

**IMPLEMENTASI DATA MINING UNTUK KLASIFIKASI
KELAYAKAN MASYARAKAT UNTUK MENDAPATKAN
BANTUAN PKH DI DESA KALIGAWA DENGAN MODEL
NAIVE BAYES DAN RANDOM FOREST**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

STEFANI NUGRAHANINGTYAS

21.11.3877

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

**IMPLEMENTASI DATA MINING UNTUK KLASIFIKASI
KELAYAKAN MASYARAKAT UNTUK MENDAPATKAN
BANTUAN PKH DI DESA KALIGAWA DENGAN MODEL
NAIVE BAYES DAN RANDOM FOREST**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh
STEFANI NUGRAHANINGTYAS
21.11.3877

Kepada
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2025

HALAMAN PERSETUJUAN

SKRIPSI

**IMPLEMENTASI DATA MINING UNTUK KLASIFIKASI KELAYAKAN
MASYARAKAT UNTUK MENDAPATKAN BANTUAN PKH DI DESA
KALIGAWA DENGAN MODEL NAIVE BAYES DAN RANDOM FOREST**

yang disusun dan diajukan oleh

Stefani Nugrahaningtyas

21.11.3877

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 19 Februari 2025

Dosen Pembimbing,



Anna Baita, S.Kom., M.Kom.

NIK. 190302290

HALAMAN PENGESAHAN

SKRIPSI

**IMPLEMENTASI DATA MINING UNTUK KLASIFIKASI KELAYAKAN
MASYARAKAT UNTUK MENDAPATKAN BANTUAN PKH DI DESA
KALIGAWA DENGAN MODEL NAIVE BAYES DAN RANDOM FOREST**

yang disusun dan diajukan oleh

Stefani Nugrahaningtyas

21.11.3877

Telah dipertahankan di depan Dewan Penguji
pada tanggal 19 Februari 2025

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Dina Maulina, S.Kom., M.Kom.
NIK. 190302250

Ike Verawati, S.Kom., M.Kom.
NIK. 190302237

Anna Baita, S. Kom., M.Kom.
NIK. 190302290



Skrripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 19 Februari 2025

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom., Ph.D.

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Stefani Nugrahaningtyas

NIM : 21.11.3877

Menyatakan bahwa Skripsi dengan judul berikut:

**IMPLEMENTASI DATA MINING UNTUK KLASIFIKASI KELAYAKAN
MASYARAKAT UNTUK MENDAPATKAN BANTUAN PKH DI DESA KALIGAWA
DENGAN MODEL NAIVE BAYES DAN RANDOM FOREST**

Dosen Pembimbing : Anna Baita, S.Kom., M.Kom.

1. Karya tulis ini adalah benar-benar **ASLI** dan **BELUM PERNAH** diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian **SAYA** sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab **SAYA**, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini **SAYA** buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka **SAYA** bersedia menerima **SANKSI AKADEMIK** dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 19 Februari 2025

Yang Menyatakan,



Stefani Nugrahaningtyas

HALAMAN PERSEMBAHAN

Segala puji dan syukur saya panjatkan kepada Tuhan Yesus Kristus, Sang Sumber Hikmat dan Kasih, yang telah memberikan kekuatan, kesabaran, dan pertolongan-Nya dalam setiap proses penyelesaian skripsi ini.

Terima kasih atas motivasi, dukungan dan doa dari semua pihak yang telah ikut serta dalam penyelesaian pembuatan laporan skripsi ini. Penulis mengucapkan banyak terima kasih kepada:

1. Tuhan Yesus Kristus, yang senantiasa menyertai, memberikan hikmat, dan menguatkan saya dalam setiap langkah kehidupan.
2. Kedua orang tua tercinta, yang selalu mendoakan, mendukung, dan memberikan kasih sayang tanpa batas.
3. Kakak dan Adik yang senantiasa memberikan semangat dan dukungan kepada penulis dalam menyelesaikan skripsi ini.
4. Ibu Anna Baita, S.Kom, M.Kom, terimakasih untukmu sebagai dosen pembimbing yang telah bersedia mengantarkanku untuk mengantungi gelar "sarjana". Semoga kebahagiaanmu juga merupakan kebahagiaanmu sebagai "guruku" yang teramat baik.
5. Teman-teman yang tidak bisa saya sebutkan namanya satu persatu, yang selalu memberikan dukungan kepada penulis dalam menyelesaikan skripsi ini.

Skripsi ini adalah bukti kasih dan pertolongan Tuhan dalam hidup saya. Kiranya dapat menjadi berkat dan bermanfaat bagi banyak orang.

KATA PENGANTAR

Segala puji dan syukur saya panjatkan kepada Tuhan Yesus Kristus, Sang Sumber Hikmat dan Kasih, yang telah memberikan kekuatan, kesabaran, dan pertolongan-Nya dalam setiap proses penyelesaian skripsi, sehingga penulis dapat menyelesaikan skripsi yang berjudul “Implementasi Data Mining Untuk Klasifikasi Kelayakan Untuk Mendapatkan Bantuan PKH Di Desa Kaligawe Dengan Model Naïve Bayes dan Random Forest”. Skripsi ini dapat diselesaikan atas bimbingan dan dukungan dari berbagai pihak, sehingga pada kesempatan kali ini dengan tulus penulis menyampaikan rasa hormat dan rasa terimakasih kepada:

1. Hanif Al Fatta, S.Kom., M.Kom., Ph.D. selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
2. Windha Mega Pradnya Dhuhiya, M.Kom. selaku ketua Program Studi Informatika Universitas Amikom Yogyakarta.
3. Ibu Anna Baita, S.Kom., M.Kom. selaku pembimbing yang telah memberi bimbingan, arahan, dan motivasi kepada penulis selama proses penyusunan skripsi ini.
4. Ibu Dina Maulina, S.Kom., M.Kom. selaku dosen penguji 1 yang telah memberikan masukan, serta saran dalam penyusunan skripsi ini.
5. Ibu Ike Verawati, S.Kom., M.Kom. selaku dosen penguji 2 yang telah memberikan masukan, serta saran dalam penyusunan skripsi ini.
6. Suryadi, Sri Rahayu selaku orangtua yang selalu memberikan doa dan dukungan serta sebagai penyemangat penulis untuk menyelesaikan skripsi ini.

Yogyakarta, 19 Februari 2025

Penulis

DAFTAR ISI

HALAMAN PERSETUJUAN.....	i
HALAMAN PENGESAHAN.....	ii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iii
HALAMAN PERSEMBAHAN.....	iv
KATA PENGANTAR	v
DAFTAR ISI.....	vi
DAFTAR TABEL.....	x
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN.....	xii
INTISARI.....	xiii
ABSTRACT.....	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan.....	5
BAB II TINJAUAN PUSTAKA.....	7
2.1 Studi Literatur	7
2.2 Dasar Teori.....	13
2.2.1 Program Keluarga Harapan (PKH)	13
2.2.2 Data Mining	14
2.2.3 Klasifikasi	16

2.2.4	EDA.....	17
2.2.5	Data Preprocessing.....	18
2.2.6	Seleksi Fitur Berbasis Korelasi	19
2.2.7	SMOTE	19
2.2.8	Split Data.....	20
2.2.9	Naive Bayes	21
2.2.10	Random Forest.....	22
2.2.11	Evaluasi dan Validasi.....	24
2.2.12	Confusion Matrix	24
2.2.13	<i>Sensitivity (Recall)</i> dan <i>Specificity</i>	25
2.2.14	Geometric Mean.....	26
2.2.15	<i>Area Under the Curve (AUC)</i>	27
2.2.16	<i>Accuracy</i>	28
2.2.17	<i>Cross Validation</i>	28
2.2.18	<i>Deployment</i>	29
BAB III METODE PENELITIAN.....		31
3.1	Objek Penelitian.....	31
3.2	Alur Penelitian	31
3.2.1	Pengumpulan Dataset.....	32
3.2.2	Memuat Dataset	32
3.2.3	EDA.....	33
3.2.3.1	Identifikasi Struktur Data	33
3.2.3.2	Identifikasi Statistik Deskriptif.....	34
3.2.3.3	Identifikasi Outlier	34
3.2.4	<i>Preprocessing</i>	35

3.2.4.1	<i>Data Cleaning</i>	35
1.	Menangani <i>Outlier</i>	35
2.	Menghapus kolom yang tidak diperlukan.....	36
3.	Membersihkan Data Kategorikal.....	38
3.2.4.2	<i>Data Tranformation</i>	39
3.2.4.3	<i>Feature Selection</i>	40
3.2.5	Split Data.....	45
3.2.6	Teknik <i>Resampling</i>	45
3.2.7	<i>Modeling</i>	48
3.2.8	Evaluasi.....	48
3.2.9	<i>Deployment</i>	48
3.3	Alat dan Bahan.....	48
3.3.1	Data Penelitian	48
3.3.2	Alat/ Instrumen.....	49
BAB IV HASIL DAN PEMBAHASAN		50
4.1	Pengujian <i>Hyperparameter Tuning</i> dengan <i>GridSearchCV</i>	50
4.1.1	Tujuan Pengujian	50
4.1.2	Hasil Pengujian <i>Hyperparameter Tuning</i>	52
4.1.3	Analisis Hasil Pengujian	53
1.	Performa <i>Naïve Bayes</i> setelah <i>Hyperparameter Tuning</i>	53
2.	Performa <i>Random Forest</i> setelah <i>Hyperparameter Tuning</i>	54
4.1.4	Kesimpulan Pengujian	54
4.2	Pengujian Performa model sebelum dan sesudah <i>SMOTE</i>	55
4.2.1	Tujuan Pengujian	55
4.2.2	Hasil Evaluasi Model Sebelum dan Sesudah <i>SMOTE</i>	56
4.2.3	Analisis Hasil Pengujian Performa Model Sebelum dan Sesudah <i>SMOTE</i>	56

1.	Perbandingan Akurasi dan Akurasi <i>Cross-Validation</i>	56
2.	<i>Sensitivity (Recall)</i> dan <i>Specificity</i>	57
3.	<i>AUC-Score</i> dan <i>G-Mean</i>	57
4.2.4	Kesimpulan Pengujian	58
4.3	<i>Deployment</i>	58
4.3.1	Hasil <i>Deployment</i>	58
4.3.2	Tampilan Antarmuka (UI) Sistem	59
4.3.3	Pengujian Aplikasi	61
4.3.3.1	Pengujian Black Box	61
4.3.3.2	Pengujian <i>User Acceptance Test (UAT)</i>	63
4.4	Diskusi Hasil	64
4.4.1	Perbandingan Akurasi dan <i>Cross-Validation</i>	65
4.4.2	Perbandingan <i>Sensitivity</i> dan <i>Specificity</i>	65
4.4.3	Perbandingan <i>AUC-Score</i> dan <i>G-Mean</i>	66
4.4.4	Perbandingan Efisiensi Waktu Otomatisasi dan Manual	67
4.4.5	Kesimpulan Perbandingan	67
BAB V PENUTUP		70
5.1	Kesimpulan	70
1.	Otomatisasi Proses Klasifikasi	70
2.	Dampak Penerapan SMOTE terhadap Performa Model	70
3.	Perbandingan Performa Naïve Bayes dan Random Forest	71
5.2	Saran	71
REFERENSI		73
LAMPIRAN		78

DAFTAR TABEL

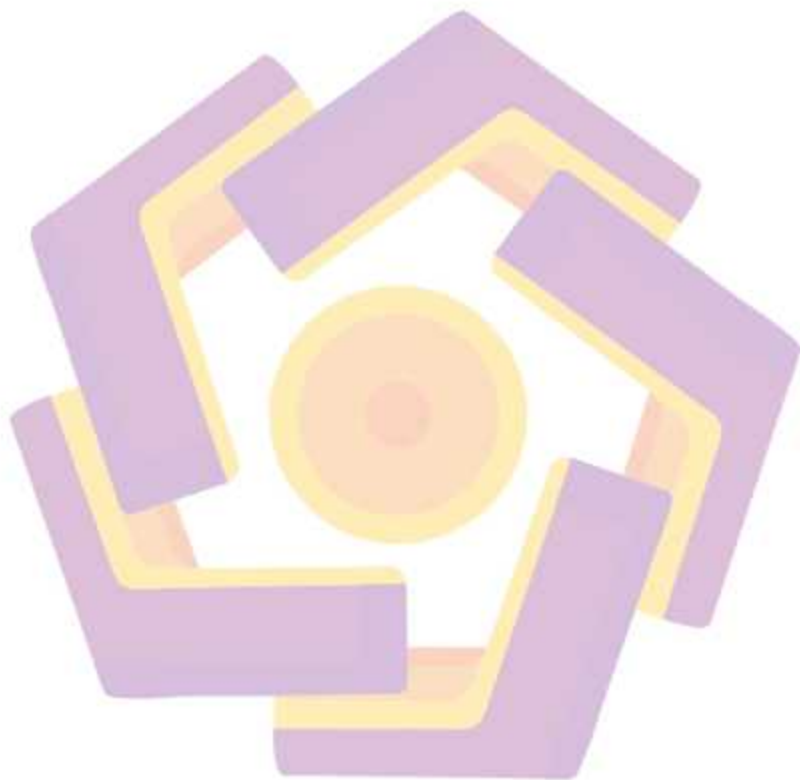
Tabel 2. 1 Tabel Keaslian.....	8
Tabel 2. 2 Kriteria Calon Penerima PKH.....	14
Tabel 2. 3 Confusion Matrix	25
Tabel 2. 4 Indikator Penilaian AUC.....	27
Tabel 3. 1 Dataset.....	32
Tabel 3. 2 Struktur Dataset PKH.....	33
Tabel 3. 3 Statistik Deskriptif	34
Tabel 3. 4 Identifikasi Outlier	35
Tabel 3. 5 Tabel Distribusi Kategorikal.....	37
Tabel 3. 6 Pemeriksaan Kategori yang Tidak Konsisten	39
Tabel 3. 7 Transformasi Data.....	40
Tabel 3. 8 Korelasi Dengan Fitur Kelayakan	41
Tabel 3. 9 Distribusi Fitur Label Dinding dan Label Lantai terhadap Kelayakan.....	43
Tabel 3. 10 Hasil Seleksi Fitur Menggunakan Korelasi	44
Tabel 3. 11 Targer dan Fitur	45
Tabel 3. 12 Split Data.....	45
Tabel 3. 13 Distribusi kelas dalam data training sebelum dan sesudah oversampling menggunakan SMOTE	47
Tabel 3. 14 Split data Sebelum dan Sesudah SMOTE.....	48
Tabel 4. 1 Search Space untuk Hyperparameter Tuning.....	50
Tabel 4. 2 Hasil Optimasi Hyperparameter.....	52
Tabel 4. 3 Performa Model Sebelum dan Sesudah Hyperparameter Tuning.....	53
Tabel 4. 4 Performa Model Sebelum dan Sesudah SMOTE.....	56
Tabel 4. 5 Pengujian Black Box.....	61
Tabel 4. 6 Hasil UAT	64
Tabel 4. 7 Perbandingan Akurasi dan Cross-validation.....	65
Tabel 4. 8 Perbandingan Sensitivity dan Specificity	65
Tabel 4. 9 Perbandingan AUC-Score dan G-Mean.....	66
Tabel 4. 10 Perbandingan Efisiensi Waktu Otomatisasi dan Manual	67

DAFTAR GAMBAR

Gambar 2. 1 Skema kfold cross validation[38].....	29
Gambar 3. 1 Alur Penelitian.....	31
Gambar 3. 2 Outlier Variabel Usia dan Penghasilan	35
Gambar 3. 3 Sebelum Menangani Outlier.....	36
Gambar 3. 4 Setelah Menangani Outlier.....	36
Gambar 3. 5 Struktur Data Setelah Menghapus Kolom.....	38
Gambar 3. 6 Distribusi Fitur Label Dinding Terhadap Label Kelayakan.....	43
Gambar 3. 7 Distribusi Fitur Label Lantai Terhadap Label Kelayakan.....	43
Gambar 3. 8 Distribusi Kelas Sebelum dan sesudah di SMOTE.....	46
Gambar 4. 1 Tampilan Hasil Deployment	60

DAFTAR LAMPIRAN

Lampiran. 1 SIP Dari Kampus	78
Lampiran. 2 Surat Balasan Izin Penelitian	79
Lampiran. 3 Surat Izin Penelitian dari Dinas Sosial	80



INTISARI

Penelitian ini mengimplementasikan data mining untuk mengklasifikasikan kelayakan masyarakat dalam menerima bantuan Program Keluarga Harapan (PKH) di Desa Kaligawe dengan menggunakan algoritma Naïve Bayes dan Random Forest. Selain itu, Synthetic Minority Over-sampling Technique (SMOTE) diterapkan untuk mengatasi ketidakseimbangan kelas dalam data. Hasil penelitian menunjukkan bahwa proses klasifikasi berhasil diotomatisasi, meningkatkan efisiensi, serta mengurangi kesalahan subjektif dibandingkan metode manual. Penerapan SMOTE pada Naïve Bayes meningkatkan akurasi dari 91.92% menjadi 93.07%, specificity dari 93.10% menjadi 100.00%, AUC-score dari 96.81% menjadi 97.11%, G-Mean dari 92.34% menjadi 95.44%, dan akurasi cross-validation dari 91.63% menjadi 95.16%, meskipun sensitivity sedikit menurun dari 91.58% menjadi 91.06%. Sementara itu, pada Random Forest, SMOTE meningkatkan akurasi dari 97.69% menjadi 98.84%, sensitivity dari 97.97% menjadi 98.51%, specificity dari 96.83% menjadi 100.00%, AUC-score dari 99.65% menjadi 99.67%, dan G-Mean dari 97.40% menjadi 99.25%. Secara keseluruhan, model Random Forest menunjukkan performa yang lebih baik dibandingkan dengan Naïve Bayes, terutama dalam hal akurasi, sensitivity, dan AUC-score, meskipun Naïve Bayes memiliki stabilitas yang lebih baik sebelum dan sesudah penerapan SMOTE. Dengan demikian, Random Forest dipilih sebagai model yang lebih unggul dalam klasifikasi kelayakan penerima bantuan PKH di Desa Kaligawe karena performanya yang lebih stabil serta specificity yang lebih tinggi. Selain itu, penggunaan SMOTE terbukti membantu menyeimbangkan distribusi kelas, memungkinkan model mengenali kedua kelas dengan lebih baik. Penelitian ini berhasil mengotomatiskan proses klasifikasi serta memberikan wawasan berharga dalam menentukan kelayakan penerima bantuan PKH.

Kata kunci: Data Mining, Klasifikasi, Program Keluarga Harapan (PKH), Naïve Bayes, Random Forest, Synthetic Minority Over-sampling Technique (SMOTE), Ketidakseimbangan Kelas, Otomatisasi Klasifikasi

ABSTRACT

This study implements data mining to classify the eligibility of communities for receiving assistance from the Family Hope Program (PKH) in Kaligawe Village using the Naïve Bayes and Random Forest algorithms. Additionally, the Synthetic Minority Over-sampling Technique (SMOTE) was applied to address class imbalance in the data. The findings indicate that the classification process was successfully automated, improving efficiency and reducing subjective errors compared to manual methods. The application of SMOTE to Naïve Bayes increased accuracy from 91.92% to 93.07%, specificity from 93.10% to 100.00%, AUC-score from 96.81% to 97.11%, G-Mean from 92.34% to 95.44%, and cross-validation accuracy from 91.63% to 95.16%, although sensitivity slightly decreased from 91.58% to 91.06%. Meanwhile, for Random Forest, SMOTE improved accuracy from 97.69% to 98.84%, sensitivity from 97.97% to 98.51%, specificity from 96.83% to 100.00%, AUC-score from 99.65% to 99.67%, and G-Mean from 97.40% to 99.25%. Overall, the Random Forest model demonstrated better performance compared to Naïve Bayes, particularly in terms of accuracy, sensitivity, and AUC-score, although Naïve Bayes exhibited greater stability before and after SMOTE application. Thus, Random Forest was chosen as the superior model for classifying PKH beneficiary eligibility in Kaligawe Village due to its more stable performance and higher specificity. Additionally, the use of SMOTE proved effective in balancing class distribution, allowing the model to better recognize both classes. This study successfully automated the classification process and provided valuable insights for determining PKH beneficiary eligibility.

Keyword: Data Mining, Classification, Family Hope Program (PKH), Naïve Bayes, Random Forest, Synthetic Minority Over-sampling Technique (SMOTE), Class Imbalance, Classification Automation