

STUDI KOMPARASI ALGORITMA NAIVE BAYES DAN K-NEAREST NEIGHBOR UNTUK KLASIFIKASI KELUARGA MISKIN DI DESA NGADIWARNO KECAMATAN SUKOREJO KABUPATEN KENDAL

¹⁾Wahyu Aji Bayu Pemungkas, ²⁾Muhamad Fuat Asnawi

^{1,2)}Universitas Sains Al-Qur'an

¹⁾ wabp100799@gmail.com, ²⁾ fuatasnawi@unsiq.ac.id

INFO ARTIKEL

Riwayat Artikel :

Diterima : 20 Mei 2024

Disetujui : 31 Mei 2024

Kata Kunci :

Naive Bayes, KNN, Klasifikasi, Keluarga Miskin

ABSTRAK

Penelitian ini bertujuan untuk mengevaluasi dan membandingkan kinerja algoritma Naive Bayes dan K-Nearest Neighbor (KNN) dalam mengklasifikasikan keluarga miskin di Desa Ngadiwarno, Kecamatan Sukorejo, Kabupaten Kendal. Data yang digunakan mencakup 675 Kepala Keluarga (KK) yang terdaftar dalam Data Terpadu Kesejahteraan Sosial (DTKS) 2023. Data ini diolah menggunakan perangkat lunak RapidMiner Studio, dengan evaluasi berdasarkan akurasi, precision, recall, dan Area Under Curve (AUC). Hasil penelitian menunjukkan bahwa algoritma Naive Bayes memiliki akurasi sebesar 99,26%, precision 100%, recall 95,41%, dan AUC 0,978, sedangkan KNN memiliki akurasi 98,52%, precision 100%, recall 90,82%, dan AUC 0,993. Meskipun KNN memiliki nilai AUC sedikit lebih tinggi, Naive Bayes menunjukkan performa keseluruhan yang lebih unggul dalam klasifikasi keluarga miskin. Penelitian ini menyimpulkan bahwa Naive Bayes lebih efektif untuk digunakan dalam konteks data yang ada, dan diharapkan dapat membantu pemerintah daerah dalam menyusun kebijakan penanggulangan kemiskinan yang lebih tepat sasaran dan efektif

ARTICLE INFO

Article History :

Received : May 20, 2024

Accepted : May 31, 2024

Keywords:

Naive Bayes, KNN, Classification, Poor Households

ABSTRACT

This study aims to evaluate and compare the performance of the Naive Bayes and K-Nearest Neighbor (KNN) algorithms in classifying poor households in Ngadiwarno Village, Sukorejo District, Kendal Regency. The data used includes 675 households registered in the Integrated Social Welfare Data (DTKS) 2023. This data was processed using RapidMiner Studio software, with evaluation based on accuracy, precision, recall, and Area Under Curve (AUC). The results show that the Naive Bayes algorithm has an accuracy of 99.26%, precision of 100%, recall of 95.41%, and AUC of 0.978, while KNN has an accuracy of 98.52%, precision of 100%, recall of 90.82%, and AUC of 0.993. Although KNN has a slightly higher AUC, Naive Bayes shows superior overall performance in classifying poor households. This study concludes that Naive Bayes is more effective for the given data context and is expected to assist local governments in formulating more accurate and effective poverty alleviation policies.

1. PENDAHULUAN

Perkembangan teknologi yang semakin pesat, khususnya dalam bidang teknik informatika, memberikan berbagai kemajuan dalam analisis data yang besar dan kompleks melalui disiplin ilmu data mining. Data mining menggabungkan konsep-konsep dari komputer dan statistika untuk mengeksplorasi dan menganalisis pola-pola tersembunyi dalam kumpulan data besar dengan tujuan utama untuk mengidentifikasi hubungan atau tren yang memberikan wawasan berharga dalam pengambilan keputusan. Teknik ini telah banyak digunakan dalam berbagai aplikasi seperti pengembangan sistem rekomendasi, deteksi anomali, dan prediksi pola perilaku konsumen.

Kemiskinan, sebagai fenomena yang kompleks, tidak hanya terbatas pada kekurangan finansial dan pendapatan yang minim, tetapi juga meliputi aspek-aspek lain seperti kesehatan yang rendah, tingkat pendidikan yang kurang memadai, ketidakadilan dalam sistem hukum, risiko terhadap kejahatan, dan keterbatasan dalam mengendalikan arah kehidupan. Faktor-faktor ekonomi seperti pendapatan rendah, keterbatasan lapangan kerja, pertumbuhan ekonomi yang lambat, dan tingkat pendidikan yang rendah juga turut berperan dalam memperparah kondisi kemiskinan (Zurisdah, 2017).

Desa Ngadiwarno di Kecamatan Sukorejo, Kabupaten Kendal, merupakan salah satu wilayah di Indonesia yang mengalami masalah kemiskinan. Menurut data Badan Pusat Statistik (BPS) Jawa Tengah, tingkat kemiskinan di Kendal meningkat dalam tiga tahun terakhir, dengan angka kemiskinan pada tahun 2020 sebesar 9,99%, tahun 2021 sebesar 10,24%, dan tahun 2022 sebesar 10,93%. Desa Ngadiwarno sendiri memiliki data kemiskinan yang mencapai 1577 jiwa dari total penduduk sebanyak 4.601 jiwa, dengan mayoritas penduduknya bekerja sebagai petani.

Dalam upaya mengurangi kemiskinan, pemerintah telah meluncurkan berbagai program penanggulangan kemiskinan. Namun, tantangan yang dihadapi adalah ketidakakuratan dalam pendataan keluarga miskin, yang sering kali mengakibatkan bantuan tidak tepat sasaran. Oleh karena itu, diperlukan penelitian yang dapat membantu pemerintah mengidentifikasi dan

mengklasifikasikan rumah tangga miskin dengan lebih tepat.

Penelitian ini memfokuskan pada penggunaan algoritma Naive Bayes dan K-Nearest Neighbor untuk mengklasifikasikan keluarga miskin di Desa Ngadiwarno. Algoritma Naive Bayes memanfaatkan probabilitas untuk memberikan prediksi berdasarkan independensi fitur, sementara K-Nearest Neighbor mengandalkan jarak antar data untuk menentukan kelasnya. Dengan penerapan teknik ini, diharapkan dapat diperoleh informasi yang lebih akurat tentang karakteristik keluarga miskin, sehingga kebijakan program pemerintah daerah dapat disusun dengan lebih tepat sasaran dan efektif dalam mengurangi angka kemiskinan (Santosa, 2022).

Beberapa penelitian sebelumnya telah membandingkan algoritma Naive Bayes dan K-Nearest Neighbor dalam berbagai konteks, menunjukkan bahwa performa kedua algoritma dapat bervariasi tergantung pada jenis data dan aplikasi yang digunakan. Maulidya Dwi Nurmalasari et al. (2021) dalam jurnal Komtika membandingkan akurasi kedua algoritma untuk mendiagnosis penyakit diabetes, dengan hasil menunjukkan bahwa K-Nearest Neighbor memiliki akurasi sebesar 91,61%, lebih tinggi dibandingkan Naive Bayes yang memiliki akurasi sebesar 85,60%. Bayu Delvika et al. (2022) dalam jurnal Malcom juga menemukan bahwa Naive Bayes lebih baik dalam mengklasifikasikan risiko diabetes pada ibu hamil, dengan akurasi sebesar 75,78%, dibandingkan K-Nearest Neighbor yang memiliki akurasi sebesar 74,48%. Sedangkan, Eka Wahyu Sholeha et al. (2022) dalam jurnal JTIM melaporkan bahwa K-Nearest Neighbor memiliki akurasi yang sedikit lebih tinggi dibandingkan Naive Bayes dalam analisis sentimen pada agen perjalanan online.

Penelitian ini bertujuan untuk mengeksplorasi dan membandingkan hasil akurasi dari algoritma Naive Bayes dan K-Nearest Neighbor dalam mengklasifikasikan keluarga miskin di Desa Ngadiwarno, Kecamatan Sukorejo, Kabupaten Kendal. Dengan demikian, diharapkan penelitian ini dapat memberikan solusi yang lebih baik dalam penanggulangan kemiskinan di wilayah tersebut.

2. METODE

Penelitian ini merupakan jenis eksperimen yang bertujuan untuk menilai tingkat akurasi yang dihasilkan oleh algoritma Naive Bayes dan K-Nearest Neighbor dalam mengklasifikasikan keluarga miskin di Desa Ngadiwarno, Kecamatan Sukorejo. Data yang digunakan adalah data primer dari hasil wawancara dengan perangkat desa dan data sekunder dari buku dan jurnal yang relevan. Sampel penelitian terdiri dari 675 warga yang terdaftar dalam Data Terpadu Kesejahteraan Sosial (DTKS) 2023. Teknik pengumpulan data meliputi observasi, studi pustaka, dan wawancara. Data kemudian diolah menggunakan perangkat lunak RapidMiner Studio dan dibandingkan hasilnya antara kedua algoritma tersebut.

Alur proses penelitian dimulai dari identifikasi masalah, pengumpulan data, pengolahan data, perbandingan hasil pengolahan, hingga penyusunan laporan. Penelitian ini menggunakan perangkat keras seperti prosesor Intel Core i3 dan perangkat lunak seperti RapidMiner untuk analisis data. Dengan demikian, penelitian ini diharapkan dapat memberikan solusi yang lebih akurat dalam penanggulangan kemiskinan di Desa Ngadiwarno melalui penggunaan teknologi data mining dan algoritma klasifikasi yang tepat.

3. HASIL DAN PEMBAHASAN

Berikut ini adalah data yang digunakan:

Tabel 1. Data

NO	Nama	Jenis Kelamin	Umur	Tanggungan	Pekerjaan	Penghasilan	BPNP	PKH	PBIJKN	Kelas Asli
1	Jumiko	L	Tua	3-4	Petani	Sedang	Tidak	Tidak	Ya	Miskin
2	Darwan	L	Tua	0-2	Buruh	Rendah	Tidak	Tidak	Ya	Miskin
3	Siti	P	T	0	B	R	T	T	Y	Mis

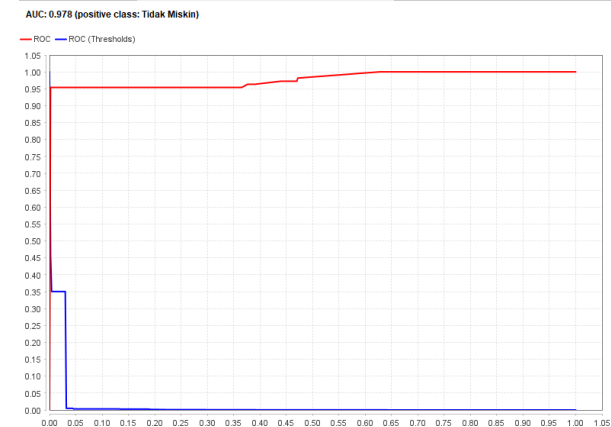
	musri fah		ua	-2	uruh	enda h	id ak	id ak	a kin	
4	Tatik rohman	P	Muda	0-2	Buruh	Rendah	Tidak	Tidak	Ya	Miskin
NO	Nama	JK	Umur	Tanggungan	Pekerjaan	Penghasilan	BPNP	PKH	PBIJKN	Kelas Asli
5	Ngati yem	P	Tua	0-2	Petani	Sedang	Tidak	Tidak	Ya	Miskin
6	Slamet mustaghfirin	L	Tua	3-4	Petani	Sedang	Tidak	Tidak	Ya	Miskin
7	Tubi	L	Tua	0-2	Petani	Sedang	Tidak	Tidak	Ya	Miskin
8	Chaerun	L	Tua	3-4	Petani	Sedang	Tidak	Tidak	Ya	Miskin
9	Kaswadi	L	Tua	0-2	Petani	Sedang	Tidak	Tidak	Tidak	Miskin
10	Lieta winarni	P	Tua	0-2	Pedagang	Sedang	Tidak	Tidak	Ya	Miskin
11	Rochimah	P	Tua	0-2	Petani	Sedang	Ya	Ya	Ya	Miskin
12	Ahmad imron	L	Tua	3-4	Petani	Sedang	tidak	Tidak	Ya	Miskin
13	Ahmad	L	Tu	0-	Pet	S	ti	Ti	Ya	Mis

	basuki		a	2	a	d	a	d		
14	Misri	L	Tua	3-4	Petani	Sedang	tidak	Tidak	Ya	Miskin
15	Zurqoni	L	Tua	0-2	Petani	Sedang	tidak	Tidak	Ya	Miskin
16	Akhmad Nurhadi	L	Tua	3-4	Buruh	Rendah	tidak	Tidak	Ya	Miskin
17	Suisman	L	Tua	0-2	Buruh	Rendah	tidak	Tidak	Ya	Miskin
18	Sumari	L	Tua	0-2	Petani	Sedang	tidak	Tidak	Ya	Miskin
19	Bejo	L	Tua	0-2	Buruh	Rendah	tidak	Tidak	Ya	Miskin
20	Sudarno	L	Tua	0-2	Buruh	Rendah	tidak	Tidak	Ya	Miskin
...
675	Surip Turiah	P	Tua	0-2	Buruh	Rendah	Tidak	Tidak	Ya	Miskin

Penelitian ini menggunakan algoritma Naive Bayes dan K-Nearest Neighbor (KNN) untuk mengklasifikasikan data keluarga miskin di Desa Ngadiwarno. Data yang digunakan mencakup 675 Kepala Keluarga (KK) yang terdaftar dalam Data Terpadu Kesejahteraan Sosial (DTKS) 2023 dan 870 KK yang tidak terdaftar. Data ini diolah menggunakan perangkat lunak RapidMiner Studio, dengan masing-masing algoritma diuji untuk menentukan akurasi, precision, recall, dan Area Under Curve (AUC) sebagai parameter evaluasi kinerja.

accuracy: 99.26%			
	true Miskin	true Tidak Miskin	class precision
pred. Miskin	566	5	99.12%
pred. Tidak Miskin	0	104	100.00%
class recall	100.00%	95.41%	

Gambar 1. Nilai Akurasi Naïve Bayes

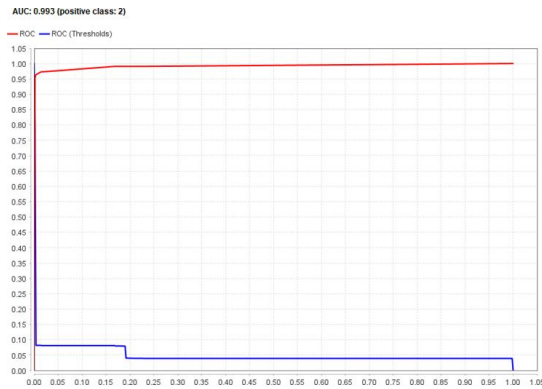


Gambar 2. Nilai AUC Naïve Bayes

Pada tahap pengujian, algoritma Naive Bayes menunjukkan hasil yang sangat baik dengan akurasi sebesar 99,26%. Selain itu, algoritma ini juga memiliki nilai precision 100%, yang berarti bahwa semua keluarga yang diprediksi sebagai miskin benar-benar miskin. Nilai recall yang mencapai 95,41% menunjukkan bahwa sebagian besar keluarga miskin berhasil diidentifikasi oleh algoritma ini. Nilai AUC sebesar 0,978 juga mengindikasikan bahwa algoritma Naive Bayes memiliki kemampuan diskriminasi yang sangat baik antara keluarga miskin dan tidak miskin. Sebanyak 566 data dari total sampel berhasil diprediksi dengan benar sebagai keluarga miskin oleh algoritma Naive Bayes, sementara hanya 5 data yang salah diprediksi sebagai tidak miskin. Hasil ini menunjukkan tingkat kesalahan yang sangat rendah dan efisiensi tinggi dari algoritma Naive Bayes dalam menangani data klasifikasi keluarga miskin di Desa Ngadiwarno.

accuracy: 98.96%			
	true 1	true 2	class precision
pred. 1	564	5	99.12%
pred. 2	2	104	98.11%
class recall	99.65%	95.41%	

Gambar 3. Nilai Akurasi KNN



Gambar 4. Nilai AUC KNN

Sebaliknya, algoritma KNN dengan nilai $K=3$ juga menunjukkan hasil yang cukup baik, meskipun sedikit lebih rendah dibandingkan Naive Bayes. Algoritma KNN mencapai akurasi sebesar 98,52%, dengan precision 100%. Ini menunjukkan bahwa algoritma KNN juga memiliki kemampuan yang sangat baik dalam mengidentifikasi keluarga miskin dengan benar. Namun, recall sebesar 90,82% menunjukkan bahwa ada beberapa keluarga miskin yang tidak berhasil diidentifikasi oleh algoritma ini.

Nilai AUC untuk algoritma KNN adalah 0,993, sedikit lebih tinggi dibandingkan dengan Naive Bayes. Hal ini menunjukkan bahwa KNN memiliki kemampuan yang sangat baik dalam membedakan antara keluarga miskin dan tidak miskin, meskipun akurasi keseluruhan sedikit lebih rendah. Sebanyak 564 data dari total sampel berhasil diprediksi dengan benar sebagai keluarga miskin, sementara 2 data salah diprediksi sebagai tidak miskin.

Berdasarkan hasil perbandingan antara kedua algoritma tersebut, dapat disimpulkan bahwa algoritma Naive Bayes lebih unggul dalam klasifikasi keluarga miskin di Desa Ngadiwarno dibandingkan dengan algoritma KNN. Hal ini terutama terlihat dari nilai akurasi dan recall yang lebih tinggi, meskipun AUC KNN sedikit lebih baik. Naive Bayes mampu memberikan prediksi yang lebih akurat dan konsisten dalam konteks penelitian ini.

Perbandingan ini menunjukkan bahwa kedua algoritma memiliki kelebihan masing-masing, namun dalam konteks data yang digunakan, Naive Bayes lebih efektif untuk klasifikasi keluarga miskin di Desa Ngadiwarno. Faktor-faktor seperti independensi fitur dan jumlah sampel yang digunakan mungkin berkontribusi pada hasil ini, di mana Naive Bayes

memanfaatkan probabilitas fitur dengan lebih baik.

4. PENUTUP

4.1. Kesimpulan

Dapat disimpulkan bahwa algoritma Naive Bayes lebih efektif dan akurat dalam mengklasifikasikan keluarga miskin di Desa Ngadiwarno dibandingkan dengan KNN. Penggunaan algoritma Naive Bayes diharapkan dapat membantu pemerintah daerah dalam menyusun kebijakan program penanggulangan kemiskinan yang lebih tepat sasaran dan efektif. Penelitian ini menunjukkan pentingnya pemilihan algoritma yang tepat dalam analisis data mining untuk mengatasi masalah sosial seperti kemiskinan.

4.2. Saran

Untuk penelitian selanjutnya, disarankan untuk mengeksplorasi penggunaan algoritma lain seperti Random Forest atau Support Vector Machine (SVM) dalam klasifikasi keluarga miskin guna membandingkan kinerja dengan algoritma Naive Bayes dan K-Nearest Neighbor. Selain itu, mempertimbangkan faktor-faktor tambahan yang mungkin mempengaruhi tingkat kemiskinan, seperti akses terhadap layanan kesehatan, pendidikan, dan infrastruktur, dapat memberikan gambaran yang lebih komprehensif. Penelitian juga dapat diperluas dengan menggunakan dataset yang lebih besar dan beragam dari berbagai daerah untuk meningkatkan generalisasi hasil. Implementasi metode validasi silang yang lebih kompleks dan analisis sensitivitas terhadap parameter algoritma juga dapat membantu dalam memahami lebih baik performa masing-masing algoritma dalam berbagai kondisi data.

5. DAFTAR PUSTAKA

Bayu Delvika, Syahida Nurhidayarnis, Puji Dwi Rinada, Naufal Abror, & Assad Hidayat. (2022). Perbandingan Klasifikasi Antara Naive Bayes dan K-Nearest Neighbor Terhadap Resiko Diabetes Pada Ibu Hamil. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 2(2), 68-75.
<https://journal.irpi.or.id/index.php/malcom/article/view/432>

- Maulidya Dwi Nurmalasari, Kusri, & Sudarmawan. (2021). Komparasi Algoritma Naïve Bayes dan K-Nearest Neighbor untuk Membangun Pengetahuan Diagnosa Penyakit Diabetes. *Jurnal Komtika (Komputasi dan Informatika)*, 5(1), 52-57. <https://doi.org/10.31603/komtika.v5i1.5140>
- Santoso, A., Utami, E., & Hartanto, A. D. (2022). Perbandingan Algoritma Naïve Bayes dan K-Nearest Neighbor dalam Menentukan Kriteria Masyarakat Miskin. *Jurnal Informa: Jurnal Penelitian dan Pengabdian Masyarakat*, 8(1), 19-23.
- Sholeha, E. W., Yunita, S., Hammad, R., Hardita, V. C., & Kaharuddin. (2022). Analisis Sentimen Pada Agen Perjalanan Online Menggunakan Naïve Bayes dan K-Nearest Neighbor. *JTIM: Jurnal Teknologi Informasi dan Multimedia*, 3(4), 203-208.
- Wang, X., Xu, X., & Zhao, Y. (2021). Runoff and Sediment Deposition Characteristics of Gravel-Mulched Land: An Experimental Study. *Land*, 10(2), 445. <https://doi.org/10.3390/land10020445>.
- Zurisdah, Z. (2017). Pengaruh Tingkat Pengangguran terbuka terhadap Tingkat Kemiskinan di provinsi Banten (Doctoral dissertation, Universitas Islam Negeri Sultan Maulana Hasanuddin BANTEN).