Analisis Perbandingan Metode Naïve Bayes dan K-NN dalam Klasifikasi Kelayakan Keluarga Terdaftar DTKS Penerimaan Bantuan Sosial di Desa Dubesi

Delviana Muti¹, Fandi Yulian Pamuji²

^{1,2}Sistem Informasi Universitas Merdeka Malang Jl. Terusan Raya Dieng No. 62-64 e-mail: delviana.muti@tudent.unmer.ac.id¹, fandi.pamuji@unmer.ac.id²

Kata Kunci:

Naïve Bayes KNN Keluarga Terdaftar DTKS

ABSTRAK

Koordinasi Bantuan Sosial Indonesia merupakan program bantuan pemerintah berdasarkan peraturan pemerintah RI No. 60 Tahun 2014 tentang penghematan anggaran berdasarkan pendapatan dan belanja negara. Pada Desa Dubesi, yang terletak di Kecamatan Nanaet Dubesi, Kabupaten Belu Atambua, Provinsi Nusa Tenggara Timur, perhitungan untuk layak atau tidak layak penerima bantuan sosial masih secara manual, terdapat beberapa masalah yang terjadi mengenai salah sasaran yang disebabkan kesalahan keluarga yang layak atau tidak layak penerima bantuan sosial. Penelitian ini bertujuan untuk menganalisis perbandingan metode Naive Bayes dan metode K-NN dalam klasifikasi kelayakan keluarga terdaftar dtks penerima bantuan sosial di Desa Dubesi. Penelitian menggunakan metode Naive Bayes dan K-NN Hasil analisis penelitian ini menunjukkan bahwa setelah menggunakan metode Naive Bayes dan K-NN untuk mengklasifikasikan kelayakan keluarga penerima bantuan sosial yang terdaftar di Desa Dubesi. Ditemukan bahwa dari 80 informasi data uji, Naive Bayes memiliki nilai akurasi 82% dan presisi 89% sedangkan K-NN (K-Nearest Neighbor) memiliki nilai akurasi 70% dan presisi 73%. Manajemen Kantor Desa Dubesi dapat menggunakan kontribusi penelitian ini untuk mengevaluasi kelayakan keluarga yang terdaftar sebagai penerima bantuan sosial di Desa Dubesi, Kecamatan Nanaet Dubesi, Kabupaten Belu Atambua.

Keyword:

Naïve Bayes KNN DTKS Registered Family

ABSTRACT

Indonesian Social Assistance Coordination is a government assistance program based on the Indonesian government regulation No. 60 of 2014 concerning budget savings based on state revenues and expenditures. In Dubesi Village, which is located in Nanaet Dubesi District, Belu Atambua Regency, East Nusa Tenggara Province, calculations for eligibility or ineligibility for social assistance recipients are still manual, there are several problems that occur regarding mistargeting caused by errors in families who are eligible or ineligible for social assistance recipients. This study aims to analyze the comparison of the Naive Bayes method and the K-NN method in the classification of the eligibility of registered families of social assistance recipients in Dubesi Village. The research uses the Naive Bayes and K-NN methods. It was found that out of 80 test data information, Naive Bayes has an accuracy value of 82% and a precision of 89% while K-NN (K-Nearest Neighbor) has an accuracy value of 70% and a precision of 73%. The management of Dubesi Village Office can use the contribution of this research to evaluate the eligibility of families registered as social

PENDAHULUAN

Koordinasi Bantuan Sosial Indonesia merupakan program bantuan pemerintah berdasarkan peraturan pemerintah RI No. 60 Tahun 2014 tentang penghematan anggaran berdasarkan pendapatan dan belanja negara. Kemiskinan bukan hanya ketidakmampuan untuk memperoleh kebutuhan dasar seperti pangan, sandang, papan dan transportasi, tetapi juga ketidakmampuan untuk berpartisipasi dalam memecahkan berbagai masalah yang berkaitan dengan pendidikan, kesehatan dan pembangunan sumber daya manusia[1]. Pemerintah telah melakukan berbagai upaya untuk mengentaskan kemiskinan, salah satunya dengan menyalurkan dana di tingkat desa. Dana desa adalah anggaran desa dan desa adat yang dialihkan melalui APBD (Anggaran Pendapatan dan Belanja Daerah) kabupaten/desa untuk membiayai penyelenggaraan pemerintahan , pembangunan, dan pemberdayaan masyarakat, dan masyarakat[2].

Pada Desa Dubesi, yang terletak di Kecamatan Nanaet Dubesi, Kabupaten Belu Atambua, Provinsi Nusa Tenggara Timur, perhitungan untuk layak atau tidak layak penerima bantuan sosial masih secara manual, terdapat beberapa masalah yang terjadi mengenai salah sasaran yang disebabkan kesalahan keluarga yang layak atau tidak layak penerima bantuan sosial[3]. Dengan adanya penelitian menggunakan metode Naive Bayes dan K-NN diharapkan bisa membantu Desa Dubesi untuk lebih mudah dalam perhitungan layak atau tidak layak dalam keluarga terdaftar DTKS penerima bantuan sosial salah satunya bisa menggunakan aplikasi orange[4].

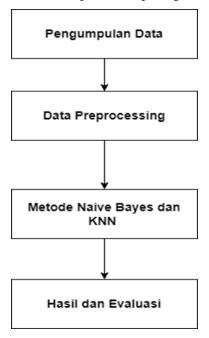
Perbandingan penduduk miskin menggunakan metode Naive bayes, dan variabel yang digunakan termasuk jenis kelamin, peserta PKH, jumlah anggota rumah tangga, status kepemilikan rumah tangga, perlengkapan rumah tangga, dan alat elektronik. Dengan menggunakan algoritma K-Nearest Neighbor (K-NN), aparat desa dapat menggunakan algoritma ini sebagai acuan dalam mengevaluasi penerima bantuan sosial dan mengurangi kesalahan informasi dalam mengidentifikasi penerima bantuan sosial[5]. Berdasarkan penjelasan di atas, penelitian ini bertujuan untuk menganalisis perbandingan metode Naive Bayes dan metode K-NN dalam klasifikasi kelayakan keluarga terdaftar DTKS penerima bantuan sosial di Desa Dubesi.

METODE

Dalam penelitian ini penerapan Metode yang digunakan dalam penelitian ini merupakan metode Naïve Bayes dan metode K-NN. Metode ini dapat dilakukan dengan mengklasifikasi kelayakan keluarga terdaftar DTKS penerima bantuan sosial di Desa Dubesi. Dalam metode penelitan ini dipelajari bagaimana proses pengumpulan data untuk kegiatan penelitian. Pada penelitian ini metode yang digunakan untuk pengambilan data ialah wawancara. Sedangkan metode

Copyright © SENASIF 2023 ISSN: 2598-0076

yang digunakan untuk analisis perbandingan Data preprocessing, Naïve Bayes dan K-NN (K-Nearest Neighbor)[6]. Berikut ini adalah alur penelitian pada gambar 1 dibawah ini:



Gambar 1. Alur Penelitian

Populasi dan Sampel

1. Populasi

Pengertian populasi merupakan wilayah generasi yang terdiri atas objek/subjek yang mempunyai kualitas dan karakteristik tertentu yang ditetapkan oleh peneliti untuk mempelajari dan dikeluarkan kesimpulannya. Populasi dalam penelitian ini adalah jumlah dataset Keluarga bantuan social 80 orang yang ada di Desa Dubesi.

2. Sampel

Sampel merupakan bagian dari jumlah dan karakteristik yang dimiliki oleh populasi tersebut. Dalam penelitian ini adalah Kelayakan keluarga penerimaan bantuan sosial di tahun 2023 di Desa Dubesi sejumlah 80 orang. Data ini berasal dari Desa Dubesi Kecamatan Nanaet Dubesi.

Teknik Pengumpulan Data

Pengumpulan data merupakan tahap yang dilakukan manfaat mendapatkan berbagai informasi secara lengkap dan rinci manfaat mencapai tujuan penelitian. Setelah data yang dibutuhkan telah terkumpul, maka dapat dilakukan transformasi data dan normalisasi data[7]. Adapun cara yang dapat dilakukan untuk pengumpulan data sebagai berikut:

1. Wawancara

Pada penelitian ini wawancara dilakukan oleh penelitian dengan bapak Andreas Atok selaku kepala desa Dubesi wawancara mengenai tentang kelayakan keluarga terdaftar DTKS penerima bantuan sosial.

2. Studi Pustaka

Metode ini dilakukan dengan cara mempelajari literatur-teori dan buku-buku serta situssitus penyediaan layanan yang berhubungan dengan objek penelitian.

3. Dataset Bantuan Sosial

Dalam dataset ini merupakan dataset yang berupa data jumlah kelayakan keluarga yang menerima Bantuan sosial seperti nama kelapa rumah tangga, Pkh, jumlah tanggungan, kondisi rumah, jumlah penghasilan, status pemilik rumah, status, jabatan, alamat, satuan, jumlah, total penduduk, yang berasal dari Desa Dubesi Kec. Nanaet Duabesi

4. Dokumentasi

Dokumentasi adalah suatu cara yang digunakan untuk memperoleh data dan informasi dalam bentuk buku, arsip, dokumen, tulisan angka dan gambar yang berupa laporan serta keterangan yang dapat mendukung penelitian dataset kelayakan keluarga terdaftar DTKS penerimabantuan social di Desa Dubesi Kecamatan Nanaet Dubesi.

Metode Naïve Bayes

Naive Bayes merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Algoritma mengunakan teorema Bayes dan mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas. Definisi lain mengatakan Naive Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya[8]. Perhitungan rumus Naïve Bayes dibawah ini:

$$P(H|X) = \frac{p(X|H).p(H)}{p(x)}$$
(1)

Keterangan:

P(H|X) = Probabilitas hipotesis H berdasarkan kondisi X (posteriori probabilitas)

H = Hipotesis data merupakan suatu class spesifik

X = Data dengan class yang belum diketahui

P(X|H) = Probabilitas X berdasarkan kondisi pada hipotesis H

P(H) = Probabilitas hipotesis H (prior probabilitas)

P(X) = Probabilitas X

Copyright © SENASIF 2023 ISSN: 2598-0076

Metode K-NN

Algoritma K-Nearest Neighbor (KNN) adalah sebuah metode untuk melakukan klasifikasi terhadap obyek berdasarkan data pembelajaran yang jaraknya paling dekat dengan obyek tersebut. Prinsip kerja dari K-Nearest Neighbor (KNN) adalah mencari jarak terdekat antara data yang akan dievaluasi dengan K tetangga (neighbor) terdekatnya dalam data pembelajaran[9]. Perhitungan rumus K-Nearest Neighbor (KNN) dibawah ini:

$$d(x,y) = \sqrt{\sum_{i=1}^{n} (x^{i} training - y^{i} testing) 2}$$
(2)

Keterangan:

d(x,y) = Jarak

 x^{i} training = Data training

 y^i testing = Data testing

n = Dimensi data

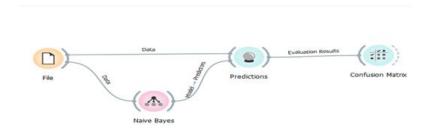
Hasil dan Evaluasi

Analisis data dalam penelitian ini merupakan tahap menyelesaikan permasalahan yang diperoleh dari hasil wawancara yang telah dipilih oleh peneliti dalam penelitian. Pada analisis data ini melakukan berdasarkan hasil wawancara dengan informasi sesuai dengan peneliti. Mengolah data Bantuan sosial dengan data preprocessing untuk mempersiapkan data mentah yang lebih cocok seperti data Transformation dengan tujuan untuk klasifikasi menggunakan metode Naïve Bayes dan metode K-NN dengan perbandingan probabilitas kelas setiap fitur dan menggabungkan probabilitas tersebut untuk mendapatkan probabilitas kelas yang mungkin[10].

HASIL DAN PEMBAHASAN

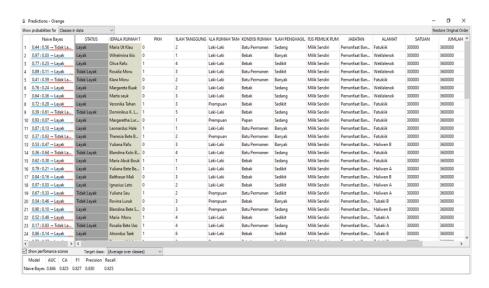
Hasil dan pembahasan dari penelitian ini merupakan penerapan dari metode Naive Bayes dan K-NN dengan menggunakan aplikasi Orange.

Klasifikasi Naïve Bayes menggunakan Aplikasi Orange



Gambar 2. Alur Naïve Bayes

Hasil Predictions Naïve Bayes



Gambar 3. Predictions Naïve Bayes

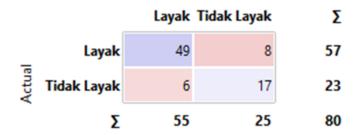
Pada Gambar 3 diatas dapat memperlihatkan metode Naïve Bayes dengan nilai AUC: **0.846**, CA: **0.825**, F1: **0.827**, Precision: **0.830**, Recall: **0.825**. Diketahui bahwa nilai AUC dengan nilai tinggi adalah metode Naïve Bayes yaitu 0.846. AUC untuk mengukur kinerja diskriminatif dengan memperkirakan probabilitas output dari data yang diseleksi secara acak dari populasi positif dan negative, semakin besarnya AUC, maka semakin baik hasil klasifikasi yang digunakan.

Hasil Confusion Matrix Naïve Bayes

Dari tabel tersebut dapat diketahui bahwa ada 49 data yang diprediksi layak dan kenyataannya ada 6 data yang dinyatakan tidak layak dan ada 17 data yang diprediksi tidak layak dan kenyataannya ada 8 yang layak seperti gambar 4 dibawah ini.

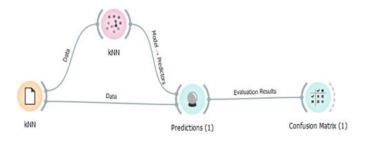
Copyright © SENASIF 2023 ISSN: 2598-0076

Predicted



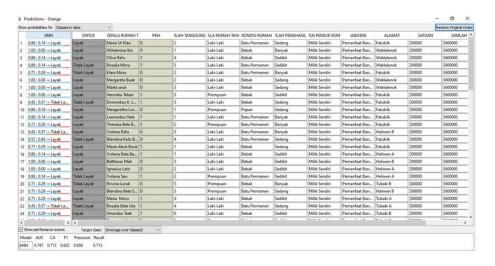
Gambar 4. Nilai Confusion Matrix Metode Naive Bayes

Klasifikasi K-Nearest Neighbor (K-NN) menggunakan Aplikasi Orange



Gambar 5. Alur K-NN

Hasil Predictions K-NN

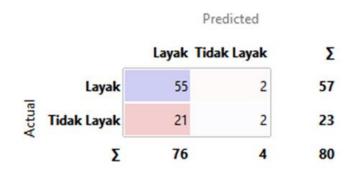


Gambar 6. Hasil Predictions

Berdasarkan Gambar 6 dapat memperlihatkan metode K-NN dengan nilai **AUC**: **0.747**, **CA**: **0.713**, **F1**: **0.632**, **Precision**: **0.659**, **Recall**: **0.713**. Diketahui bahwa nilai AUC dengan nilai tinggi adalah mote Naïve Bayes yaitu 0.747.

Hasil Confusions Matrix K-NN

Dari gambar tersebut dapat diketahui bahwa ada 55 data yang diprediksi layak dan kenyataannya ada 21 data yang dinyatakan tidak layak dan ada 2 data yang diprediksi tidak layak dan kenyataannya ada 2 yang layak. Seperti pada gambar 7 dibawah ini.



Gambar 7. Nilai Confusion Matrix metode K-NN

Pada hasil evaluasi ini dengan menggunakan Confusion Matrix yang diperoleh nilai perbandingan Accuracy, Precision dan Recall dari kedua metode Naive Bayes dan K-NN terlihat pada tabel dibawah ini.

 Metode
 Accuracy
 Precision
 Recall

 Naïve Bayes
 82%
 89%
 85%

 K-NN
 71%
 72%
 92%

Tabel 1. Perbandingan Kinerja

Pada Tabel 1 diatas kinerja metode Naive Bayes memiliki nilai akurasi 82%, Precision 89% dan Recall 85% sementara metode K-NN akurasi 71%, Precision 72% dan Recall 92%.

SIMPULAN

Hasil analisis penelitian ini menunjukkan bahwa setelah menggunakan metode Naive Bayes dan K-NN untuk mengklasifikasikan kelayakan keluarga penerima bantuan sosial yang terdaftar di Desa Dubesi. Tujuan untuk melakukan peningkatan akurasi penerima bantuan social dengan menggunakan metode Naïve Bayes dan metode K-NN, ditemukan bahwa dari 80 informasi data uji layak atau tidak layak menerima bantuan social di Desa Dubesi. Untuk menghasilkan peningkatan akurasi klasifikasi keluarga layak atau tidak layak penerima bantuan social dengan perbandingan metode Naïve Bayes dan metode K-NN memiliki nilai metode Naïve Bayes akurasi 82% dan presisi 89% sedangkan K-NN (K-Nearest Neighbor) memiliki nilai akurasi 71% dan

Copyright © SENASIF 2023 ISSN: 2598-0076 presisi 72%. Dari perbandingan metode Naïve Bayes dan metode K-NN yang memiliki akurasi yang tinggi adalah metode Naïve Bayes.

DAFTAR RUJUKAN

- [1] F. P.-S. N. S. I. (SENASIF) and undefined 2022, "Pengujian Metode SMOTE Untuk Penanganan Data Tidak Seimbang Pada Dataset Binary," *Jurnalfti.Unmer.Ac.Id*, vol. 2022, no. September, pp. 3200–3208, 2022, [Online]. Available: https://jurnalfti.unmer.ac.id/index.php/senasif/article/view/403
- [2] F. Y. Pamuji and V. P. Ramadhan, "Komparasi Algoritma Random Forest dan Decision Tree untuk Memprediksi Keberhasilan Immunotheraphy," *J. Teknol. dan Manaj. Inform.*, vol. 7, no. 1, pp. 46–50, 2021, doi: 10.26905/jtmi.v7i1.5982.
- [3] F. Y. Pamuji, S. Dwi, and A. Putri, "Komparasi Metode SMOTE dan ADASYN Untuk Penanganan Data Tidak Seimbang MultiClass," pp. 331–338, 2021.
- [4] V. P. Ramadhan and F. Y. Pamuji, "Jurnal Teknologi dan Manajemen Informatika Analisis Perbandingan Algoritma Forecasting dalam Prediksi Harga Saham LQ45 PT Bank Mandiri Sekuritas (BMRI)," vol. 8, no. 1, pp. 39–45, 2022.
- [5] T. Arifin and S. Syalwah, "Prediksi Keberhasilan Immunotherapy Pada Penyakit Kutil Dengan Menggunakan Algoritma Naïve Bayes," *J. Responsif*, vol. 2, no. 1, pp. 38–43, 2020.
- [6] F. L. D. Cahyanti, W. Gata, and F. Sarasati, "Implementasi Algoritma Naïve Bayes dan K-Nearest Neighbor Dalam Menentukan Tingkat Keberhasilan Immunotherapy Untuk Pengobatan Penyakit Kanker Kulit," *J. Ilm. Univ. Batanghari Jambi*, vol. 21, no. 1, p. 259, 2021, doi: 10.33087/jiubj.v21i1.1189.
- [7] A. Supriyatna and W. P. Mustika, "Komparasi Algoritma Naive bayes dan SVM Untuk Memprediksi Keberhasilan Imunoterapi Pada Penyakit Kutil," *J-SAKTI (Jurnal Sains Komput. dan Inform.*, vol. 2, no. 2, p. 152, 2018, doi: 10.30645/j-sakti.v2i2.78.
- [8] O. P. Barus and T. Sanjaya, "Prediksi Tingkat Keberhasilan Pengobatan Kanker Menggunakan Imunoterapi Dengan Metode Naive Bayes," vol. 5, no. 1, pp. 1–6, 2020.
- [9] B. Harto and Rino, "Comparison of Data Mining Methods Using the Naïve Bayes Algorithm and K Nearest Neighbor in Predicting Immunotherapy Success," *J. Tech-E*, vol. 2, 2019.
- [10] X. Mao, G. Zhao, and R. Sun, "Naive Bayesian algorithm classification model with local attribute weighted based on KNN," Proc. 2017 IEEE 2nd Inf. Technol. Networking, Electron. Autom. Control Conf. ITNEC 2017, vol. 2018-Janua, pp. 904–908, 2018, doi: 10.1109/ITNEC.2017.8284867.