

PERBANDINGAN METODE KNN, NAIVE BAYES, DAN ENSEMBLE CLASSIFIER (*AVERAGE OF PROBABILITIES*) UNTUK KLASIFIKASI KELAYAKAN PENERIMA PKH STUDI KASUS (DESA ENTALSEWU, KECAMATAN BUDURAN, KAB. SIDOARJO)

Oleh :
Handraswara Pratama,
Dr. Uce Indahyanti, S.Kom., M.Kom
Program Studi Informatika
Universitas Muhammadiyah Sidoarjo
2026

Latar Belakang

Kemiskinan adalah suatu kondisi ketidakmampuan seseorang secara ekonomi untuk memenuhi standar hidup rata-rata masyarakat di suatu daerah.



Permasalahan yang sering terjadi Adalah tidak tepat sasaran penerima program keluarga harapan, Masalah ketidaktepatan sasaran Program Keluarga Harapan (PKH) di Sidoarjo dan wilayah lain sering muncul karena data yang tidak selalu mutakhir (DTKS perlu pembaruan)



Rumusan Masalah

Bagaimana membangun Model KNN dan Naive Bayes dalam mengklasifikasi kelayakan penerima PKH pada data yang digunakan yaitu data Masyarakat desa entalsewu Kecamatan Buduran Kabupaten Sidoarjo?

Bagaimana mengevaluasi Model Klasifikasi mana yang paling baik dan cocok untuk klasifikasi Program PKH pada data yang digunakan yaitu data Masyarakat desa entalsewu Kecamatan Buduran Kabupaten Sidoarjo berdasarkan metrik evaluasi (akurasi, *precision*, *recall*, *F1 Score*)?

Tujuan

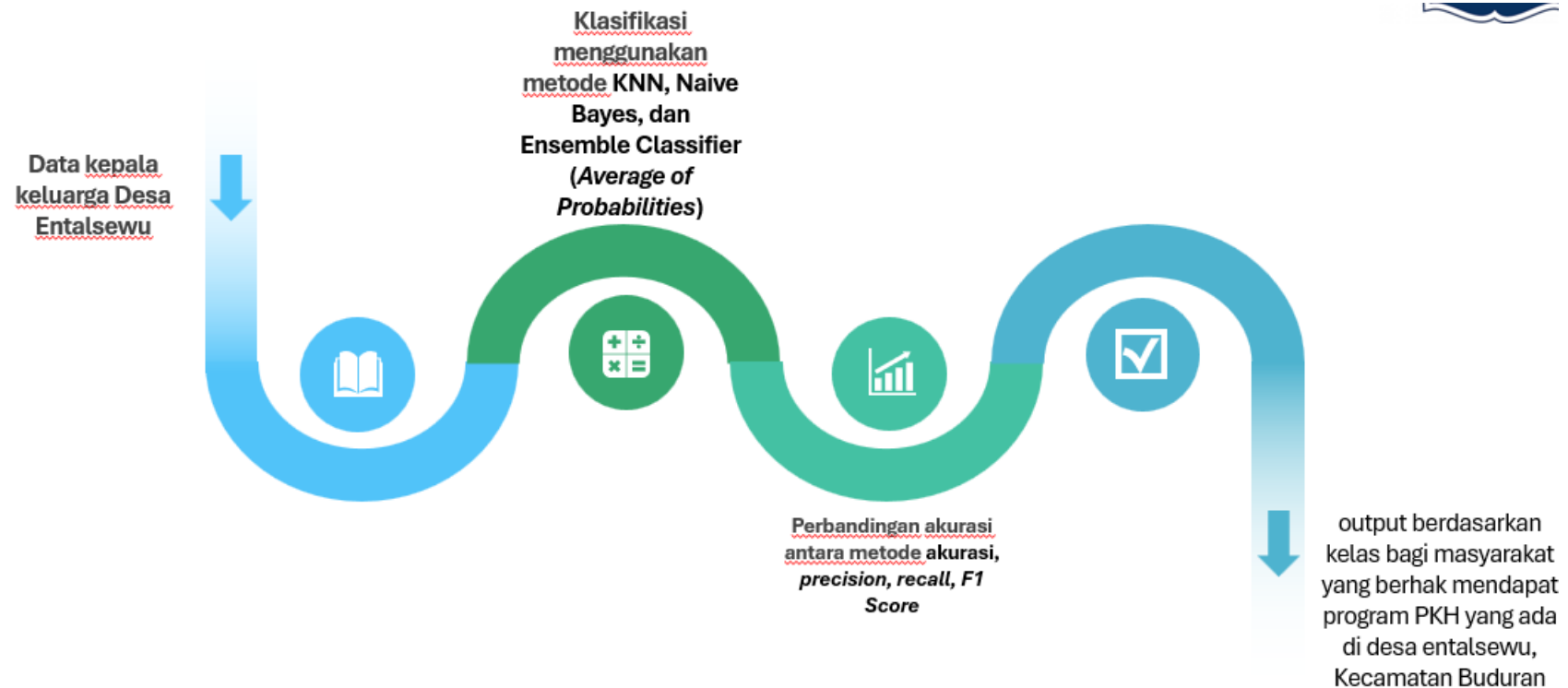
Membangun sebuah model klasifikasi KNN dan Naive Bayes untuk klasifikasi kelayakan PKH dengan studi kasus Desa Entalsewu, Kecamatan Buduran, Kabupaten Sidoarjo.

Mengevaluasi dan membandingkan performa model klasifikasi, dan menentukan model terbaik berdasarkan pengujian klasifikasi dalam kelayakan PKH dengan studi kasus Desa Entalsewu, Kecamatan Buduran, Kabupaten Sidoarjo. Target akurasi yang diharapkan berkisar antara 70% sampai 100%.

Batasan Masalah

1. Data yang digunakan adalah data kepala keluarga di Desa Entalsewu, Kec. Buduran, Kab. Sidoarjo.
2. Pendefinisian label tidak mampu didapatkan dari penetapan Keluarga yang mendapatkan PKH sedangkan untuk label mampu didapatkan dari penetapan Keluarga yang tidak mendapatkan PKH
3. Ada 12 parameter yang digunakan untuk klasifikasi pada penelitian ini.
4. Kelas target hanya 2 kelas (mampu/tidak mampu) sesuai label data pada program PKH.
5. Laporan komparasi performa model dari metode klasifikasi terbaik berdasarkan pengujian model yang dilakukan. Model klasifikasi dan laporan hasil evaluas. Serta file skrip untuk pelatihan dan pengujian menggunakan metode KNN, Naive Bayes, dan Ensemble Classifier (*Average of Probabilities*)

Metode



Metodologi Penelitian

Studi Literatur dan Pengumpulan Data

Pengumpulan data yang berkaitan dengan penelitian. Data masyarakat ental sewu dengan variabel – variabel yang akan dipakai .

Preprocessing

Proses Preprocessing adalah proses data mentah diolah menjadi data yang siap untuk diklasifikasikan terdapat beberapa tahapan.

Pembagian data latih

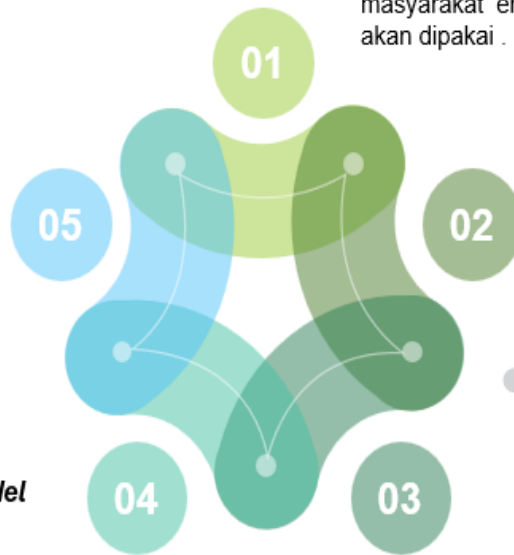
Pembagian data latih (training data) adalah proses krusial dalam klasifikasi. Metode dan rasio pembagiannya tergantung pada ukuran kumpulan data (dataset).

Evaluasi Model (Model Evaluation)

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

Pembuatan dan Pelatihan Model (Model Development and Training)

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 mengabungkan metode KNN dan Naïve Bayes.



Pembagian Data Latih

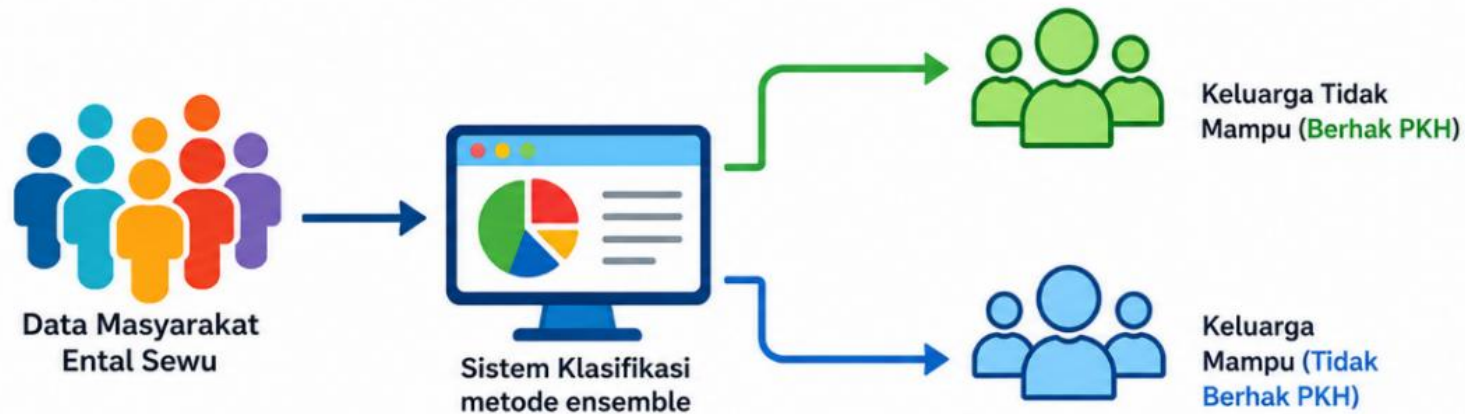
Pembagian data latih (training data) adalah proses krusial dalam klasifikasi. Metode dan rasio pembagiannya tergantung pada ukuran kumpulan data (dataset). Pada penelitian ini akan menggunakan data Train-test split. Teknik Train-test split untuk mengevaluasi atau menguji data yang sebelumnya telah digunakan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

Pelatihan Model

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 $K = 1, 3, 5, 7, 9$.

Pengumpulan Data

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 TANGGUGAN	KONDISI KESEHATAN	JUMLAH KELUARGA BEKERJA	GAJI KELUARGA BEKERJA	KONDISI RUMAH	KEPEMILIKAN RUMAH	PEKERJAN	KELAS
3	SEHAT	2	4.000.000	LAYAK	MILIK SENDIRI	NOTARIS	MAMPU



Preprocessing Data



Cleaning Data



Missing Value



Encoding Categorical



Normalisasi Data
Numerikal (Min Max)

Implementasi Program



SISTEM KLASIFIKASI KELAYAKAN PENERIMA PKH STUDI KASUS (DESA ENTALSEWU, KECAMATAN BUDURAN, KABUPATEN SIDOARJO) METODE KNN, NAIVE BAYES, DAN ENSEMBLE CLASSIFIER (AVERAGE OF PROBABILITIES)

Login

SILAKAN LOGIN DENGAN USERNAME DAN PASSWORD ANDA.

Username:

Password:

LOGIN

© 2026

Form Testing KNN

Jumlah Data Training : 77 Jumlah Data Testing : 20

Pilih Nilai K:

KIRUNG KNN

NO.	NAMA	KELAS	PREDIKSI NAIVE BAYES	NILAI NAIVE BAYES KELAS MAMPU	NILAI NAIVE BAYES KELAS TIDAK MAMPU
1	ANDHIKA PRAMUDITYA	TIDAK MAMPU	TIDAK MAMPU	0.000000720396	0.000019129204
2	FANNY AMRIZAL	MAMPU	MAMPU	0.000195400178	0.000000000000
3	GI LANG DARMAWANGSA	TIDAK MAMPU	TIDAK MAMPU	0.000012783515	0.00003423242
4	SUPRPTO	TIDAK MAMPU	MAMPU	0.000107766398	0.000075117605
5	ARI SUWONDO	TIDAK MAMPU	TIDAK MAMPU	0.000000000000	0.000000038532
6	ITA NOVARIITA	MAMPU	TIDAK MAMPU	0.000020572656	0.000080975966
7	NUNIK	MAMPU	TIDAK MAMPU	0.0000000627145	0.000000030081
8	BOEDI ADIWIDJAJA	MAMPU	MAMPU	0.000994799309	0.000000000000
9	HARI SUBANDI	TIDAK MAMPU	TIDAK MAMPU	0.000001146706	0.000239115572

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 = 1.7119260 Second
Waktu Eksekusi KNN = 39.3355310 Second
Waktu Eksekusi Esemble = 7.3381628 Second

NO.	NAMA	KELAS	PREDIKSI NAIVE BAYES	NILAI NAIVE BAYES KELAS MAMPU	NILAI NAIVE BAYES KELAS TIDAK MAMPU	PREDIKSI KNN	NILAI KNN KELAS MAMPU	NILAI KNN KELAS TIDAK MAMPU	TETANGGA TERDEKAT	PREDIKSI ESEMBLE	NILAI ESEMBLE KELAS MAMPU	NILAI ESEMBLE 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	WIDIAYA 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

Hasil Pengujian

	Rasio	Data Latih	Data Uji	Akurasi(%)		
				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

No.	Rasio	Data Latih	Data Uji	Waktu Eksekusi(Second)		
				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

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

F1 Score

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

Kesimpulan

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

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.

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.

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%.

Referensi

- Aisyah Chairani P, F. E. (Januari 2021). Klasifikasi Rumah Tangga Miskin Menggunakan Metode Naive Bayes (Studi Kasus: Provinsi Papua Tahun 2017). Jurnal Sains Matematika dan Statistika, Vol. 7, No. 1, 89-95.
- BPS/Badan Pusat Statistik dan Depsos/Departemen Sosial (2002), Penduduk Fakir Miskin Indonesia 2002, Jakarta: BPS.
- Badan Pusat Statistik Sidoarjo , 2024. Profil Kemiskinan di Kabupaten Sidoarjo Maret 2024. Sidoarjo Pusat : Badan Pusat Statistik Sidoarjo.
- Chairul Fadlan1, S. N. (Juni 2018). Penerapan Metode Naïve Bayes Dalam Klasifikasi Kelayakan Keluarga Penerima Beras Rastra. Jutim (Jurnal Teknik Infoematika Musiwaras), Vol 3 No.1, 1-8.
- I. K. Yoga, S. S. Prasetyowati, dan Y. Sibaroni, “Prediction And Mapping Rainfall Classification Using Naive Bayes And Simple Kriging,” JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform., vol. 7, no. 4, hal. 1244–1253, 2022.
- Jose, S. (2024). Pendekatan Metode Ensemble Learning untuk Prakiraan Cuaca menggunakan Soft Voting Classifier. Journal Of Applied Computer Science And Technology (JACOST) Vol. 5 No. 1 (2024) : Hal 72 – 80.
- Lestary Fitriany, “Pengaruh Produk Domestik Regional Bruto (PDRB) Per Kapita, Belanja Daerah dan Zakat Terhadap Kemiskinan di Kota Dumai,” Al-Hisbah Jurnal Ekonomi Syariah, vol. 1, no. 1, pp. 12–23, Dec. 2021, doi: 10.57113/his.v1i1.75.
- Muluk, F. A. (2024). Efektivitas Program Keluarga Harapan (Pkh) Dalam Mengurangi Garis Kemiskinan Kecamatan Sidoarjo, Kabupaten Sidoarjo. *JURNAL MEDIA AKADEMIK (JMA)*, Vol.2, No.7Juli 2024.
- Peraturan Menteri Sosial Republik Indonesia Nomor 1 Tahun 2018 Tentang Program Keluarga Harapan. Sumber : BN.2018/NO.187, jdih.kemsos.go.id : 35 hlm.
- Pristyanto, Y. (2019). Penerapan Metode Ensemble Untuk Meningkatkan Kinerja Algoritme Klasifikasi Pada Imbalanced Dataset. Jurnal TEKNOINFO, Vol. 13, No. 1, 11-16.
- Riyan Latifahul Hasanah, M. H. (Maret 2019). Klasifikasi Penerima Dana Bantuan Desa Menggunakan Metode Knn (K-Nearest Neighbor). Jurnal TECHNO Nusa Mandiri Maret 2019, Vol. 16, No. 1, 1-6.
- Rohma, N. (2022). Perbandingan Algoritma Klasifikasi Data Kesejahteraan Sosial Kabupaten Bantul. Jurnal PROCESSOR –VOL. 17, No. 2, Oktober 2022.
- Utami dkk. (2022). Klasifikasi Kelayakan Penerima Bantuan Program Keluarga Harapan (Pkh) Menggunakan Metode Weighted Naïve Bayes Dengan Laplace Smoothing. JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika) Vol. 07, Nomor 04, Desember 2022 : Hal 1373 - 1384.

