

Comparison of Ensemble Classifier Methods, using the K-NN, and Naive Bayes algorithms for the Eligibility of PKH Recipients Case Study (Entalsewu Village, Buduran District, Sidoarjo Regency)

[Perbandingan Metode Ensemble Classifier, dengan menggunakan algoritma K-NN, dan Naive Bayes Kelayakan Penerima PKH Studi Kasus (Desa Entalsewu, Kecamatan Buduran, Kabupaten Sidoarjo)]

Handraswara Pratama ¹⁾, Uce Indahyanti ²⁾

¹⁾ Program Studi Teknik Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia

²⁾ Program Studi Teknik Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia

*Email Penulis Korespondensi: uceindahyanti@umsida.ac.id

Abstract. Poverty represents a socioeconomic challenge where individuals cannot fulfill the standard baseline requirements for basic living within a community. Etymologically derived from 'poor', it signifies material deprivation and severe limitations. To address this, the Indonesian government implemented the Family Hope Program (PKH) as a conditional social assistance scheme aimed at boosting the economic standing of vulnerable and Extremely Poor Households (RTSM). However, mistargeting remains a major obstacle, where facilitators encounter logistical and structural barriers, preventing the optimal implementation of PKH in Entalsewu Village, Buduran, Sidoarjo. To resolve target accuracy issues, this research introduces an Ensemble Classifier approach integrating K-Nearest Neighbor (K-NN) and Naive Bayes models via a Soft Voting mechanism, which averages predicted class probability outputs. Empirical validation demonstrates that the proposed Ensemble Framework effectively yields superior performance, securing 93% classification accuracy under a 50:50 train-test split configuration. Furthermore, execution benchmark analysis reveals an efficient computational overhead with a 6.425-second average run time. Precision metrics achieved 96% for the eligible target group and 91% for ineligible records, accompanied by recall levels of 93% and 95% respectively, culminate in F1-scores of 94% (eligible) and 93% (ineligible).

Keywords – Poverty, PKH, Ensemble, K-NN, Naive Bayes, Soft Voting

Abstrak. Problematika kemiskinan mencerminkan ketidakmampuan kondisi finansial warga dalam memenuhi taraf hidup minimal pada suatu kawasan. Dari sudut pandang etimologi, akar kata 'miskin' mengidentifikasikan kebiasaan hidup tanpa aset berharga serta mengalami keterbatasan materiil. Menyikapi persoalan tersebut, pihak pemerintah menginisiasi skema Program Keluarga Harapan (PKH) berupa jaminan bantuan sosial bersyarat guna mendongkrak taraf perekonomian masyarakat berpenghasilan rendah dan Rumah Tangga Sangat Miskin (RTSM). Kendati demikian, masalah penyaluran yang tidak tepat sasaran masih membayangi pelaksanaan di lapangan, di mana petugas pendamping menghadapi aneka hambatan operasional sehingga capaian program PKH di wilayah Desa Entalsewu, Kecamatan Buduran, Kabupaten Sidoarjo belum menunjukkan hasil optimal. Sebagai solusi teknis, studi ini menerapkan pendekatan Ensemble Classifier yang memadukan model Klasifikasi K-Nearest Neighbor (K-NN) dan Naive Bayes melalui mekanisme Soft Voting berbasis perataan estimasi nilai probabilitas kelas. Hasil pengujian empiris mengonfirmasi keunggulan metode gabungan ini dengan torehan akurasi 93% pada skema pembagian data 50:50. Dari sisi efisiensi waktu, model ensemble sanggup memproses klasifikasi dalam tempo rata-rata 6,425 detik. Capaian presisi mencatatkan angka 96% untuk kelompok layak dan 91% untuk kelompok tidak layak, serta perolehan recall berturut-turut 93% dan 95%, yang menghasilkan F1-score sebesar 94% (layak) dan 93% (tidak layak).

Kata Kunci - Kemiskinan, PKH, Ensemble, K-NN, Naive Bayes, Soft Voting

I. PENDAHULUAN

Penetapan status kemiskinan merujuk pada situasi keterbatasan finansial perseorangan yang membatasi pemenuhan kebutuhan pokok harian secara layak. Mengacu pada penjelasan etimologis, kata dasar 'miskin' mengekspresikan kondisi nir-aset serta serba kekurangan. Lembaga Kementerian Sosial bersama Badan Pusat Statistik mengartikan ranah ini sebagai ketidakmampuan warga memenuhi standar minimum penghidupan [1]. Realita tersebut tecermin dari minimnya tingkat daya beli untuk memperoleh pangan, sandang, serta hunian [2]. Dalam sudut pandang lain, keterbatasan ini diukur melalui besarnya proporsi warga yang gagal menjangkau fasilitas dasar mendasar [3]. Kegagalan penyediaan sarana vital serta jangkauan sumber daya oleh instansi pemerintah kerap memicu eskalasi angka kemiskinan, yang lantas membawa implikasi luas bagi kehidupan publik. Komplikasi kerentanan ekonomi ini

amat rentan menyulut timbulnya gejolak sosial, perilaku menyimpang, peningkatan kriminalitas, gangguan psikologis, hingga penurunan kualitas kesehatan warga [4]. Disamping itu, tertutupnya akses pendidikan bermutu turut memperparah lingkaran kemiskinan ekstrem, mengingat jenjang pendidikan tinggi berbanding lurus dengan peningkatan pendapatan personal [5].

Indonesia sebagai negara yang berkembang dari sisi bidang ekonomi masih terus bergelut memecahkan problematika kemiskinan melalui kerja sama sinergis antara pemerintah dan jajaran elemen masyarakat. Langkah penanganan ini diwujudkan lewat pengguliran program proteksi sosial maupun agenda pemberdayaan ekonomi masyarakat miskin [6]. Salah satu kebijakan strategis pemerintah dalam menekan taraf kemiskinan ditandai dengan peluncuran Program Keluarga Harapan (PKH). Jaminan sosial bersyarat ini menasar perbaikan taraf hidup warga berpenghasilan rendah serta Rumah Tangga Sangat Miskin (RTSM), sebagaimana tertuang pada Undang-Undang Nomor 13 Tahun 2011 dan Permensos RI Nomor 1 Tahun 2018. Penyelenggaraan bantuan PKH diresmikan demi meringankan beban finansial rumah tangga, menjamin pemenuhan jangkauan hidup layak, serta memutus mata rantai kemiskinan antargenerasi [7]. Fokus utama PKH diarahkan pada pengikisan angka kemiskinan ekstrem dan jurang ketimpangan melalui penguatan kapasitas Sumber Daya Manusia (SDM). Hal ini ditempuh dengan mensyaratkan akses layanan kesehatan (ibu hamil dan anak balita), pemenuhan pendidikan (anak usia sekolah), serta pemeliharaan kesejahteraan bagi penyandang disabilitas dan lansia demi terciptanya kemandirian keluarga miskin [8].

Kabupaten Sidoarjo merupakan salah satu kabupaten yang mengimplementasikan skema PKH guna menanggulangi angka kemiskinan kawasan [9]. Laporan BPS Sidoarjo Maret 2024 mencatatkan populasi warga miskin dengan pengeluaran di bawah garis kemiskinan mencapai 109,39 ribu jiwa. Ambang Garis Kemiskinan di Kabupaten Sidoarjo berada pada angka Rp 597.284 per kapita tiap bulan, mengalami kenaikan sebesar Rp 25.588 (4,48%) dibanding Maret 2023 [10]. Desa Entalsewu merupakan salah satu desa yang berada di Kecamatan Buduran Kabupaten Sidoarjo dengan jumlah penduduknya yang berjumlah 7.167 jiwa dan 2.151 KK, pemerintah Desa Entalsewu memiliki permasalahan dalam pengklasifikasian masyarakat tidak mampu. Tingkat kemiskinan yang tergolong cukup rendah di Desa Entalsewu membuat beberapa permasalahan terkait masyarakat tidak mampu terlewat begitu saja akibat dalam beberapa kasus, tingkat ekonomi masyarakat Desa Entalsewu disamaratakan. Permasalahan yang sering terjadi adalah tidak tepat sasaran penerima PKH, masalah ketidaktepatan sasaran PKH di Sidoarjo dan wilayah lain sering muncul karena Data Terpadu Kesejahteraan Sosial (DTKS) yang tidak selalu mutakhir (perlu pembaruan), faktor kelalaian pendamping, hingga adanya penerima yang dianggap mampu namun tetap menerima bantuan. Aparatur masih belum mampu melaksanakan program dengan baik dalam memahami program. Ketidaktepatan sasaran dalam menerima bantuan PKH. Ketidakteraturan penyaluran bantuan juga dapat menimbulkan kecemburuan sosial, ketidakpuasan, dan menurunkan kepercayaan masyarakat terhadap pemerintahan khususnya pemerintah daerah [11]. Pendamping PKH masih memiliki beberapa kendala yang menyebabkan tidak tercapainya sasaran tersebut. Hal ini membuat efektivitas PKH di Desa entalsewu Kecamatan Buduran kabupaten Sidoarjo belum efektif. Perlunya Sebuah sistem klasifikasi yang mampu mendukung proses pendataan keluarga penerima PKH secara lebih akurat merupakan jawaban dari persoalan tersebut. Sistem klasifikasi tersebut dapat digunakan untuk menentukan kategori keluarga mampu dan tidak mampu berdasarkan variabel atau indikator yang digunakan dalam proses klasifikasi.

Pendekatan proses klasifikasi berperan dalam memetakan pola data sampel ke dalam kategori target yang telah ditentukan [12], sehingga dapat memprediksi label data baru di masa mendatang. Penggabungan algoritma K-Nearest Neighbor (K-NN) dan Naive Bayes dalam kerangka Ensemble Classifier terbukti mampu mendongkrak performa akurasi sekaligus efisiensi waktu pemrosesan. Teknik ensemble bekerja membangun sekumpulan model pembelajaran dasar (base classifiers) untuk kemudian dikombinasikan melalui skema pembobotan, bagging, boosting, atau stacking [13]. Muklisah dan kawan-kawan menyebutkan bahwa tahap klasifikasi setiap data akan dihitung menggunakan *K-Nearest Neighbour* (K-NN), dan *Naive Bayes* [14]. Penelitian mengenai implementasi teknik *Ensemble Stacking* pada klasifikasi penyakit anemia memperoleh nilai akurasi sebesar 95,77%. Nilai akurasi tersebut tergolong tinggi untuk sebuah teknik klasifikasi yang diterapkan pada jumlah data mencapai ribuan. Penelitian yang berjudul Pemilihan model optimal untuk ensemble k-nearest neighbours melalui sub-bagging dan sub-sampling dengan pembobotan fitur menghasilkan, hasilnya menunjukkan dengan jelas bahwa metode ensemble yang diusulkan mengungguli prosedur standar lainnya seperti weighted K-NN dan Random K-NN [15] pada sebagian besar dataset. Penggunaan resampling dan pembobotan fitur yang dikombinasikan dengan pemilihan model menjamin keragaman dan akurasi model, yang merupakan intuisi dasar di balik efektivitas metode yang diusulkan tersebut. Penerapan Ensemble juga dapat mengungguli metode K-NN Tunggal dengan memberikan tingkat MAPE 2,1% pada penelitian lain Penggabungan model ini juga mampu menurunkan tingkat galat MAPE hingga 2,1% pada analisis Indeks Pembangunan Manusia [16]. Metode Naive Bayes pun juga memiliki keunggulan jika memakai teknik ensemble, Penelitian Dewa Raka Krisna Saputra(2024) [17]. yang berfokus pada transaksi keuangan, teknik naive bayes yang dipadukan ensemble dalam mendeteksi anomali transaksi keuangan dapat menghasilkan Tingkat *Precision* sebesar 97,56% dibandingkan dengan penelitian sebelumnya [18]. Tingkat *Precision* sebesar 77,4% jika menggunakan 1 model klasifikasi dengan menggunakan Naive Bayes.

Pada penelitian ini, hasil perhitungan klasifikasi K-NN dan Naïve Bayes diproses kembali menggunakan metode *Ensemble Classifier* dengan teknik *Soft Voting* (rata-rata nilai klasifikasi) yang mengambil nilai rata – rata pada klasifikasi dengan menggabungkan hasil dari metode K-NN dan Naïve Bayes. Teknik *Soft Voting* sering diunakan untuk mengoptimasikan hasil klasifikasi. Pada penelitian [19]. Teknik *Soft Voting* digunakan untuk mengoptimalkan hasil akurasi dari klasifikasi. Teknik *Soft Voting* bekerja dengan cara setiap model klasifikasi menghasilkan nilai probabilitas untuk setiap kelas (*Average of Probabilities*). Nilai probabilitas dari masing-masing model tersebut kemudian dihitung rata-ratanya untuk menentukan hasil prediksi akhir berdasarkan keluaran dari model *Ensemble Classifier*.

II. METODE

Klasifikasi merupakan proses evaluasi objek data untuk dikelompokkan ke dalam kategori tertentu yang telah didefinisikan. Tujuan utama dari tugas klasifikasi dalam pembelajaran mesin adalah memetakan fitur masukan terhadap kelas target secara akurat dan presisi [20]. Pemrosesan klasifikasi mencakup dua tahapan inti: pembentukan model prototype sebagai basis pengetahuan, serta penerapannya untuk memprediksi label pada data baru [21]. Gambar 1 merupakan proses penerapan model klasifikasi ensemble yang terdiri dari 5 proses.



Gambar 1. Proses Pembentukan Model Klasifikasi

A. Pengumpulan Data

Tahap awal penelitian diawali dengan penghimpunan data primer dari masyarakat Desa Entalsewu dengan variabel – variabel yang akan dipakai yaitu : Jenis Kelamin Kepala Keluarga, Umur, Pendidikan, Pekerjaan dari Kepala Keluarga, Penghasilan, Tanggungan dari Kepala Keluarga, Status Perkawinan dari Kepala Keluarga, Kondisi Rumah, Status Pemilik Rumah, Jumlah Anggota Keluarga Yang Bekerja, Penghasilan Anggota Keluarga Yang Bekerja. Kemudian mengumpulkan jurnal serta studi literatur yang berkaitan dengan permasalahan kemiskinan, PKH dan metode klasifikasi penelitian yang digunakan dalam penelitian ini yaitu Naive Bayes, K-NN dan *ensemble clasifier (Soft Voting)*.

B. Preprocessing

Proses Preprocessing adalah proses data mentah diolah menjadi data yang siap untuk diklasifikasikan terdapat beberapa tahapan seperti cleaning, handling missing value (penanganan nilai hilang), encoding kategorikal, normalisasi (untuk K-NN) untuk jenis data label yang bersifat numerical menggunakan normalisasi min max scaling dengan skala 0 sampai 1. Adanya proses preprocessing membuat model klasifikasi lebih mudah mengenali pola, memiliki proses pelatihan yang lebih stabil, dan berpotensi menghasilkan akurasi serta kemampuan generalisasi yang lebih baik[22].

C. Pembagian Data Latih

Pembagian data latih (training data) adalah proses krusial dalam klasifikasi [23]. Metode dan rasio pembagiannya tergantung pada ukuran kumpulan data (dataset). Pada penelitian ini akan menggunakan data Train-test split. Teknik Train-test split untuk mengevaluasi atau menguji data yang sebelumnya telah digunakan untuk melatih model [24].

D. Implemeptasi Pembuatan Dan Pelatihan Model (*Model Development and Training*)

Pelatihan Model pada Klasifikasi Naïve Bayes akan menggunakan Train-test split dengan Rasio 50:50, 60:40, dan 75:25 untuk data latih : data ujinya. Pelatihan Model pada Klasifikasi K-NN akan menggunakan Train-test split dengan Rasio 50:50, 60:40, dan 75:25 untuk data latih : data ujinya dengan nilai K = 1,3,5,7,9.

Penentuan parameter K yang presisi sangat krusial untuk mencegah kecenderungan overfitting maupun underfitting [25].

Pelatihan Model pada Klasifikasi ensemble Classifier menggunakan teknik Average of Probabilities yaitu Salah satu Teknik *Soft Voting* yang mengambil nilai rata – rata pada klasifikasi dengan menggabungkan metode K-NN

dan Naïve Bayes. Ketika memprediksi kelas untuk data baru, setiap model menghasilkan probabilitas untuk setiap kelas. Probabilitas ini kemudian dirata-ratakan untuk menentukan prediksi akhir.

Pada proses ini juga digunakan sebuah implementasi dengan melakukan pembuatan program untuk mengklasifikasikan menggunakan teknik Average of Probabilities yaitu Salah satu Teknik Soft Voting yang mengambil nilai rata – rata pada klasifikasi dengan mengabungkan metode K-NN dan Naïve Bayes. Selain nilai akurasi digunakan juga perbandingan performa dari setiap metode dengan menggunakan waktu respons (response time) sebagai acuan dalam memproses klasifikasi baik dengan data sedikit ataupun banyak.

E. Evaluasi Model (Model Evaluation)

Pengujian model klasifikasi nantinya dilakukan memakai akurasi, Confusion Matrix(precision, recall)I dan F1 Score serta balanced accuracy bila kelas tidak seimbang.

III. HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Pada tahap pengumpulan data peneliti melakukan pengumpulan data profil warga Desa Entalsewu sebagai basis pengujian klasifikasi kelayakan penerima bantuan PKH. Kategori keluaran kelas dibagi menjadi dua kelas: 'Mampu' (Tidak Berhak Menerima PKH) dan 'Tidak Mampu' (Berhak Menerima PKH). Data tersebut dikumpulkan berdasarkan variabel-variabel yang digunakan dalam proses klasifikasi, yaitu nama, nomor Kartu Keluarga (KK), jenis kelamin kepala keluarga, umur, tingkat pendidikan, pekerjaan kepala keluarga, penghasilan, jumlah tanggungan kepala keluarga, status perkawinan kepala keluarga, kondisi rumah, status kepemilikan rumah, jumlah anggota keluarga yang bekerja, serta penghasilan anggota keluarga yang bekerja. Contoh data masyarakat Desa Entalsewu yang digunakan dalam penelitian ini disajikan sebagai berikut.

Tabel 1 Data Masyarakat Entalsewu

NOMOR	NAMA	NO KK	JENIS KELAMIN	UMUR	PENDIDIKAN	GAJI	STATUS KAWIN
1	Rendi Kurniawan	-	Laki-Laki	46	D3/S1/S2	4.500.000	Kawin
JUMLAH TANGGUNGAN	KONDISI KESEHATAN	JUMLAH KELUARGA BEKERJA	GAJI KELUARGA BEKERJA	KONDISI RUMAH	KEPEMILIKAN RUMAH	PEKERJAAN	KELAS
3	SEHAT	2	4.000.000	LAYAK	MILIK SENDIRI	NOTARIS	MAMPU

Tabel 1 merupakan Tabel data masyarakat entalsewu yang nantinya diolah terlebih dahulu sebelum dimasukkan ke dalam sistem untuk diklasifikasikan menggunakan metode K-NN dan Naïve Bayes dan ensemble Classifier menggunakan Teknik soft voting. Berikut adalah Gambaran sekaligus alur dari proses klasifikasi pada penelitian. Terdapat Beberapa variabel terhadap data yang nantinya diolah dan diproses dalam klasifikasi Kelayakan data tersebut dalam kelas mampu atau tidak mampu.



Gambar 2. Proses Klasifikasi

Pada gambar 2 merupakan gambaran sistem yang akan dibuat dalam penelitian dimana data masyarakat entalsewu akan dimasukkan ke dalam sistem klasifikasi metode ensemble yang nantinya hasil outputnya adalah Status kelas Berhak Menerima PKH (Tidak Mampu) atau Tidak berhak Menerima (Mampu).

B. Preprocessing Data

Proses ini merupakan serangkaian teknik untuk mengubah data mentah (raw data) menjadi data yang bersih dan terstruktur sehingga dapat dipahami dengan baik oleh algoritma baik oleh Naïve Bayes dan K-NN.

- **Cleaning Data**

Pada proses ini akan dilakukan cleaning data Membuang baris (record) atau kolom (fitur) jika nilai hilangnya sangat sedikit contoh pada Data Masyarakat Entalsewu, tidak adanya no KK tidak memberikan pengaruh pada variabel dan perhitungan klasifikasi kelas sehingga kolom tersebut dapat dihilangkan. Menghapus data yang identik atau hampir identik untuk menghindari bias pada model Pada data yang dihimpun data yang identik dengan nama yang sama akan dihapus.

- **Handling Missing Value**

Pada proses ini data akan diproses menggunakan Imputation Mengisi nilai hilang dengan nilai pengganti contoh pada data Masyarakat Entalsewu pada variabel jumlah dan gaji keluarga yang bekerja jika data kosong maka dapat digantri dengan nilai 0 untuk proses variabel numerical. Pada pekerjaan misalkan terdapat data dengan pekerjaan notaris akan ditransformasikan menjadi pekerjaan dengan kategori pegawai swasta. Begitu juga dengan data lain seperti mekanik akan ditransformasikan menjadi pegawai swasta.

- **Encoding Kategorikal**

Pada proses ini data akan diubah data teks (kategori) menjadi angka karena pada poses Klasifikasi menggunakan K-NN hanya paham angka. Berikut adalah beberapa contoh variabel yang bersifat kategorikal diubah menjadi angka dalam bentuk table

Tabel 2. Tabel Kategori Dan Kode Data

VARIABEL	KATEGORI	KODE
Jenis Kelamin	Laki Laki	1
	Perempuan	0
Status Kawin	Sudah	1
	Belum	0
Kondisi Rumah	Baik	1
	Tidak	0
Kepemilikan Rumah	Milik Sendiri	1
	Kontrak/Tidak Milik Sendiri	0

VARIABEL	KATEGORI	KODE
Kondisi Kesehatan	Sehat	1
	Tidak Sehat	0
Pendidikan (Ordinal)	Sd	0
	Smp	1
	Sma	2
	S1	3
Pekerjaan (Binary)	Wirawasta	11
	Pegawai Swasta	01
	Pengawai Negeri Sipil/Bumn	10
	Tidak Bekerja	00

Pada tabel 2 data kategori diubah menjadi bentuk numerik agar dapat diproses oleh metode klasifikasi. Variabel dengan dua kondisi menggunakan kode biner (0 dan 1). Variabel pendidikan menggunakan skala ordinal (0–3) karena memiliki tingkatan dari rendah ke tinggi. Sedangkan variabel pekerjaan menggunakan kode biner 2 digit untuk membedakan beberapa kategori tanpa menunjukkan urutan antar pekerjaan.

- **Normalisasi Data**

Pada tahap ini menggunakan Min-Max dalam klasifikasi berfungsi seperti menyamakan variabel yang bersifat numerikal. Dengan menyamakan skala semua fitur ke 0-1, algoritma dapat memproses dan membandingkan informasi secara lebih adil dan efisien, yang pada akhirnya meningkatkan kualitas dan akurasi model klasifikasi. Contoh pada data gaji akan dinormalisasikan sebagai berikut

Gaji Tertinggi Pada data Masyarakat Entalsewu : Rp. 10.000.000, Gaji Terendah Pada data Masyarakat Entalsewu : Rp. 0 , Contoh Data yang dipakai adalah gaji dengan data bernama Rendi Kurniawan dengan nilai gaji Rp. : 4.000.000 Maka proses min-max scalling sebagai berikut.

$$\frac{4.000.000 - 0}{10.000.000 - 0} = 0,4.$$

Maka nilai normalisasi Gaji dari data Rendi Kurniawan adalah 0,4 Hal ini juga berlaku untuk variabel Umur, kategori tanggungan, jumlah keluarga bekerja dan gaji keluarga yang bekerja.

C. Pembagian Data

Dataset yang mencakup 98 rekaman data masyarakat dibagi menjadi subset data latih (training) dan data uji (testing) menggunakan teknik Train-Test Split [23][24]. Penelitian ini menguji tiga rasio pembagian untuk mengevaluasi konsistensi model, seperti tercantum pada Tabel 3.

Tabel 3 Pembagian Data Latih dan Data Uji

NO	RASIO	DATA LATIH	DATA UJI
1	50:50	49	49
2	60:40	59	39
3	75:25	73	25

D. Implementasi Pembuatan dan Pelatihan Model (Model Development and Training)

Pada proses ini akan dibuatkan implementasi sebuah model klasifikasi dan sekaligus pelatihan model dengan Train-test split serta pembagian Metode data Train-test split dengan Rasio 50:50, 60:40, dan 75:25 untuk data latih : data ujinya, menjadikannya ideal untuk data yang tidak seimbang (imbalanced data) agar evaluasi model lebih adil dan representative.

Gambar 3. Form Sistem Klasifikasi

Pada gambar 3 merupakan form sistem klasifikasi, user masuk dengan memasukkan username serta password kemudian user akan masuk ke dalam aplikasi.

Form Atur Jumlah Data Testing

Jumlah Data testing

20

**Jumlah Data Testing Tidak boleh Melebihi 50% (49) dari data Penduduk
Jumlah Data Training Saat Ini : 77 Data**

SET DATA

Gambar 4. Form Atur Jumlah Data Testing

Pada gambar 4 merupakan form untuk menentukan jumlah data latih dan data testing yang akan dipakai untuk proses klasifikasi.

Form Klasifikasi

Jumlah Data Training 77 Jumlah Data Testing 20

PROSES NAIVE BAYES

Gambar 5. Form Klasifikasi Naive Bayes

Pada gambar 5 merupakan form sistem klasifikasi untuk naïve bayes user terlebih menentukan jumlah data latih dan data uji pada gambar 4 kemudian klik proses naïve bayes dan user akan diarahkan ke gambar 6.

Form Testing KNN

Jumlah Data Training : 77 Jumlah Data Testing : 20

Pilih Nilai K ▼

HITUNG KNN

NO.	NAMA	KELAS	PREDIKSI NAIVE BAYES	NILAI NAIVE BAYES KELAS MAMPU	NILAI NAIVE BAYES KELAS TIDAK MAMPU
1	ANDHIKA PRAMUDIYA	TIDAK MAMPU	TIDAK MAMPU	0.000000720396	0.000019129204
2	FANNY AMRIZAL	MAMPU	MAMPU	0.000195400178	0.000000000000
3	GILANG DARMAWANGSA	TIDAK MAMPU	TIDAK MAMPU	0.000012783515	0.000034223242
4	SUPRAPTO	TIDAK MAMPU	MAMPU	0.000107766398	0.000075117605
5	AJI SUWONDO	TIDAK MAMPU	TIDAK MAMPU	0.000000000000	0.000000038532
6	ITA NOVARITA	MAMPU	TIDAK MAMPU	0.000020572656	0.000080975966
7	NUNIK	MAMPU	TIDAK MAMPU	0.000000627145	0.000000930081
8	BOEDI ADIWIDJAJA	MAMPU	MAMPU	0.000994799309	0.000000000000
9	HARI SUBANDI	TIDAK MAMPU	TIDAK MAMPU	0.000001146706	0.000239115572

Gambar 6. Hasil Klasifikasi Menggunakan Naïve Bayes

Pada gambar 6 merupakan hasil keluaran sistem klasifikasi menggunakan metode naïve bayes, untuk melanjutkan proses perhitungan klasifikasi K-NN user terlebih dahulu memilih Nilai K kemudian klik hitung K-NN dan user akan diarahkan ke form berikut.

Hasil Klasifikasi Esemble(Average)												
Jumlah Data Training = 58 Nilai Akurasi Naive = 82.05% Nilai Error Naive = 17.95%			Jumlah Data Testing = 39 Nilai Akurasi KNN = 74.36% Nilai Error KNN = 25.64%			Nilai K = 3 Nilai Akurasi Esemble = 74.36% Nilai Error Esemble = 25.64%			Waktu Eksekusi NAIVE BAYES = 1.7119280 Second Waktu Eksekusi KNN = 39.335310 Second Waktu Eksekusi Esemble = 7.3381828 Second			
NO.	NAMA	KELAS	PREDIKSI NAIVE BAYES	NILAI NAIVE KELAS MAMPU	NILAI NAIVE KELAS TIDAK MAMPU	PREDIKSI KNN	NILAI KNN KELAS MAMPU	NILAI KNN KELAS TIDAK MAMPU	TETANGGA TERDEKAT	PREDIKSI EMSEMBLE	NILAI EMSEMBLE KELAS MAMPU	NILAI EMSEMBLE KELAS TIDAK MAMPU
1	MOCH HUSEIN	MAMPU	MAMPU	0.000000201425	0.000000000000	MAMPU	0.6667	0.3333	●	MAMPU	0.3334	0.1666
2	ADYATMA SHOHAN WIRYAWAN	TIDAK MAMPU	TIDAK MAMPU	0.000031670021	0.000209339574	TIDAK MAMPU	0.3333	0.6667	●	TIDAK MAMPU	0.1667	0.3335
3	IRWAN PURNOMO	MAMPU	MAMPU	0.000031126125	0.000000014761	MAMPU	1.0000	0.0000	●	MAMPU	0.5000	0.0000
4	WIDIJAYA SEOHARJO	TIDAK MAMPU	TIDAK MAMPU	0.000000000000	0.000000001194	TIDAK MAMPU	0.0000	1.0000	●	TIDAK MAMPU	0.0000	0.5000
5	BAGUS HADI WIJAYA	MAMPU	MAMPU	0.000009448221	0.000000000000	MAMPU	1.0000	0.0000	●	MAMPU	0.5000	0.0000
6	HIDAYAT	TIDAK MAMPU	TIDAK MAMPU	0.000000000000	0.000001347867	TIDAK MAMPU	0.0000	1.0000	●	TIDAK MAMPU	0.0000	0.5000
7	M. SYAFIUL	MAMPU	MAMPU	0.000048492377	0.000001410774	MAMPU	1.0000	0.0000	●	MAMPU	0.5000	0.0000

Gambar 7. Hasil Klasifikasi Ensemble

Pada gambar 7 merupakan hasil sistem klasifikasi untuk ensemble menggunakan soft-voting dengan menghitung rata – rata hasil keluaran dari klasifikasi naïve bayes dan K-NN, terdapat akurasi serta error terhadap hasil klasifikasi data uji pada gambar 7. Berikut adalah hasil pelatihan model berdasarkan Metode data Train-test split dengan Rasio 50:50, 60:40, dan 75:25 untuk data latih : data ujinya serta pengukuran akurasi perbandingan antara metode naïve bayes , K-NN serta metode ensemble.

Tabel 3 Tabel Perbandingan Pengujian

Akurasi(%)						
	Rasio	Data Latih	Data Uji	Naïve Bayes	K-NN	Ensemble
1	50:50	49	49	88	72	93
2	60:40	59	39	82	71	88
3	75:25	73	25	82	71	87

Pada tabel 3 memperlihatkan tabel perbandingan pengujian akurasi, dari pengujian akurasi didapatkan nilai akurasi tertinggi terdapat pada Rasio 50:50 dengan masing – masing metode menghasilkan nilai akurasi pada naïve bayes sebesar 88%, Pada K-NN 72% sedangkan pada ensemble di angka 93%.

Tabel 4 Perbandingan Waktu Eksekusi Per metode
Waktu Eksekusi(Second)

Waktu Eksekusi(Second)						
No.	Rasio	Data Latih	Data Uji	Naïve Bayes	K-NN	Ensemble
1	50:50	49	49	7,455	26,831	7,402
2	60:40	59	39	7,118	21,253	6,112
3	75:25	73	25	7,945	28,067	5,762
Rata - Rata				7,506	25,384	6,425

Pada tabel 4 memperlihatkan waktu eksekusi per metode, Pada pengujian tersebut diperoleh rata – rata untuk eksekusi menggunakan ensemble memiliki waktu eksekusi paling cepat dibandingkan dengan 2 metode lainnya yaitu dengan rata – rata waktu sebesar 6.425 Detik. Dapat dipastikan juga jumlah data uji mempengaruhi waktu

eksekusi dari setiap metode, yang apabila semakin besar datanya maka waktu yang dibutuhkan untuk memproses semakin Panjang.

Tabel 5 Perbandingan Precision dan Recall Per metode

Jenis	Kelas	Naïve		
		Bayes	K-NN	Ensemble
Precision	Mampu	88%	72%	96%
	Tidak	88%	70%	91%
	Mampu	88%	78%	93%
Recall	Tidak	88%	64%	95%

Pada tabel 5 merupakan tabel perbandingan precision dan recall per metode didapatkan Nilai tertinggi pada metode ensemble dengan nilai precision untuk kelas mampu 96% dan kelas tidak mampu sebesar 91%, Sedangkan Recall kelas mampu sebesar 93% dan Recall kelas tidak mampu sebesar 95%. Perhitungan F1 Score per metode sebagai berikut.

Tabel 6 Perhitungan F1 Score

Kelas	Naïve		
	Bayes	K-NN	Ensemble
Mampu	88%	75%	94%
Tidak	88%	67%	93%

Pada tabel 6 merupakan tabel perbandingan Perhitungan F1 Score per metode didapatkan Nilai tertinggi pada metode ensemble dengan nilai F1 Score untuk kelas mampu 94% dan kelas tidak mampu sebesar 93%.

V. SIMPULAN

Berdasarkan serangkaian pengujian dan analisis yang telah dilakukan, dapat ditarik beberapa poin kesimpulan utama:

1. Metode Ensemble dapat digunakan secara baik untuk melakukan proses klasifikasi ini dibuktikan dengan nilai akurasi yang mencapai 93% pada rasio 50:50 Data Latih terhadap data uji, jika menggunakan ensemble lebih baik dibandingkan dengan kedua metode lainnya.
2. Pada waktu eksekusi metode ensemble lebih cepat jika dibandingkan dengan kedua metode lainnya kecepatan eksekusi jika menggunakan ensemble mencapai rata – rata 6,425 detik, Berbeda tipis dengan klasifikasi jika menggunakan metode naïve bayes atau K-NN saja. Kecepatan dalam waktu eksekusi juga dipengaruhi banyaknya jumlah data uji yang diujikan.
3. Pada pengujian precision dan recall didapatkan nilai tertinggi pada metode ensemble dengan nilai precision untuk kelas mampu 96% dan kelas tidak mampu sebesar 91%, Sedangkan Recall kelas mampu sebesar 93% dan Recall kelas tidak mampu sebesar 95%. Pada pengujian F1 Score metode ensemble dengan nilai F1 Score untuk kelas mampu 94% dan kelas tidak mampu sebesar 93%.

UCAPAN TERIMA KASIH

Apresiasi dan ucapan terima kasih disampaikan kepada Universitas Muhammadiyah Sidoarjo atas fasilitasi dan kesempatan akademis selama masa studi. Penghargaan juga diberikan kepada tim pengelola laboratorium informatika atas bantuan teknis selama pelaksanaan praktikum dan pengujian sistem, serta kepada Ketua Program Studi Teknik Informatika atas kepemimpinan dan arahan yang inspiratif.

REFERENSI

- [1] A. C. Putri, F. E. Hariyanto, N. L. E. Andini, and Z. C. S. Zulkarnaen, "Klasifikasi Rumah Tangga Miskin di Provinsi Papua Tahun 2017 Menggunakan Metode Naive Bayes," *Jurnal Sains Matematika dan Statistika*, vol. 7, no. 1, p. 89, Mar. 2021, doi: 10.24014/jsms.v7i1.11924.

- [2] A. N. Wongkar¹, T. O. Rotinsulu², M. T. B. Maramis^{3 1, 2, 3}program Studi, E. Pembangunan, F. Ekonomi, and D. Bisnis, “Pengaruh Ipm, Tingkat Pengangguran Terbuka Dan Belanja Pemerintah Terhadap Kemiskinan Di Kabupaten Bolaang Mongondow.”
- [3] M. P. Toadro and S. C. Smith, *economic-development-todaro-and-smith*, 11th ed. Pearson, 2011.
- [4] N. E. Khairiroh, “Kemiskinan di indonesia: Analisis penyebab, dampak, dan solusi kebijakan,” 2025. [Online]. Available: <http://urj.uin-malang.ac.id/index.php/mij/index>
- [5] H. Hindun, A. Soejoto, and H. Hariyati, “Pengaruh Pendidikan, Pengangguran, dan Kemiskinan terhadap Ketimpangan Pendapatan di Indonesia,” *Jurnal Ekonomi Bisnis dan Kewirausahaan*, vol. 8, no. 3, p. 250, Dec. 2019, doi: 10.26418/jebik.v8i3.34721.
- [6] C. Clara Puspita and L. S. Uguy, “Sustainability in Action: Transformative Strategies in Management and Accounting,” 2024.
- [7] B. Setiawati *et al.*, “Evaluasi Pelaksanaan Prorgam Keluarga Harapan (Pkh) Bidang Pendidikan Di Desa Lumbang Kecamatan Muara Uya Kabupaten Tabalong.”
- [8] M. L. Rohmi, P. Studi, A. Keuangan, P. N. Jakarta, and A. A. Ramadhani, “Program Keluarga Harapan (PKH) Dalam Kaitannya Dengan Pengembangan Kualitas Sumber Daya Manusia Sebagai Upaya Pengentasan Kemiskinan Desyria Pratiwi.”
- [9] O. : Farid, A. Muluk, and I. D. Rahmawati, “Pt. Media Akademik Publisher Efektivitas Program Keluarga Harapan (Pkh) Dalam Mengurangi Garis Kemiskinan Kecamatan Sidoarjo, Kabupaten Sidoarjo,” *JMA*, vol. 2, no. 7, pp. 3031–5220, 2024, doi: 10.62281.
- [10] Badan Pusat Statistik Sidoarjo, “Profil Kemiskinan di Kabupaten Sidoarjo Maret 2024,” Sidoarjo, Mar. 2024.
- [11] Y. D. Arimbi, “Kecemburuan Sosial Masyarakat Non Penerima Manfaat PKH di Dusun Pringroto Desa Punjung Kecamatan Kebonagung Kabupaten Pacitan,” *Jurnal Pengabdian UntukMu NegeRI*, vol. 6, no. 1, pp. 163–167, Jun. 2022, doi: 10.37859/jpumri.v6i1.3280.
- [12] R. L. Hasanah *et al.*, “Klasifikasi Penerima Dana Bantuan Desa Menggunakan Metode Knn (K-Nearest Neighbor),” *Jurnal TECHNO Nusa Mandiri*, vol. 16, no. 1, p. 1, 2019, [Online]. Available: <http://nusamandiri.ac.id/>
- [13] H. Chen, G. Zhang, X. Pan, and R. Jia, “Using dual evolutionary search to construct decision tree based ensemble classifier,” *Complex and Intelligent Systems*, vol. 9, no. 2, pp. 1327–1345, Apr. 2023, doi: 10.1007/s40747-022-00855-x.
- [14] N. Muklisah, R. Indira, A. Desiani, N. R. Dewi, and Y. Andriani, “Implementasi Teknik Ensemble Stacking pada Klasifikasi Penyakit Anemia.”
- [15] N. Gul, W. K. Mashwani, M. Aamir, S. Aldahmani, and Z. Khan, “Optimal model selection for k-nearest neighbours ensemble via sub-bagging and sub-sampling with feature weighting,” *Alexandria Engineering Journal*, vol. 72, pp. 157–168, Jun. 2023, doi: 10.1016/j.aej.2023.03.075.
- [16] Y. Pristyanto, “Penerapan Metode Ensemble Untuk Meningkatkan Kinerja Algoritme Klasifikasi Pada Imbalanced Dataset,” 2019. [Online]. Available: <https://archive.ics.uci.edu/ml/datasets/User+Knowledge>
- [17] D. R. K. Saputra, Y. V. Via, and A. N. Sihananto, “Deteksi Anomali Menggunakan Ensemble Learning Dan Random Oversampling Pada Penipuan Transaksi Keuangan,” *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3, Aug. 2024, doi: 10.23960/jitet.v12i3.4910.
- [18] R. Sudiyarno, A. Setyanto, and E. T. Luthfi, “Peningkatan Performa Pendeteksian Anomali Menggunakan Ensemble Learning dan Feature Selection Anomaly Detection Performance Improvement Using Ensemble Learning and Feature Selection,” *Citec Journal*, vol. 7, no. 1, 2020.
- [19] T. Misriati, R. Aryanti, and H. Leidiyana, “Soft Voting Based Optimized Ensemble for Migraine Type Classification,” vol. 6, no. 2, pp. 266–275, 2025, doi: 10.33365/jatika.v6i3.861.
- [20] M. Ibnu and C. Rachmatullah, “IMPROVE Ensemble Learning untuk Klasifikasi: Tinjauan Komprehensif Metode, Aplikasi, dan Perkembangan Terkini,” 2025.
- [21] E. Prasetyo, *Data Mining : Konsep dan Aplikasi Menggunakan MATLAB*. Yogyakarta: Andi, 2012.
- [22] I. Dagal, A. Harrison, A. W. Ibrahim, and W. F. Mbasso, “Comprehensive evaluation of data preprocessing and visualization techniques for enhanced classification and sampling,” *Cluster Comput.*, vol. 28, no. 7, Sep. 2025, doi: 10.1007/s10586-025-05512-9.
- [23] Erfin Nur Rohma, “Comparison of the Social Welfare Data Classification Algorithm for Bantul Regency,” *Jurnal PROCESSOR*, vol. 17, no. 2, pp. 91–100, Oct. 2022, doi: 10.33998/processor.2022.17.2.1222.
- [24] E. Nirmala and S. Kom, *Metode Klasifikasi Modern DalamMachine Learning Menggunakan Python*. 2026. [Online]. Available: www.mediapustakaindo.com
- [25] N. F. Fahrudin, K. R. Putra, S. Umaroh, and Lautan Gamas Bloory, “Influence of Data Scaling and Train/Test Split Ratios on LightGBM Efficacy for Obesity Rate Prediction,” *MIND Journal*, vol. 9, no. 2, pp. 220–234, Dec. 2024, doi: 10.26760/mindjournal.v9i2.220-234.

Conflict of Interest Statement:

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.