Si selaku pembimbing tugas akhir atas bimbingan, motivasi serta meluangkan waktu dalam pelaksanaan penelitian serta penyusunan skripsi;
4. Bapak Nur Salam, S.Si., M.Sc dan Ibu Dr. Meitria Syahadatina Noor, dr. M.Kes selaku penguji yang telah memberikan masukan dan saran dalam perbaikan skripsi;
5. Kedua orang tua dan keluarga besar yang telah memberikan motivasi, dukungan, do'a serta pengertian yang besar kepada penulis;
6. Maridha Istiqhomah, S.Stat, M. Ryan Afrizal, S.Kom, dan semua pihak yang tidak dapat disebutkan satu persatu yang telah memberikan dukungan serta motivasi dalam penulisan skripsi ini;

Akhir kata penulis berharap semoga skripsi ini dapat memberikan manfaat bagi pembacanya.

Banjarbaru, 16 Mei 2023



Gusti Rahimatul Husna  
NIM. 1811017320028

## DAFTAR ISI

SKRIPSI.....	iii
PERNYATAAN .....	iv
ABSTRAK.....	v
ABSTRACT .....	vi
PRAKATA .....	vii
DAFTAR ISI.....	viii
DAFTAR TABEL .....	x
DAFTAR GAMBAR .....	xi
DAFTAR LAMPIRAN .....	xii
ARTI LAMBANG DAN SINGKATAN .....	xiii
BAB I PENDAHULUAN .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah.....	4
1.3 Tujuan Penelitian.....	4
1.4 Manfaat Penelitian.....	4
BAB II TINJAUAN PUSTAKA .....	5
2.1 Kajian Peneliti Terdahulu.....	5
2.2 Kajian Teori.....	6
2.2.1 Diabeteses Melitus.....	6
2.2.2 <i>Data Mining</i> .....	7
2.2.3 Statistika Deskriptif .....	9

2.2.4	Normalisasi Data .....	7
2.2.5	Data <i>Imbalance</i> .....	10
2.2.6	<i>Naive Bayes Classifier</i> .....	11
2.2.7	Kriteria Evaluasi Kinerja Metode Klasifikasi.....	13
BAB III METODOLOGI PENELITIAN.....		15
3.1	Sumber Data.....	15
3.2	Variabel Penelitian .....	15
3.3	Tahapan Penelitian .....	16
3.4	Alur Penelitian .....	18
BAB IV HASIL DAN PEMBAHASAN .....		19
4.1	Uji Statistika Deskriptif .....	19
4.2	<i>Preprocessing</i> Data .....	22
4.3	Normalisasi Data .....	30
4.4	Pembagian <i>Dataset</i> .....	34
4.5	Proses Klasifikasi <i>Naive Bayes</i> Tanpa SMOTE.....	36
4.6	Proses Klasifikasi <i>Naive Bayes</i> Dengan SMOTE.....	42
4.7	Evaluasi.....	46
BAB V PENUTUP .....		48
5.1	Kesimpulan.....	48
5.2	Saran .....	49
DAFTAR PUSTAKA .....		50
LAMPIRAN.....		53
RIWAYAT HIDUP .....		70

## DAFTAR TABEL

	Halaman
Tabel 1 <i>Confusion Matrix</i> .....	13
Tabel 2 Definisi Operasional Variabel Penelitian.....	15
Tabel 3 <i>Dataset</i> Pima Indians Diabetes.....	20
Tabel 4 Ukuran Penyebaran Data .....	20
Tabel 5 Hasil Deteksi Missing Value.....	23
Tabel 6 <i>Dataset</i> Setelah <i>Preprocessing</i> .....	30
Tabel 7 Normalisasi Data Pima Indians Diabetes .....	34
Tabel 8 Hasil <i>Training</i> Splitting Data .....	34
Tabel 9 Hasil Perhitungan Mean.....	37
Tabel 10 Hasil Perhitungan Standar Deviasi .....	38
Tabel 11 Contoh Data <i>Testing</i> .....	39
Tabel 12 Probabilitas Setiap Variabel pada data ke-38.....	40
Tabel 13 <i>Confusion Matrix Naive Bayes</i> .....	41
Tabel 14 <i>Confusion Matrix Naive Bayes</i> dengan SMOTE.....	45
Tabel 15 Perbandingan Performa Model Klasifikasi .....	46

## DAFTAR GAMBAR

	Halaman
Gambar 1 <i>Flowchart Naive Bayes</i> .....	12
Gambar 2 <i>Flowchart</i> Penelitian.....	18
Gambar 3 Distribusi Kelas pada Pima Indians Diabetes <i>Database</i> .....	19
Gambar 4 <i>Outliers</i> pada variabel <i>Pregnancies</i> .....	24
Gambar 5 <i>Outliers</i> pada variabel <i>Glucose</i> .....	24
Gambar 6 <i>Outliers</i> pada variabel <i>Blood Pressure</i> .....	25
Gambar 7 <i>Outliers</i> pada variabel <i>Skin Thickness</i> .....	26
Gambar 8 <i>Outliers</i> pada variabel <i>Insulin</i> .....	26
Gambar 9 <i>Outliers</i> pada variabel BMI .....	27
Gambar 10 <i>Outliers</i> pada variabel <i>Diabetes Pedigree Function</i> .....	28
Gambar 11 <i>Outliers</i> pada variabel <i>Age</i> .....	29
Gambar 12 Hasil <i>Removing Outliers</i> .....	29
Gambar 13 Distribusi Pembagian <i>Dataset</i> .....	35
Gambar 14 Hasil Sintetis Data dengan SMOTE .....	44

## DAFTAR LAMPIRAN

	Halaman
Lampiran 1. Hasil Prediksi Data <i>Testing</i> Dengan <i>Naive Bayes</i> .....	53
Lampiran 2. Hasil Prediksi Data <i>Testing</i> Dengan <i>Naive Bayes</i> + SMOTE.....	54
Lampiran 3. <i>Dataset</i> Penelitian .....	56
Lampiran 4. <i>Source Code Import Library</i> .....	57
Lampiran 5. <i>Source Code</i> dan <i>Output Import Data</i> .....	59
Lampiran 6. <i>Source Code</i> dan <i>Output Deteksi Missing Value</i> .....	59
Lampiran 7. <i>Source Code</i> dan <i>Output Outliers Detection</i> .....	59
Lampiran 8. <i>Source Code</i> dan <i>Output Removing Outliers Dengan IQR</i> .....	60
Lampiran 9. <i>Source Code</i> dan <i>Output Removing Outliers</i> .....	61
Lampiran 10. <i>Source Code</i> dan <i>Output Analisis Deskriptif</i> .....	62
Lampiran 11. <i>Source Code</i> dan <i>Output Normalisasi Data</i> .....	63
Lampiran 12. <i>Source Code</i> dan <i>Output Data Splitting</i> .....	64
Lampiran 13. <i>Source Code</i> dan <i>Output Klasifikasi Naive Bayes</i> .....	64
Lampiran 14. <i>Source Code</i> dan <i>Output SMOTE</i> .....	66
Lampiran 15. <i>Source Code</i> dan <i>Output Klasifikasi Naive Bayes + SMOTE</i> .....	67