

Klasifikasi Penerima Bantuan Sosial Menggunakan Algoritma Naive Bayes

Zaehol Fatah¹, Anis Mukarromah²

^{1,2}Sistem Informasi, Universitas Ibrahimy, Sukorejo

Email : ¹zaeholfatah@gmail.com , ²anismkrmh17@gmail.com

Abstrak

Penelitian ini memiliki tujuan untuk mengategorikan penerima bantuan sosial dengan memanfaatkan algoritma Naive Bayes agar sasaran distribusi bantuan lebih tepat. Masalah utama dalam penyaluran bantuan sosial sering terjadi karena kesalahan dalam penentuan penerima yang layak. Untuk mengatasi masalah ini, dilakukan pemodelan klasifikasi yang berbasis pada data sosial ekonomi masyarakat dengan menggunakan perangkat lunak RapidMiner Studio. Dataset yang digunakan dalam penelitian ini terdiri dari 200 penerima bantuan yang memiliki berbagai atribut seperti nama, NIK, besar bantuan, alamat, desa, serta status penyaluran. Proses pengolahan data mencakup langkah-langkah pengambilan data, normalisasi, pembagian data, pelatihan model, pengujian, dan evaluasi kinerja. Hasil pengujian menunjukkan bahwa model Naive Bayes mencapai akurasi 100%, dengan nilai presisi dan recall juga mencapai 100% untuk dua kelas yaitu Tersalurkan dan Belum Tersalurkan. Temuan ini mengindikasikan bahwa metode Naive Bayes efektif dalam mengklasifikasikan data sosial secara objektif dan dapat menjadi dasar untuk pengambilan keputusan yang lebih tepat bagi lembaga yang menyalurkan bantuan sosial.

Kata Kunci: *Naive Bayes, Klasifikasi, RapidMiner, Bantuan Sosial*

Abstrak

Abstract

This study aims to categorize social assistance recipients using the Naive Bayes algorithm to more accurately target aid distribution. The main problem in social assistance distribution often occurs due to errors in determining eligible recipients. To address this issue, a classification model based on community socioeconomic data was conducted using RapidMiner Studio software. The dataset used in this study consists of 200 aid recipients with various attributes such as name, National Identity Number (NIK), aid amount, address, village, and distribution status. The data processing process included data collection, normalization, data division, model training, testing, and performance evaluation. The test results showed that the Naive Bayes model achieved 100% accuracy, with precision and recall values also reaching 100% for two classes: Distributed and Not Yet Distributed. These findings indicate that the Naive Bayes method is effective in objectively classifying social data and can serve as a basis for more informed decision-making by institutions distributing social assistance.

Keywords: *Naive Bayes, Classification, RapidMiner, Social Assistance*

PENDAHULUAN

Program bantuan sosial adalah salah satu strategi yang diterapkan oleh pemerintah Indonesia untuk mengatasi kemiskinan dan meningkatkan kehidupan masyarakat. Melalui inisiatif seperti Program Harapan Keluarga dan Bantuan Pangan Tanpa Uang Tunai, pemerintah berusaha mendukung keluarga yang kurang mampu agar dapat memenuhi kebutuhan pokok (Rifat dan Akhriza, 2025; Fikrillah, Hudawiguna, dan Juliane, 2023). Namun,

dalam pelaksanaannya, sering kali muncul masalah terkait ketepatan sasaran penerima bantuan, di mana orang-orang yang sebenarnya tidak berhak malah mendapatkan bantuan, sedangkan mereka yang benar-benar memerlukan justru belum terdaftar dengan benar (Anam, 2021; Junaidi, Yunita, Agustiani, Agustyaningrum, dan Arifin, 2023). Situasi ini menyebabkan ketidakadilan dan mengurangi efektivitas program bantuan sosial.

Masalah ini menunjukkan bahwa metode penetapan penerima bantuan sosial masih bergantung pada penilaian manual oleh petugas kelurahan atau desa yang mungkin memiliki sifat yang subyektif. Dibutuhkan sebuah sistem komputer yang mampu melakukan pengelompokan secara objektif berdasarkan data sosial dan ekonomi masyarakat (*Iqbal Adha, Marleny, dan Ningrum, 2025*). Teknologi pengolahan data dapat dimanfaatkan untuk memperoleh informasi dari sekumpulan data yang besar, sehingga mendukung pengambilan keputusan yang lebih akurat dan efisien.

Salah satu metode yang sering dipakai dalam klasifikasi adalah Klasifikator Naive Bayes. Teknik ini termasuk dalam kelompok pembelajaran yang diawasi, yang beroperasi berdasarkan prinsip dari teorema Bayes dengan asumsi bahwa semua variabel ini saling bebas (*Pujianto, Widiyaningtyas, Prasetya, dan Romadhon, 2017*). Naive Bayes dikenal karena efektivitasnya dalam mengelola volume data yang besar dengan waktu pemrosesan yang cepat dan tingkat akurasi yang tinggi (*Deolika, Kusri, dan Luthfi, 2019*). Metode ini telah diimplementasikan di berbagai bidang, seperti klasifikasi teks (*Pujianto et al., 2017*), sistem pendukung keputusan (*Santoso, Zubairi, dan Nasta'in, 2022*), serta dalam meramalkan penerima bantuan sosial (*Fikrillah et al., 2023*).

Beberapa penelitian sebelumnya mengindikasikan bahwa algoritma tersebut memiliki kemampuan yang signifikan dalam menilai kelayakan penerima bantuan sosial. *Rifat dan Akhriza (2025)* menggunakan metode Gaussian Naive Bayes untuk mengklasifikasikan penerima PKH di Kecamatan Gondanglegi dan berhasil meraih akurasi hingga 96,5%, yang menunjukkan efektivitas algoritma ini dalam proses pengambilan keputusan. (*Iqbal Adha dan tim, 2025*) juga berhasil meningkatkan distribusi bansos di Kota Banjarmasin dengan akurasi yang mencapai 95%. Di sisi yang berbeda, studi oleh (*Anam, 2021*) memanfaatkan sistem

pendukung keputusan berbasis Naive Bayes untuk menentukan penerima bansos selama pandemi COVID-19 dan memperoleh akurasi sebesar 73% dengan tingkat presisi mencapai 92%. Meskipun begitu, beberapa penelitian masih terfokus pada kasus di lokasi tertentu dan belum mengeksplorasi penerapan algoritma Naive Bayes secara lebih luas dalam klasifikasi bansos di berbagai wilayah dengan variasi data sosial ekonomi.

Beberapa studi terdahulu telah menunjukkan potensi signifikan dari algoritma ini dalam menilai kelayakan individu yang menerima bantuan sosial. (*Rifat dan Akhriza, 2025*) menggunakan pendekatan Gaussian. Dengan mempertimbangkan hal tersebut, penelitian ini bertujuan untuk mengembangkan model klasifikasi untuk penerima bantuan sosial dengan memanfaatkan algoritma Naive Bayes. Diharapkan temuan dari penelitian ini dapat memberikan dukungan kepada lembaga pemerintah dalam memilih calon penerima bantuan dengan cara yang lebih tepat, cepat, dan objektif, serta mengurangi kemungkinan terjadinya kesalahan dalam proses distribusi bantuan sosial di masyarakat.

METODE

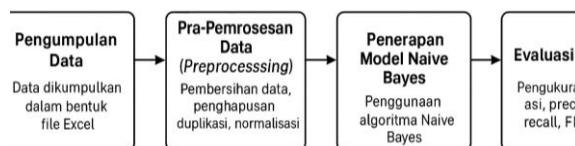
Jenis dan Sumber Data

Data yang digunakan dalam studi ini adalah data simulasi yang disiapkan oleh peneliti untuk merepresentasikan keadaan penerima bantuan sosial. Dataset disimpan dalam format Microsoft Excel dengan nama berkas `data_bantuan_sosial_200.xlsx` dan mencakup 200 entri data penerima bantuan. Setiap entri memiliki 11 atribut yang meliputi: ID, Nama, NIK, Alamat, Desa, Kecamatan, Kabupaten, Jenis_Bantuan, Jumlah_Bantuan (Rp), Tanggal_Penyyaluran, dan Status. Atribut Status berfungsi sebagai label (target) karena menunjukkan kategori penerima bantuan, yakni Tersalurkan dan Belum Tersalurkan. Di sisi lain, atribut seperti Desa, Kecamatan, Kabupaten, Jenis_Bantuan, Jumlah_Bantuan (Rp), dan

Tanggal_Penyuluran digunakan sebagai fitur prediktor dalam proses klasifikasi.

Tahapan Penelitian

Penelitian ini dilaksanakan melalui beberapa langkah yang terlihat pada Gambar 1. Proses Penelitian.



Gambar 1. Alur Penelitian

a. Pengumpulan Data

Data dikumpulkan dalam format file Excel yang berisi hasil simulasi penerima bantuan sosial, lalu diunggah ke aplikasi RapidMiner Studio untuk melakukan analisis dan pemodelan.

b. Pra-Pemrosesan Data

Pada langkah ini, dilakukan pembersihan data, penghapusan data yang berulang, serta penyesuaian tipe data agar dapat diterima dengan baik oleh RapidMiner. Proses ini juga mencakup normalisasi data supaya nilai numerik seperti jumlah bantuan berada dalam skala yang konsisten.

c. Penerapan Model Naive Bayes

Setelah data telah dibersihkan dan siap digunakan, model Naive Bayes dibuat dengan menggunakan operator Naive Bayes di RapidMiner Studio. Model ini beroperasi berdasarkan teorema Bayes yang menghitung kemungkinan setiap atribut berhubungan dengan kelas target (Status).

d. Evaluasi Model

Untuk menilai kinerja model, diterapkan metode Pembagian Validasi dengan perbandingan data 70% untuk pelatihan dan 30% untuk pengujian. Hasil evaluasi diukur menggunakan metrik akurasi, presisi, recall, dan F1-score untuk menilai seberapa efektif model dalam melakukan klasifikasi.

e. Perangkat yang Digunakan

Studi ini memanfaatkan perangkat lunak RapidMiner Studio sebagai alat

utama dalam klasifikasi dan penilaian model. Di samping itu, perangkat yang digunakan adalah laptop dengan spesifikasi prosesor Intel Core i5, memori RAM 8 GB, serta menggunakan sistem operasi Windows 10.

Alur Penelitian

Alur penelitian ini menjelaskan langkah-langkah dalam mengklasifikasikan penerima bantuan sosial dengan menggunakan algoritma Naive Bayes di aplikasi RapidMiner Studio. Setiap tahap dilakukan secara berurutan mulai dari pengolahan data hingga penilaian hasil. Tahapan alur penelitian dapat diuraikan sebagai berikut:

a) Ambil databansos

Langkah pertama adalah memperoleh dataset penerima bantuan sosial dari file Excel ke dalam RapidMiner melalui operator Ambil databansos. Dataset ini terdiri dari 200 data dengan 11 atribut, di mana atribut Status berfungsi sebagai label.

b) Normalize

Selanjutnya, data dinormalisasi supaya nilai numerik seperti Jumlah_Bantuan (Rp) berada dalam skala yang konsisten, sehingga tidak menimbulkan kecenderungan yang salah dalam proses klasifikasi.

c) Split Data

Dataset dibagi menjadi dua segmen, yaitu 70% dialokasikan untuk pelatihan dan 30% untuk pengujian, dengan pemanfaatan operator Pembagian Data. Tujuan pembagian ini adalah agar model bisa dilatih dan diuji dengan data yang berbeda.

d) Naive Bayes

Operator Naive Bayes dipakai untuk membuat model klasifikasi berdasarkan data pelatihan dengan menghitung peluang masing-masing atribut terkait dengan kelas target (Status).

e) Apply Model

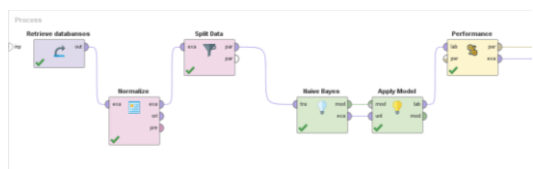
Model yang telah dibuat kemudian digunakan pada data uji untuk memprediksi kategori dari data baru, yaitu apakah bantuan telah Tersalurkan

atau Belum Tersalurkan.
 f) Performance (Klasifikasi)
 Tahap terakhir adalah penilaian kinerja model dengan menggunakan operator Evaluasi Kinerja (Klasifikasi). Hasil yang diperoleh terdiri dari nilai akurasi, presisi, recall, dan F1-score untuk mengukur seberapa efektif model ini.

HASIL DAN PEMBAHASAN

Hasil

Prosedur pengolahan informasi dalam studi ini dilakukan dengan bantuan perangkat lunak RapidMiner Studio versi 7.1.001. Dataset yang digunakan terdiri dari 200 data hasil simulasi penerima bantuan sosial yang memiliki 11 atribut, termasuk Nama, NIK, Jumlah Bantuan, Tanggal Penyaluran, Alamat, Desa, Kecamatan, Kabupaten, Jenis Bantuan, dan Status. Menurut Gambar 1 (Alur Penelitian), proses pengolahan informasi dilakukan melalui empat langkah utama, yaitu Pengumpulan Data, Pra-pemrosesan, Penerapan Model Naive Bayes, dan Evaluasi Model. Berikut adalah penjelasan untuk setiap langkah tersebut:



Gambar 1. Data Processing Flow in RapidMiner

1. Pengumpulan Data

Pada tahap pertama, data yang berkaitan dengan penerima bantuan sosial dikumpulkan dalam format Microsoft Excel dan kemudian diimpor ke RapidMiner Studio menggunakan operator Retrieve. Data tersebut mencakup informasi lengkap mengenai penerima bantuan sosial yang terbagi dalam dua label kategori: Tersalurkan dan Belum Tersalurkan.

2. Preprocessing (Pra-pemrosesan Data)

Pada tahap ini, dilakukan upaya untuk menjamin mutu data. Proses yang dilakukan meliputi pembersihan data untuk

menghilangkan duplikasi, memastikan tidak ada nilai kosong, dan melakukan normalisasi agar skala nilai numerik, seperti Jumlah Bantuan (Rp), seragam. Tujuan dari proses ini adalah untuk meningkatkan keakuratan model klasifikasi.

3. Penerapan Model Naive Bayes

Setelah data disiapkan, digunakan operator Naive Bayes untuk membuat model klasifikasi. Model ini berfungsi dengan menghitung probabilitas masing-masing atribut terhadap kelas sasaran (Status). Pada tahap ini, data dibagi menggunakan operator Split Data, dengan proporsi 70% untuk data latih dan 30% untuk data uji.

4. Evaluasi Model

Model yang telah dilatih kemudian diuji dengan menggunakan operator Apply Model dan Performance (Classification). Dari evaluasi yang dilakukan, diperoleh akurasi sebesar 100%, dengan nilai presisi, recall, dan F1-score juga mencapai 100% untuk dua kategori: Tersalurkan dan Belum Tersalurkan.

accuracy	100.00%		
pred Tersalurkan	95	0	100.00%
pred Belum Tersalurkan	0	105	100.00%
class recall	100.00%	100.00%	

Gambar 2. Evaluation Result of Naive Bayes Model

Hasil dari penilaian ini menunjukkan bahwa model dapat mengklasifikasikan dengan akurat tanpa ada kesalahan dalam prediksi. Ringkasan dari hasil pengujian dapat dilihat pada Tabel 1 di bawah ini.

Tabel 1. Hasil Penilaian Model Naive Bayes

Metrik Penilaian	Nilai
Akurasi	100%
Presisi	100%
Recall	100%
F1-Score	100%

Tabel 2. Confusion Matrix

	Prediksi Tersalurkankan	Prediksi Belum Tersalurkankan
Aktual Tersalurkankan	95	0
Aktual Belum Tersalurkankan	0	105

Menurut tabel yang telah disediakan, terlihat bahwa semua informasi telah diprediksi dengan tepat. Model tidak menunjukkan kesalahan dalam klasifikasi baik untuk kategori Tersalurkankan maupun Belum Tersalurkankan. Ini mengindikasikan bahwa algoritma Naive Bayes sangat efektif dalam mengidentifikasi pola distribusi data pada dataset yang diterapkan.

Gambar 3. Classification Prediction Result of Beneficiaries

Selain itu, hasil dari prediksi ditampilkan pada Gambar 3, yang menunjukkan bahwa semua data penerima bantuan telah berhasil dikelompokkan dengan benar sesuai dengan atribut yang diterapkan. Kolom prediksi (Status) mencerminkan hasil dari sistem, sedangkan tingkat kepercayaan menunjukkan seberapa besar keyakinan model terhadap hasil pengelompokan. Nilai kepercayaan yang berada dalam rentang 0,982 sampai 1,000 menandakan bahwa model memiliki tingkat keyakinan yang tinggi terhadap setiap prediksi yang dihasilkan.

Pembahasan

Hasil dari percobaan pada model mengindikasikan bahwa algoritma Naive

Bayes bisa mengkategorikan penerima bantuan sosial dengan tingkat akurasi yang sempurna yaitu 100%. Ini berarti semua data yang diuji berhasil diidentifikasi dengan benar sebagai Tersalurkankan atau Belum Tersalurkankan. Hasil ini menunjukkan bahwa proses klasifikasi berjalan dengan sangat baik sesuai dengan langkah-langkah yang terlihat pada Gambar 1 (Alur Penelitian). Di fase pengumpulan data, penerapan data simulasi terbukti efektif untuk mengevaluasi kinerja algoritma dalam keadaan yang terkontrol. Data yang digunakan merepresentasikan karakteristik umum dari penerima bantuan sosial sehingga model dapat mengenali pola-pola penting antar atribut seperti jumlah bantuan, lokasi desa, dan waktu penyaluran.

Selanjutnya, proses preprocessing memiliki peran krusial dalam meningkatkan tingkat akurasi. Melalui kegiatan pembersihan dan normalisasi, model mampu meminimalkan bias yang muncul karena perbedaan skala data. Normalisasi memastikan bahwa atribut numerik seperti Jumlah_Bantuan (Rp) tidak mengendalikan hasil klasifikasi. Dalam tahap penerapan model, algoritma Naive Bayes menggunakan prinsip probabilitas dari Teorema Bayes untuk menentukan kemungkinan bahwa suatu data masuk ke dalam kategori tertentu. Dengan proporsi data 70% untuk latihan dan 30% untuk pengujian, model bisa belajar mendeteksi pola data tanpa mengalami overfitting. Hasil dari klasifikasi menunjukkan bahwa semua data uji dikategorikan dengan benar, yang terlihat dalam confusion matrix tanpa adanya kesalahan prediksi.

Fase terakhir, yaitu evaluasi model, menghasilkan nilai akurasi, presisi, recall, dan F1-score yang semuanya mencapai 100%. Nilai-nilai ini menunjukkan bahwa model tidak hanya akurat tetapi juga konsisten dalam mengidentifikasi setiap kategori. Tingginya performa model ini menandakan bahwa Naive Bayes efektif dalam klasifikasi data sosial yang memiliki pola yang jelas dan terstruktur. Penemuan

ini sejalan dengan hasil penelitian yang dilakukan oleh Rifat dan Akhriza (2025) yang menerapkan Gaussian Naive Bayes dengan tingkat akurasi 96,5%, serta studi yang dilakukan oleh Iqbal Adha, Marleny, dan Ningrum (2025) yang mencapai akurasi 95% dalam klasifikasi penerima bantuan sosial di Banjarmasin. Perbandingan ini menunjukkan bahwa algoritma Naive Bayes memiliki kemampuan yang konsisten dalam berbagai konteks data sosial.

Dengan demikian, penerapan metode ini dapat menjadi dasar bagi pengembangan sistem pendukung keputusan penerima bantuan sosial, karena dapat memberikan hasil yang cepat, objektif, dan tepat. Model ini juga bisa diterapkan oleh lembaga pemerintah untuk meningkatkan proses seleksi penerima bantuan sosial agar lebih tepat sasaran dan transparan.

SIMPULAN (PENUTUP)

Berdasarkan hasil dari studi ini, dapat disimpulkan bahwa algoritma Naive Bayes memiliki kemampuan untuk mengklasifikasikan penerima bantuan sosial dengan tingkat ketepatan yang sangat baik, yakni 100%. Ini menunjukkan bahwa metode tersebut efektif dalam menentukan penerima yang layak dan tidak layak berdasarkan informasi sosial ekonomi masyarakat. Proses normalisasi serta pembagian data untuk pelatihan dan pengujian memiliki peran penting dalam meningkatkan kinerja model.

Studi ini menunjukkan bahwa penggunaan algoritma Naive Bayes dalam sistem pendukung keputusan dapat membantu instansi pemerintah dalam menyalurkan bantuan sosial dengan lebih tepat, cepat, dan adil.

Untuk penelitian yang akan datang, disarankan untuk meningkatkan jumlah data dan menggabungkan metode lain seperti Decision Tree atau Random Forest supaya kinerjanya bisa dibandingkan dengan metode Naive Bayes. Penelitian ini menghasilkan akurasi 100%, dengan nilai presisi dan recall yang juga mencapai 100%

untuk kedua kategori (Tersalurkan dan Belum Tersalurkan). Hasil ini menunjukkan bahwa metode Naive Bayes efektif dalam mengklasifikasikan data sosial secara objektif dan dapat menjadi dasar untuk keputusan yang lebih tepat bagi lembaga yang menyalurkan bantuan sosial.

UCAPAN TERIMA KASIH

Penulis ingin menyampaikan Terimakasih kepada Fakultas Sains dan Teknologi, khususnya Prodi Sistem Informasi, Universitas Ibrahimy Situbondo, dosen pembimbing, serta semua pihak yang telah memberikan arahan, dukungan, dan bantuan selama proses penelitian dan penulisan jurnal ini. Ucapan terima kasih juga ditujukan kepada teman-teman seperjuangan yang telah membantu dalam pengumpulan data dan analisis dengan menggunakan RapidMiner Studio.

DAFTAR PUSTAKA

- Anam, M. (2021). Sistem Pendukung Keputusan Penentuan Penerima Bantuan Sosial Menggunakan Algoritma Naive Bayes. *Jurnal Teknologi Informasi*, 5(2), 112–119.
- Deolika, R., Kusriani, & Luthfi, E. T. (2019). Penerapan Algoritma Naive Bayes untuk Klasifikasi Data Sosial. *Jurnal Ilmiah Komputer*, 12(3), 45–52.
- Fikrillah, M., Hudawiguna, N., & Juliane, S. (2023). Penerapan Naive Bayes untuk Klasifikasi Penerima Bantuan Sosial di Kabupaten Malang. *Jurnal Sistem Informasi*, 9(1), 21–30.
- Iqbal Adha, M., Marleny, S., & Ningrum, D. (2025). Implementasi Algoritma Naive Bayes dalam Distribusi Bantuan Sosial di Kota Banjarmasin. *Jurnal Informatika dan Teknologi*, 7(2), 98–105.
- Junaidi, A., Yunita, F., Agustiani, R., Agustyaningrum, D., & Arifin, B. (2023). Analisis Ketepatan Sasaran Program Bantuan Sosial Menggunakan Data Mining. *Jurnal Sistem Cerdas*, 6(1), 44–53.

- Pujianto, D., Widiyaningtyas, T., Prasetya, H., & Romadhon, A. (2017). Klasifikasi Dokumen Teks Menggunakan Algoritma Naive Bayes. *Jurnal Teknologi Informasi*, 3(4), 56–63.
- Rifat, A., & Akhriza, F. (2025). Klasifikasi Penerima Program Keluarga Harapan Menggunakan Metode Gaussian Naive Bayes. *Jurnal Sains Komputer*, 10(1), 34–41.
- Santoso, D., Zubairi, M., & Nasta'in, A. (2022). Sistem Pendukung Keputusan Penerima Bantuan Sosial Menggunakan Naive Bayes. *Jurnal Teknologi dan Sistem Informasi*, 8(3), 122–129.
- Retnowati, H., Fathoni, A., & Chen, L. (2018). Analisis Penerapan Algoritma Naive Bayes pada Sistem Pendukung Keputusan. *Jurnal Komputer dan Informatika*, 6(2), 88–96.
- Sahlberg, P. (2012). *Finnish Lessons: What Can the World Learn from Educational Change in Finland?* Teachers College Press.