KESATRIA: Jurnal Penerapan Sistem Informasi (Komputer & Manajemen)
Terakreditasi Nomor 204/E/KPT/2022 | Vol. 6, No. 2, April (2025), pp. 461-467

Klasifikasi Tingkat Kemiskinan di Indonesia Menggunakan Metode *Naïve Bayes*

Fathoni[1], Zahwa Aulia Prayetno[2], Michael Joenathan Darwin[3], Liza Athalya Nurjannah[4]
[1]Fakultas Ilmu Komputer, Sistem Informasi, Universitas Sriwijaya, Palembang, Indonesia
E-mail: [1,]fathoni@unsri.ac.id, [2]zahwaauliaprayetno@gmail.com, [3]michaeljdarwin9@gmail.com, [4]liza.athalya@gmail.com*

Abstract

The purpose of this study is to design a poverty level classification model in Indonesia using this Naïve Bayes algorithm, with a specific focus on identifying key socio-economic variables that significantly influence poverty status. The development of this model is aimed at laying the groundwork for supporting policy decisions in the distribution of social assistance, ensuring that aid is more accurately targeted toward communities truly in need. The dataset employed in this research includes various relevant socio-economic indicators, and the model's performance is validated through a cross-validation method to assess its accuracy and generalization capability. The outcomes reveal that the model obtained an accuracy of 92.40% ± 2.99%, with high precision for the majority class. However, the model still shows limitations in classifying the minority class or vulnerable groups. Overall, the findings suggest that the Naïve Bayes algorithm holds strong potential for poverty classification and can be utilized as a decision-support tool in the formulation of data-driven social policies, particularly in optimizing the fair and effective distribution of social aid.

Keywords: *Naïve Bayes, Poverty, Classification, Indonesia, Cross-validation*

Abstrak

Penelitian ini bertujuan merancang model klasifikasi tingkat kemiskinan di Indonesia dengan memanfaatkan algoritma Naïve Bayes, serta mengidentifikasi variabel-variabel sosial ekonomi yang berperan signifikan dalam menentukan status kemiskinan. Pengembangan model ini ditujukan sebagai dasar dalam mendukung pengambilan kebijakan penyaluran bantuan sosial secara lebih tepat sasaran, guna memastikan bantuan diterima oleh kelompok masyarakat yang benar-benar membutuhkan. Data yang digunakan dalam penelitian ini mencakup sejumlah indikator sosial ekonomi yang relevan, dengan proses validasi memanfaatkan metode cross-validation dalam menguji tingkat akurasi dan keterampilan generalisasi model. Hasil pengujian menunjukkan bahwa model menghasilkan akurasi sebesar 92,40% ± 2,99%, dengan tingkat presisi yang tinggi pada kelas mayoritas. Meski demikian, model masih menunjukkan keterbatasan dalam mengklasifikasikan kelas minoritas atau kelompok rentan. Secara keseluruhan, hasil penelitian menunjukkan algoritma Naïve Bayes mempunyai potensi yang baik untuk mengklasifikasikan tingkat kemiskinan dan dapat dimanfaatkan sebagai alat bantu dalam penyusunan kebijakan sosial berbasis data, khususnya dalam rangka optimalisasi distribusi bantuan sosial yang lebih adil dan efektif.

Kata Kunci: *Naïve Bayes, Kemiskinan, Klasifikasi, Indonesia, Cross-validation*

1. Pendahuluan

Kemiskinan di Indonesia yaitu permasalahan kompleks yang berdampak signifikan terhadap kesejahteraan masyarakat, pertumbuhan ekonomi, dan stabilitas sosial.

ISSN: 2720-992X
Copyright © KESATRIA: Jurnal Penerapan Sistem Informasi (Komputer & Manajemen)

Meskipun pemerintah telah meluncurkan berbagai program penanggulangan kemiskinan, seperti bantuan sosial, subsidi, dan dana desa, tantangan dalam mengatasi kemiskinan masih tetap besar. Hal tersebut dikarenakan beberapa faktor, berupa ketidaktepatan sasaran program, ketimpangan pendapatan, serta keterbatasan jangkauan pada pendidikan dan layanan kesehatan bermutu [1]. Berbagai penelitian telah dilakukan untuk mendukung efektivitas program penanggulangan kemiskinan, salah satunya melalui penerapan metode klasifikasi dalam pengolahan data kemiskinan. Penelitian yang dilakukan oleh Farhan Afrian menunjukkan bahwa algoritma Decision Tree dapat digunakan untuk mengelompokkan tingkat kemiskinan di Indonesia berdasarkan sejumlah variabel sosial ekonomi, namun masih menghadapi keterbatasan pada akurasi model akibat ketidakseimbangan data serta kesulitan dalam validasi hasil [2]. Studi lain yang menggunakan algoritma Decision Tree juga menemukan bahwa meskipun akurasi model tinggi, pemilihan atribut yang kurang tepat dapat menurunkan kinerja klasifikasi [3]. Di sisi lain, metode Naïve Bayes juga telah diterapkan pada data kemiskinan dan menunjukkan performa yang baik, tetapi studi tersebut mencatat kurangnya variasi variabel input dan belum dilakukannya validasi silang yang memadai [4]

Berbagai penelitian telah menerapkan metode klasifikasi untuk mengidentifikasi status kemiskinan guna mendukung kebijakan penyaluran bantuan sosial yang lebih tepat sasaran. Dalam hal tersebut, algoritma Naïve Bayes muncul sebagai pilihan utama karena kesederhanaannya dalam mengolah data dan efisiensinya dalam memberikan hasil klasifikasi yang cukup baik [5]. Beberapa studi sebelumnya telah menerapkan Naïve Bayes untuk mengklasifikasikan status kesejahteraan dan kemiskinan di berbagai daerah. Desa Gunungsari berhasil menerapkan Naïve Bayes dalam mengelompokkan kesejahteraan masyarakatnya, meskipun penelitian tersebut mencatat tantangan dalam diversifikasi variabel yang digunakan [4]. Metode yang sama diterapkan pada studi kasus di Desa Karangasem dengan mendapatkan tingkat akurasi yang tinggi, walaupun masih terdapat kekurangan dalam hal validasi model dan penanganan data tidak seimbang [3]. Penelitian di Provinsi Papua juga menegaskan bahwa pemilihan variabel yang representatif sangat penting untuk meningkatkan sensitivitas model Naïve Bayes dalam mendeteksi kelompok masyarakat yang rentan [6]. Faktor lain seperti tingkat pendidikan, kondisi tempat tinggal, akses terhadap pelayanan kesehatan, dan lapangan pekerjaan menjadi hal penting yang harus dipertimbangkan dalam pemodelan agar hasil klasifikasi lebih mencerminkan realitas sosial dan ekonomi masyarakat setempat.

Penelitian ini berfokus pada penerapan algoritma *Naïve Bayes* dalam mengklasifikasikan status kemiskinan, dengan tambahan metode validasi *cross-validation*. Pendekatan *cross-validation* digunakan dalam menjamin model didapati mempunyai kekuatan konsisten juga tidak hanya bergantung pada pembagian data tertentu. Dengan menggunakan metode ini, penelitian diharapkan dapat mengidentifikasi variabel apa saja yang memiliki pengaruh tinggi dalam klasifikasi kemiskinan serta menghasilkan model yang efektif dan akurat. Hasil dari model ini nantinya diharapkan dapat memberikan rekomendasi yang lebih tepat untuk mendukung perumusan kebijakan penyaluran bantuan sosial yang adil dan efisien, serta memberikan gambaran yang lebih mendalam tentang dinamika kemiskinan di Indonesia.

2. Metodologi Penelitian

2.1. Pengumpulan Data

Penelitian ini memanfaatkan data yang diperoleh dari *kaggle*. Data dilakukan analisis lagi menggunakan RapidMiner dengan beberapa metode yaitu *data selection, data pre-processing, data transformation, data mining,* dan *evaluation*.

2.2. *Data Selection*

Data Selection proses memilih subset data yang sejalan pada kumpulan data lebih besar guna analisis lebih lanjut. Tujuan utama dari data selection adalah memastikan bahwa data yang digunakan dalam analisis sesuai dengan tujuan penelitian atau aplikasi tertentu, sehingga meningkatkan efisiensi dan akurasi hasil yang diperoleh. Proses otomatisasi seleksi data digunakan dalam misi luar angkasa untuk memastikan hanya data *burst* yang paling informatif yang diproses lebih lanjut [7].

2.3. *Data Pre-Processing*

Data preprocessing yaitu langkah awal proses *data mining* berfungsi dalam merancang data mentah supaya siap dianalisis secara efektif. Data yang dikumpulkan dari berbagai sumber umumnya memiliki banyak permasalahan, seperti nilai yang hilang (*missing values*), data duplikat, kesalahan entri, skala data tidak konsisten, atau bahkan data tidak sesuai. Oleh karena itu, sebelum proses *mining* dilakukan, data harus melalui tahap preprocessing agar hasil analisis menjadi akurat dan bermakna. Pengelolaan data yang baik termasuk proses preprocessing adalah dasar dari manajemen kualitas data yang dapat mendukung inovasi dan pengambilan keputusan berbasis data [8].

2.4. *Data Transformation*

Data transformation mencakup konversi data mentah jadi format yang selaras dalam analisis dan pemodelan, serta memastikan data bebas dari kesalahan dan inkonsistensi [8]. *Data transformation* dalam konteks *data mining* adalah proses mengubah data mentah ke format atau struktur yang selaras dalam analisis lebih lanjut. Langkah ini merupakan bagian penting dari tahap *data preprocessing*, yang bertujuan untuk meningkatkan kualitas data sehingga algoritma *data mining* dapat bekerja secara optimal.

2.5. *Data Mining*

Data mining merupakan tahap utama proses *Knowledge Discovery in Databases (KDD)*, berupa seleksi data, pembersihan data, transformasi data, penambangan pola, evaluasi pola, presentasi hasil. Proses ini dapat digunakan dalam berbagai bidang seperti pemasaran, keuangan, kesehatan, manufaktur, dan ilmu sosial [5]. *Data mining* adalah proses mengekstraksi informasi atau pola yang berguna pada kumpulan data besar, memanfaatkan teknik statistik, matematika, kecerdasan buatan, dan pembelajaran mesin. Tujuan penting *data mining* yaitu mendapati pengetahuan tersembunyi yang tidak langsung terlihat dari data mentah yang tersedia, serta membantu dalam pengumpulan ketetapan berlandaskan data (*data-driven decision making*).

2.6. *Evaluation*

Evaluation atau evaluasi adalah tahap penting yang dilakukan untuk menilai kualitas dan kegunaan pola atau model yang dihasilkan dari proses *data mining*. *Evaluation* dilakukan setelah model dibangun, untuk memastikan bahwa hasil yang diperoleh tidak hanya secara *statistik valid*, tetapi juga relevan dan bermanfaat untuk tujuan yang diinginkan.

3. Hasil dan Pembahasan

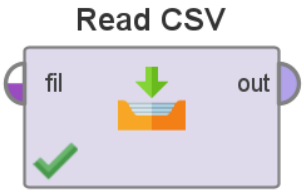
3.1. Pengumpulan Data

Tabel 1. Sample *Dataset* Persentase Penduduk Miskin (P0) Merujuk Provinsi dan Daerah 2024

Provinsi	Kab/Kota	Persentase Penduduk Miskin (P0) Menurut kabupaten/kota	…	PDRB atas Dasar Harga Konstan menurut Pengeluaran (Rupiah)	Klasifikasi Kemiskinan
1648096	Simeulue	18,98	…	1648096	0

Provinsi	Kab/Kota	Persentase Penduduk Miskin (P0) Menurut kabupaten/kota	…	PDRB atas Dasar Harga Konstan menurut Pengeluaran (Rupiah)	Klasifikasi Kemiskinan
1780419	Aceh Singkil	20,36	…	1780419	1
4345784	Aceh Selatan	13,18	…	4345784	0
…	…	…	…	…	…
PAPUA	Intan Jaya	41,66	…	767101	1
PAPUA	Deiyai	40,59	…	841296	1
PAPUA	Jayapura	11,39	…	22852202	0

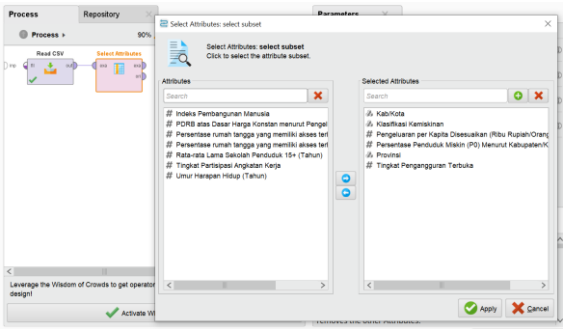
Tabel 1. menyajikan sampel data mengenai persentase penduduk miskin (P0) berdasarkan provinsi serta daerah pada tahun 2024. Data mencakup beberapa wilayah dari berbagai provinsi di Indonesia, termasuk Aceh dan Papua. Setiap baris dalam tabel merepresentasikan satu daerah (kabupaten/kota) dengan informasi mengenai tingkat kemiskinan yang dikategorikan ke dalam klasifikasi tertentu.



Gambar 1. *Read* CSV

3.2. *Data Selection*

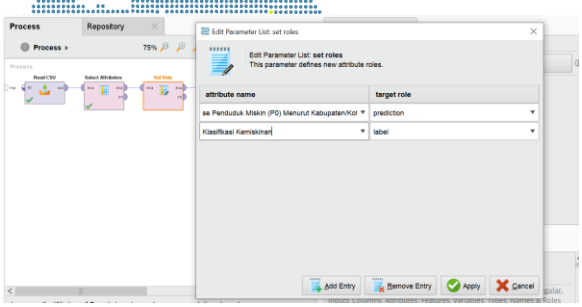
Tahapan ini merupakan proses pemilihan data dengan tujuan untuk memilih atribut-atribut yang tidak penting agar proses analisis pada RapidMiner dapat dilakukan dengan lebih efisien. Dalam tahap seleksi atribut ini, hanya beberapa variabel yang dipilih untuk dianalisis lebih lanjut, yaitu Kode Kota, Klasifikasi Kemiskinan, Pengeluaran per Kapita Disesuaikan, Persentase Penduduk Miskin, dan Tingkat Pengangguran Terbuka. Dengan melakukan *Data Selection* ini, data yang dianalisis akan lebih fokus dan relevan terhadap tujuan penelitian, sehingga dapat menghasilkan wawasan yang lebih akurat dalam tahap analisis berikutnya.



Gambar 2. *Data Selection*

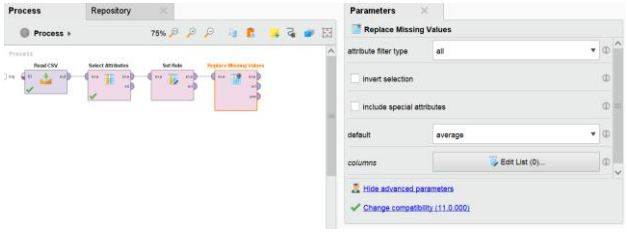
3.3. *Data Pre-Processing*

Tahapan ini, dilakukan *Set Role* untuk mendukung proses analisis di RapidMiner. Atribut "Persentase Penduduk Miskin (P0) Menurut Kabupaten/Kota" ditetapkan sebagai *prediction*, yang berarti variabel ini akan diprediksi. Sementara itu, "Klasifikasi Kemiskinan" ditetapkan sebagai label, yang berfungsi sebagai target klasifikasi. Langkah ini memastikan model dapat memahami struktur data dengan benar sebelum analisis lebih lanjut dilakukan.



Gambar 3. *Data Pre-Processing*

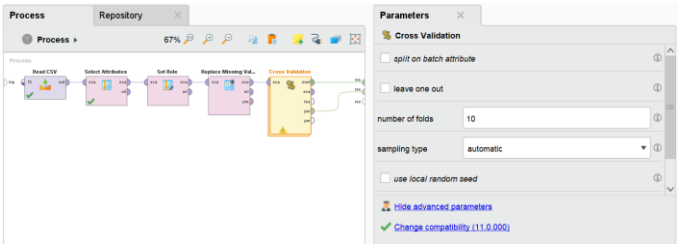
Selanjutnya dilakukan proses *Replace Missing Values* dalam menangani data hilang dalam *dataset*. Semua atribut dipilih untuk diproses, dan nilai yang hilang diubah dengan rata-rata dari atribut terkait. Langkah ini memastikan data tetap lengkap dan dapat digunakan dalam analisis tanpa mengganggu hasil pemodelan.



Gambar 4. *Replace Missing Values*

3.4. *Data Transformation*

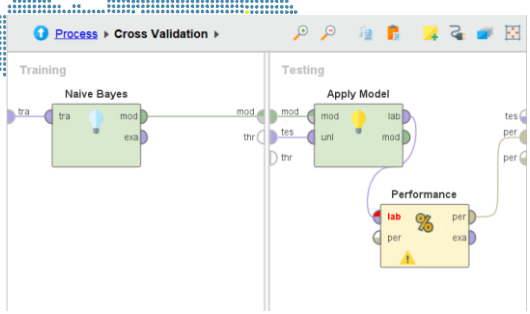
Tahap *Cross Validation*, yang bertujuan untuk mengevaluasi performa model secara akurat dengan metode pembagian data pelatihan dan pengujian.



Gambar 5. Data *Transformation*

3.5. *Data Mining*

Tahap ini, *Cross Validation* di RapidMiner yang digunakan untuk mengevaluasi performa model. Dalam tahap *Training*, model *Naive Bayes* dilatih menggunakan data latih. Setelah itu, dalam tahap *Testing*, model sudah dilatih diimplementasikan dalam data uji melalui *Apply Model*. Temuan prediksi lalu dievaluasi memanfaatkan *Performance*, yang menampilkan metrik kinerja model, berupa akurasi atau *error rate*. Proses ini memastikan model diuji secara objektif sebelum digunakan pada data baru.



Gambar 6. *Data Mining*

3.6. *Evaluation*

Hasil evaluasi performa model dengan menggunakan *confusion matrix*. Menunjukkan bahwa model yang diuji memiliki akurasi sebesar 92.40% ± 2.99%, dengan *micro average* mencapai 92.41%, yang menunjukkan tingkat prediksi yang cukup baik secara keseluruhan. Dari hasil evaluasi, model berhasil mengklasifikasikan 421 sampel dengan benar sebagai kelas 0, sementara hanya 8 sampel salah diklasifikasikan sebagai kelas 1. Namun, terdapat 31 sampel kelas 1 salah dikelompokkan menjadi kelas 0, sedangkan 54 sampel berhasil diklasifikasikan dengan benar menjadi kelas 1. *Precision* dalam kelas 0 mencapai 98.14%, berarti hampir semua prediksi positif untuk kelas ini akurat, sementara precision untuk kelas 1 lebih rendah, 63.53%, melihatkan masih ada kesalahan mengidentifikasi kelas ini. Dari sisi *recall*, kelas 0 memiliki nilai 93.14%, sedangkan kelas 1 memiliki *recall* 87.10%, yang menunjukkan bahwa model lebih baik mendeteksi kelas 0 dibandingkan kelas 1. Hasil ini mengindikasikan bahwa meskipun model memiliki akurasi tinggi, masih terdapat kelemahan dalam membedakan kelas 1 dengan benar, yang dapat diperbaiki dengan teknik optimasi lebih lanjut.

accuracy: 92.40% +/- 2.99% (micro average: 92.41%)

	true 0	true 1	class precision
pred. 0	421	8	98.14%
pred. 1	31	54	63.53%
class recall	93.14%	87.10%	

Gambar 7. *Evaluation*

4. Kesimpulan

Penelitian ini bertujuan untuk membangun model klasifikasi tingkat kemiskinan di Indonesia menggunakan algoritma *Naïve Bayes* dan mengevaluasi performanya berdasarkan data sosial ekonomi. Hasil penelitian menunjukkan bahwa model *Naïve Bayes* mampu mengklasifikasikan status kemiskinan dengan akurasi yang tinggi, yaitu sebesar 92,40% ± 2,99%. Hal ini menunjukkan bahwa algoritma tersebut efektif dalam mengenali pola-pola yang {Formatting Citation} membedakan kelompok masyarakat miskin dan tidak miskin berdasarkan indikator sosial ekonomi yang tersedia. Selain itu, model memberikan performa terbaik pada kategori mayoritas, meskipun masih terdapat tantangan dalam mengklasifikasikan kelompok rentan yang jumlahnya lebih sedikit. Oleh karena itu, perlu dilakukan pengembangan lebih lanjut, seperti penyeimbangan data dan pengujian dengan algoritma lain, guna meningkatkan akurasi pada seluruh kelas secara merata. Secara keseluruhan, penelitian ini membuktikan bahwa *Naïve Bayes* dapat digunakan sebagai pendekatan yang andal untuk mendukung pengambilan keputusan dalam penyaluran bantuan sosial secara lebih tepat sasaran dan berbasis data.

Daftar Pustaka

[1] A. Sarjito, "Efektivitas Kebijakan Sosial Dalam Mengurangi Ketimpangan Pendapatan Dan Angka Kemiskinan," *J. Ilmu Sos. Polit. Hum.*, Vol. 6, P. 2023, 2023.

[2] F. Afrian And K. Latifah, "Klasifikasi Tingkat Kemiskinanddi Indonesia Menggunakan Algoritma Decision Tree," *Semin. Nas. Inform. Upgris*, Vol. 2, 2024, [Online]. Available: Https://Www.Kaggle.Com/Datasets/Ermila/Klasifikasi-Tingkat-Kemiskinan-Di-Indonesia

[3] M. Wilda Al -Aluf And Z. Fatah, "Klasifikasi Algoritma Decision Tree Untuk Tingkat Kemiskinan Di Indonesia," *J. Comput. Sci. Technol.*, Vol. 3, Pp. 55–62, Jan. 2025, Doi: 10.59435/Jocstec.V3i1.404.

[4] C. Fuadi Ahmad, N. Suarna, And G. Dwilestari, "Klasifikasi Data Kemiskinan Menggunakan Metode Naïve Bayes Untuk Mengetahui Tingkat Kemiskian Studi Kasus: Desa Karangasem Kecamatan Leuwimunding Majalengka," *J. Inform. Dan Teknol. Inf.*, Vol. 2, No. 2, Pp. 203–208, Nov. 2023, Doi: 10.56854/Jt.V2i2.190.

[5] J. S. Komputer, K. Buatan, And A. Ridwan, "Penerapan Algoritma Naïve Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus," *J. Sist. Komput. Dan Kecerdasan Buatan*, Sep. 2020.

[6] A. C. P, F. E. Hariyanto, N. L. E. Andini, And Z. C. S, "Klasifikasi Rumah Tangga Miskin Menggunakan Metode Naive Bayes (Studi Kasus: Provinsi Papua Tahun 2017)," *J. Sains Mat. Dan Stat.*, Jan. 2021.

[7] M. R. Argall *Et Al.*, "Mms Sitl Ground Loop: Automating The Burst Data Selection Process," *Front. Astron. Sp. Sci.*, Vol. 7, Sep. 2020, Doi: 10.3389/Fspas.2020.00054.

[8] B. M. V. Bernando, H. S. Mamede, J. M. P. Barroso, And V. M. P. D. Dos Santos, "Data Governance & Quality Management - Innovation And Breakthroughs Across Different Fields," *J. Innov. Knowl.*, Oct. 2024.

[9] M. F. M. Khalik And F. Arifin, "Klasifikasi Indeks Kedalaman Kemiskinan Provinsi Sulawesi Selatan Berbasis Decision Tree, K-Nearest Neighbor, Naive Bayes, Neural Network, Dan Random Forest," *J. Edukasi Dan Penelit. Inform.*, Aug. 2023.

[10] J. Mulya Kecamatan Sepatan Timur *Et Al.*, "Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes : Studi Kasus Di Desa Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes : Studi Kasus Di Desa Jati Mulya Kecamatan Sepatan Timur," *J. Multinetics*, May 2024.

[11] W. P. Nurmayanti, Di. A. L. Saky, M. Malthuf, M. Gazali, And R. H. Hirzi, "Penerapan Naive Bayes Dalam Mengklasifikasikan Masyarakat Miskin Di Desa Lepak," *Geodika J. Kaji. Ilmu Dan Pendidik. Geogr.*, Jun. 2021.

[12] A. Ifon Purnama, A. Aziz, A. Sartika Wiguna, And K. Kunci, "Penerapan Data Mining Untuk Mengklasifikasi Penerima Bantuan Pkh Desa Wae Jare Menggunakan Metode Naïve Bayes," *Kurawal J. Teknol. Inf. Dan Ind.*, Oct. 2020, [Online]. Available: Https://Jurnal.Machung.Ac.Id/Index.Php/Kurawal

[13] E. Aditya, I. G. K. Astawa, K. G. Limbong, G. Indrawan, G. Indrawan, And M. A. O. Gunawan, "Analisis Sentimen Pengguna Sistem E-Kinerja Desa Kabupaten Jembrana Menggunakan Metode Naive Bayes," *J. Teknol. Dan Sist. Inf. Bisnis*, Vol. 7, No. 1, Pp. 8–14, Jan. 2025, Doi: 10.47233/Jteksis.V7i1.1693.

[14] A. Duwo Jiwo Saputro, A. Darmawan, And B. Nurina Sari, "Klasifikasi Persentase Kemiskinan Di Jawa Barat Menggunakan Data Mining Algoritma K-Nearest Neighbor (Knn)," 2023.

[15] E. S. Utami And Y. Setyawan, "Klasifikasi Kabupaten/Kota Di Indonesia Berdasarkan Tingkatkedalaman Dan Keparahan Kemiskinan Menggunakan Naive Bayes Classifier Dan K-Nearest Neighbor," *Pros. Semin. Nas. Apl. Sains Teknol.*, Nov. 2022.