PERBANDINGAN ALGORITMA KNN, C4.5 DAN NAÏVE BAYES UNTUK KLASIFIKASI PENERIMA BANTUAN PROGRAM KELUARGA HARAPAN

La Ode Alyandi, Kariyamin, Samsul Arif
Teknologi Informasi, Institut Teknologi dan Bisnis Muhammadiyah Wakatobi
Jln. Adhyaksa. No. 37 Desa Numana Kec. Wangi-wangi Selatan Kab. Wakatobi
laodealyandi@gmail.com

ABSTRAK

Program Keluarga Harapan (PKH) merupakan bentuk bantuan tunai bersyarat dari pemerintah yang ditujukan kepada masyarakat miskin agar memiliki akses lebih baik terhadap layanan pendidikan dan kesehatan. Namun, tantangan utama dalam pelaksanaan program ini adalah memastikan ketepatan sasaran penerima bantuan. Penelitian ini bertujuan untuk membandingkan performa tiga algoritma klasifikasi K-Nearest Neighbor (KNN), C4.5, dan Naïve Bayes (NB) dalam menentukan kelayakan penerima PKH di Kabupaten Wakatobi. Data penelitian diambil dari Dinas Sosial setempat dengan total 582 entri dan 33 atribut. Metode validasi yang digunakan adalah 7-Fold Cross Validation melalui perangkat lunak RapidMiner. Hasil evaluasi menunjukkan bahwa algoritma Naïve Bayes memberikan akurasi tertinggi sebesar 96,39%, disusul C4.5 dengan 96,22%, dan KNN dengan 81,44%. Temuan ini menunjukkan bahwa Naïve Bayes lebih unggul dalam hal ketepatan klasifikasi, sedangkan C4.5 memiliki keunggulan dalam interpretasi melalui pohon keputusan. Penelitian ini diharapkan dapat menjadi acuan dalam pengambilan keputusan berbasis data untuk meningkatkan efisiensi penyaluran bantuan sosial.

Kata kunci : *PKH, RapidMiner, KNN, C4.5, Naïve Bayes, data mining*

1. PENDAHULUAN

Salah satu bentuk intervensi pemerintah Indonesia dalam upaya mengurangi kemiskinan dan meningkatkan kualitas hidup masyarakat adalah melalui Program Keluarga Harapan (PKH). Program ini dirancang sebagai bantuan tunai bersyarat yang diberikan kepada keluarga yang memenuhi kriteria tertentu, dengan tujuan utama meningkatkan akses terhadap layanan pendidikan, kesehatan, dan kesejahteraan sosial [1].

Selama beberapa tahun pelaksanaannya, PKH telah menunjukkan kontribusi positif dalam penurunan angka kemiskinan dan peningkatan kesejahteraan masyarakat miskin. Namun demikian, salah satu tantangan terbesar dalam implementasi program ini adalah memastikan bahwa bantuan tersebut benar-benar sampai kepada mereka yang paling membutuhkan [2].

Dalam konteks ini, proses klasifikasi penerima manfaat menjadi aspek yang sangat krusial. Keberhasilan program sangat bergantung pada akurasi dalam menentukan calon penerima yang layak, karena hal ini berdampak langsung terhadap efektivitas dan efisiensi penyaluran bantuan. Oleh karena itu, pemanfaatan teknologi informasi dan pendekatan berbasis data menjadi semakin penting untuk membantu proses klasifikasi secara objektif dan sistematis [3].

Secara khusus di Kabupaten Wakatobi, PKH telah dijalankan dengan tujuan utama yang sama, yakni untuk menurunkan tingkat kemiskinan dan memperbaiki kesejahteraan masyarakat. Meski demikian, tantangan pelaksanaan di daerah ini terbilang kompleks, terutama dalam hal ketepatan sasaran. Tidak jarang bantuan diberikan kepada pihak yang kurang layak, sementara keluarga miskin yang sebenarnya membutuhkan justru tidak terdata. Kondisi geografis yang terpencil juga menjadi kendala dalam proses identifikasi keluarga miskin secara menyeluruh [4].

Dalam beberapa tahun terakhir, perkembangan teknologi informasi dan ilmu data telah memberikan peluang baru dalam pengelolaan data sosial. Salah satu pendekatan yang potensial adalah pemanfaatan algoritma pembelajaran mesin (machine learning) untuk melakukan klasifikasi penerima bantuan. Dengan menggunakan data historis dan karakteristik penerima sebelumnya, algoritma ini mampu meningkatkan akurasi dalam proses penentuan penerima yang berhak [5].

Bidang data mining merupakan salah satu cabang ilmu komputer yang secara khusus menangani pengolahan data berskala besar, termasuk proses klasifikasi secara cepat dan tepat. Beberapa metode umum yang digunakan dalam klasifikasi antara lain Decision Tree, Random Rules, Neural Networks (NN), Naive Bayes, dan K-Nearest Neighbor (KNN) [6].

Dalam penelitian ini, digunakan tiga algoritma utama, yaitu K-Nearest Neighbor (KNN), C4.5, dan Naive Bayes. KNN dikenal sebagai algoritma yang sederhana namun cukup efektif karena bekerja dengan mengklasifikasikan data berdasarkan kedekatannya dengan data lain yang telah diketahui kelasnya. Kelebihan utama dari KNN adalah fleksibilitasnya dalam menangani data dengan distribusi kompleks tanpa perlu asumsi khusus tentang distribusi tersebut. Meski demikian, algoritma ini cukup sensitif terhadap ukuran dataset dan membutuhkan komputasi tinggi jika diterapkan pada data berukuran besar [7].

Sementara itu, algoritma C4.5 yang dikembangkan oleh Ross Quinlan dikenal karena kemampuannya membentuk model klasifikasi dalam bentuk pohon keputusan yang mudah diinterpretasikan [8].

C4.5 memanfaatkan pembagian data berdasarkan atribut yang memiliki nilai informasi tertinggi, sehingga membentuk struktur klasifikasi yang logis dan mudah dipahami. Namun, algoritma ini juga rentan terhadap overfitting, terutama ketika pohon keputusan yang dihasilkan terlalu kompleks [9].

Algoritma terakhir yang digunakan adalah Naive Bayes. Metode ini bekerja berdasarkan prinsip probabilistik dengan asumsi bahwa tiap atribut dalam data bersifat independen. Naive Bayes dikenal cepat dan efisien, serta cocok digunakan pada dataset yang besar. Namun, karena sifatnya yang sederhana, akurasinya dapat menurun pada data yang memiliki relasi kompleks antar atribut.

Mengingat pentingnya akurasi dalam menentukan penerima bantuan PKH, pemilihan algoritma klasifikasi yang tepat menjadi hal krusial. Oleh karena itu, penelitian ini difokuskan untuk membandingkan kinerja tiga algoritma populer dalam data mining K-Nearest Neighbor (KNN), C4.5, dan Naïve Bayes dalam mengklasifikasikan kelayakan penerima bantuan. Penelitian ini bertujuan untuk menganalisis performa masing-masing algoritma secara menyeluruh serta mengevaluasi keefektifan dan tingkat akurasi dari ketiganya guna mendukung penyaluran bantuan yang lebih tepat sasaran.

2. TINJAUAN PUSTAKA

2.1. Pennelitian Terdahulu

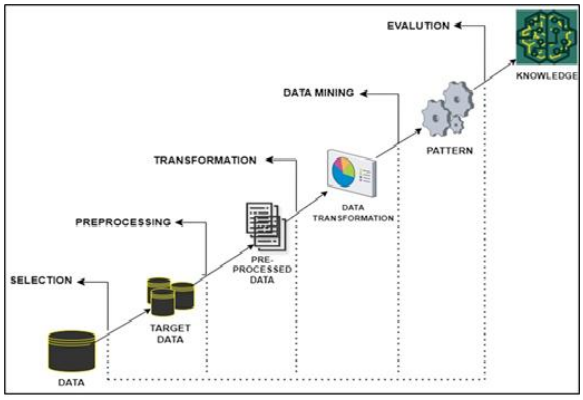
Penelitian sebelumnya membahas upaya peningkatan ketepatan penyaluran bantuan dalam Program Keluarga Harapan (PKH), yang sering kali tidak tepat sasaran. Untuk mengatasi hal tersebut, mereka membangun model klasifikasi penerima bantuan menggunakan tiga algoritma: Naïve Bayes, K-Nearest Neighbor (K=3), dan C4.5, dengan validasi menggunakan 10-Fold Cross Validation. Dari 378 data calon penerima yang terdiri atas 33 atribut, algoritma C4.5 menunjukkan akurasi tertinggi sebesar 80,16%, disusul NBC sebesar 77,51% dan K-NN sebesar 76,72%. Selain itu, C4.5 mampu menyederhanakan model dengan mereduksi atribut menjadi delapan yang paling relevan, seperti jumlah anggota rumah tangga, fasilitas sanitasi, dan jenis dinding rumah [1].

Penelitian terdahulunya yang membandingkan dua algoritma klasifikasi, yakni C4.5 dan Naïve Bayes, untuk mengevaluasi kelayakan penerima bantuan PKH dengan bantuan perangkat lunak RapidMiner. Hasil penelitian menunjukkan bahwa C4.5 memiliki performa lebih baik dengan akurasi 91,25% dan AUC 0,930, sedangkan Naïve Bayes mencatat akurasi 87,11% dan AUC 0,923. Keduanya termasuk dalam kategori klasifikasi "Excellent", namun C4.5 dinilai lebih unggul dalam menentukan kelayakan penerima bantuan [10].

Penelitian lainnya terkait klasifikasi penerima PKH dengan menerapkan metode Naïve Bayes berbasis data mining. Penelitian ini menggunakan data PKH tahun 2020 dari 82 sampel dan mempertimbangkan beberapa kriteria seperti penghasilan, kepemilikan rumah, luas bangunan, serta jenis lantai, atap, dinding, dan sumber air. Hasil pemodelan menunjukkan akurasi sebesar 88%, dengan ketepatan kelas positif mencapai 100% dan kelas negatif 77%. Sementara itu, nilai recall kelas positif sebesar 80% dan kelas negatif 100%, serta f1-score masing-masing 89% (positif) dan 87% (negatif) [11].

2.2. Data Mining dan Knowledge Discovery In Database (KDD)

KDD adalah suatu metode pencarian pengetahuan atau informasi yang belum diketahui dalam suatu database. KDD adalah nama lain dari penambangan data. Meskipun kedua istilah ini sebenarnya memiliki konsep yang berbeda, namun keduanya saling berkaitan dan data mining, salah satu tahapan proses KDD secara keseluruhan, merupakan inti dari proses KDD [12]. Berikut pada Gambar 1 adalah proses KDD.



Gambar 1. Proses KDD

Data mining adalah proses menganalisis kumpulan data besar untuk menemukan pola, tren, dan hubungan tersembunyi. Proses ini tidak hanya mengambil data, tetapi juga membangun model prediktif yang membantu memprediksi perilaku di masa depan. Dengan memanfaatkan teknik statistik, algoritma pembelajaran mesin dan data mining digunakan di berbagai bidang seperti pemasaran, kesehatan, dan deteksi penipuan untuk mendukung pengambilan keputusan yang lebih tepat [13].

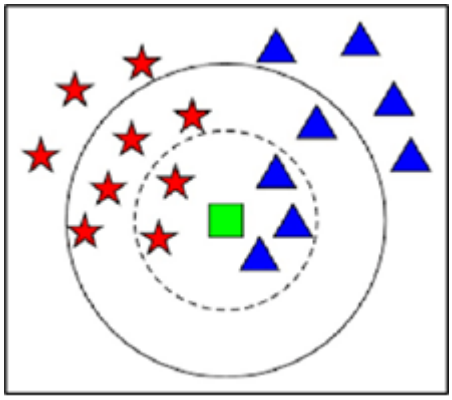
2.3. PKH

Program Keluarga Harapan (PKH) merupakan bantuan tunai yang diberikan kepada Rumah Tangga Sangat Miskin (RTSM) dengan kewajiban memenuhi ketentuan di bidang pendidikan dan kesehatan. Tujuan utama program ini adalah meningkatkan kualitas hidup keluarga miskin melalui akses yang lebih baik terhadap layanan pendidikan, kesehatan, dan sosial. Secara jangka pendek, PKH diharapkan mampu

meringankan beban ekonomi keluarga, sementara dalam jangka panjang, program ini bertujuan memutus rantai kemiskinan dengan membuka peluang peningkatan kualitas hidup masyarakat [10].

2.4. K-Nearest Neighbor

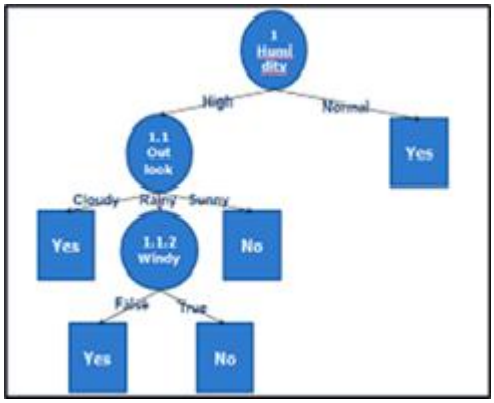
K-Nearest Neighbor (KNN) merupakan salah satu algoritma pembelajaran terawasi (supervised learning) dalam data mining. Algoritma ini bekerja dengan mengukur jarak antara data yang akan diklasifikasikan dan sejumlah K tetangga terdekat dalam data latih. Dalam penelitian ini, nilai K ditetapkan sebanyak 5. Nilai K sendiri menunjukkan jumlah tetangga yang dijadikan acuan dalam proses klasifikasi data baru. Karena bersifat non-parametrik, performa KNN sangat bergantung pada pemilihan nilai K yang tepat [14]. Gambaran algoritma KNN dapat di lihat pada Gambar 2.



Gambar 2. Gambaran KNN

2.5. C4.5

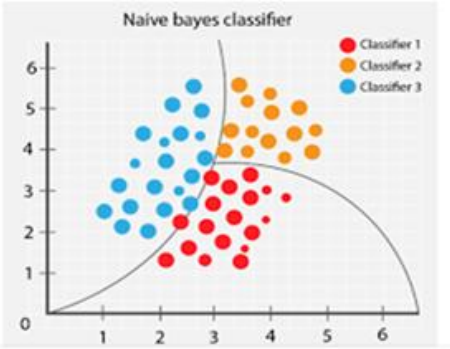
Algoritma C4.5 merupakan pengembangan dari algoritma ID3 yang diperkenalkan oleh Ross Quinlan. Algoritma ini digunakan untuk mengklasifikasikan data yang memiliki atribut numerik maupun kategorikal. C4.5 menjadi salah satu metode paling populer dalam membangun pohon keputusan untuk menyelesaikan berbagai permasalahan klasifikasi [15]. Gambar 3 adalah gambaran umum C4.5.



Gambar 3. Gambaran umum C4.5

2.6. Naïve Bayes

Naive Bayes adalah algoritma klasifikasi yang didasarkan pada Teorema Bayes, dengan asumsi sederhana bahwa setiap fitur dianggap saling independen terhadap kelasnya. Termasuk dalam kelompok klasifikasi berbasis probabilitas, algoritma ini memprediksi kelas suatu data dengan memilih probabilitas tertinggi yang dihitung dari data latih. Secara matematis, Naive Bayes tergolong efisien dan efektif, terutama dalam menangani dataset berukuran besar serta data dengan fitur kategorikal atau berbasis teks. Keunggulannya terletak pada proses perhitungan yang sederhana, kecepatan pelatihan model, dan performa yang cukup kompetitif, sehingga banyak digunakan dalam aplikasi nyata seperti filter spam, analisis sentimen, dan klasifikasi teks lainnya[16]. Gambaran umum algoritma NB dapat di lihat pada Gambar 4.



Gambar 4. Gambaran umum NB

3. METODE PENELITIAN

3.1. Jenis dan Objek Penelitian

Penelitian kuantitatif merupakan metode yang menekankan pada pengumpulan data numerik yang dianalisis menggunakan teknik statistik. Dalam buku *Metode Penelitian Kuantitatif*, dijelaskan bahwa pendekatan ini digunakan untuk menyelidiki fenomena secara sistematis melalui data yang dapat diukur dengan statistik, matematika, atau komputasi. Studi ini mengadopsi pendekatan kuantitatif untuk membandingkan efektivitas algoritma KNN, C4.5, dan Naïve Bayes dalam mengklasifikasikan penerima bantuan Program Keluarga Harapan (PKH). Analisis dilakukan secara mendalam untuk menilai tingkat akurasi, presisi, dan recall masing-masing algoritma, dengan parameter evaluasi yang disesuaikan agar relevan dengan konteks penerima bantuan.

3.2. Studi Literatur

Tinjauan literatur berfungsi sebagai landasan awal untuk menemukan referensi yang relevan dengan topik penelitian. Sumber-sumber tersebut diambil dari buku, jurnal, atau media daring yang mendukung proses penyusunan studi. Fokus utama penelitian mencakup algoritma K-Nearest Neighbor, C4.5, Naïve Bayes, serta pendekatan klasifikasi dan penelitian sejenis yang telah dilakukan sebelumnya.

3.3. Pengumpulan dan Praproses Data

Data dalam penelitian ini diperoleh dari Dinas Sosial Kabupaten Wakatobi, Wangi-Wangi Selatan, berdasarkan data tahun 2023. Pengumpulan data dilakukan melalui wawancara dan pengambilan langsung dari pengurus PKH di kantor Dinas Sosial. Pada tahap praproses, dilakukan pembersihan data dengan menghapus baris yang mengandung nilai null dan menggantinya dengan nilai rata-rata menggunakan operator Filter Example di RapidMiner. Beberapa analisis lanjutan, seperti identifikasi jumlah dan jenis atribut, dilakukan menggunakan Google Colab.

3.4. Pembagian Data Latih dan Data Uji

Pembagian data latih dan data uji dilakukan menggunakan metode K-Fold Cross Validation dengan nilai K = 7, menggunakan operator Cross Validation di RapidMiner. Metode ini digunakan untuk memperkirakan tingkat kesalahan dalam prediksi model, dengan cara membagi seluruh dataset menjadi beberapa bagian untuk pelatihan dan pengujian secara bergantian.

3.5. Pembentukan Model

3.3.1. KNN

Pada algoritma K-NN, dilakukan pengujian dengan menggunakan nilai K = 5. Proses ini memanfaatkan operator K-NN di RapidMiner pada data training, serta operator *apply model* dan *performance* pada data testing. Parameter pengukuran jarak yang digunakan adalah *mixed measure* dengan tipe *Euclidean Distance* untuk menangani atribut numerik.

3.3.2. C4.5

Pada algoritma C4.5, digunakan operator Decision Tree untuk data training, serta operator apply model dan performance untuk data testing di RapidMiner. Parameter pruning dan prepruning diaktifkan guna membantu meningkatkan akurasi hasil klasifikasi.

3.3.3. NB

Pada algoritma Naïve Bayes, digunakan operator Naïve Bayes untuk data training, serta operator apply model dan performance untuk menguji hasil pada data testing di RapidMiner.

3.6. Perbandingan dan Evaluasi Model

Evaluasi setiap model klasifikasi dilakukan menggunakan *confusion matrix* untuk menilai kinerja model. Nilai akurasi, presisi, dan recall juga dihitung dengan bantuan operator *performance* di RapidMiner. Selanjutnya, hasil dari ketiga algoritma (K-NN, C4.5, dan Naïve Bayes) dibandingkan berdasarkan metrik tersebut. Perhitungan akurasi, presisi, dan recall mengacu pada rumus yang ditunjukkan dalam persamaan (1.1), (1.2), dan (1.3).

$$akurasi = \frac{TP+TN}{\text{TP+TN+FP+FN}} \times 100\% \quad (1.1)$$

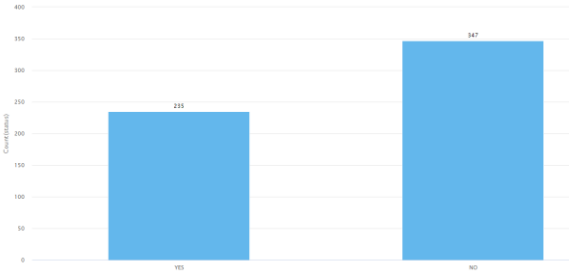
$$presisi = \frac{TP}{\text{TP+FP}} \times 100\% \quad (1.2)$$

$$recall = \frac{TP}{\text{TP+FN}} \times 100\% \quad (1.3)$$

4. HASIL DAN PEMBAHASAN

4.1. Pengambilan Data

Pengambilan data dalam penelitian ini dilakukan melalui dua tahap utama, yaitu wawancara dengan pihak terkait dan pengumpulan dataset PKH. Wawancara dilakukan dengan tenaga pendamping sosial PKH serta pihak Dinas Sosial Kabupaten Wakatobi untuk memahami mekanisme seleksi penerima bantuan dan atribut yang digunakan dalam proses tersebut. Setelah memperoleh pemahaman yang lebih mendalam, penelitian ini mengumpulkan dataset penerima PKH yang mencakup beberapa atribut sosial-ekonomi yang menjadi faktor penentu kelayakan bantuan. Hasil pengumpulan data dapat dilihat pada Gambar 5.

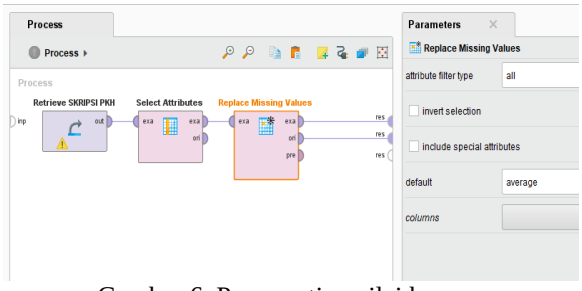


Gambar 5. Perolahan data

Gambar 5 merupakan diagram batang yang menunjukkan jumlah distribusi data sebanyak 582 data berdasarkan dua kategori dalam variabel status, yaitu YES dan NO. Terlihat bahwa jumlah data dengan status NO lebih banyak, yaitu sebanyak 347, dibandingkan dengan data yang berstatus YES yang berjumlah 235 dengan jumlah atribut sebanyak 33 dengan 1 label yakni status penerimaan.

4.2. Pra-Proses Data

Pada tahap praproses data, dilakukan penanganan nilai kosong dengan metode imputasi menggunakan rata-rata untuk menjaga proporsi dan karakteristik data. Selain itu, kolom yang tidak relevan terhadap proses klasifikasi, seperti kolom "NO", dihapus karena hanya berfungsi sebagai penomoran tanpa kontribusi analisis. Gambar 6 adalah pengisian nilai kosong dengan nilai rata-rata.

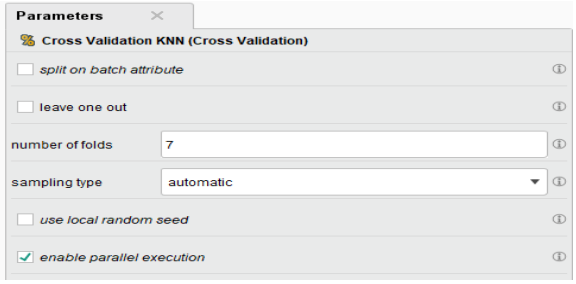


Gambar 6. Penggantian nilai kosong

Gambar 6 merupakan rangkaian proses penggantian nilai kosong ke nilai rata-rata dengan Langkah awal menggunakan fungsi select attributes, yang berfungsi untuk memilh atribut yang memiliki nilai missing atau kosong sehingga bisa di lanjutkan ke fungsi replace missing values dan parameternya menggunakan fungsi average yang bertujuan untuk mengganti nilai kosong ke nilai rata-rata.

4.3. Pembagian Data Latih Dan Data Uji

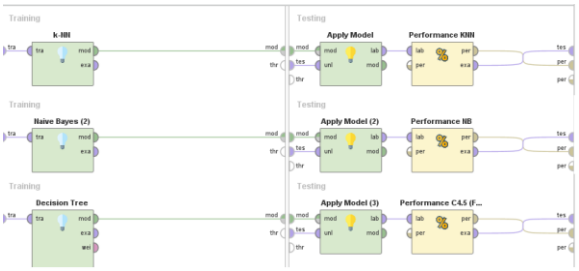
Penelitian ini menggunakan metode K-Fold Cross Validation dengan nilai K = 7 untuk mengevaluasi performa algoritma KNN, C4.5, dan NB. Data dibagi menjadi tujuh bagian, di mana setiap bagian secara bergantian digunakan sebagai data uji, sementara enam bagian lainnya sebagai data latih. Pendekatan ini memastikan bahwa seluruh data digunakan secara merata, sehingga hasil evaluasi model menjadi lebih objektif dan tidak bergantung pada satu skenario pembagian data.



Gambar 7. Penggunaan k-fold

Gambar 7 menampilkan pengaturan parameter untuk proses Cross Validation dalam sebuah alat analisis data, Pada software RapidMiner. Cross Validation adalah teknik evaluasi model dengan membagi data menjadi 7 subset (folds) guna menguji seberapa baik model dapat memprediksi data yang belum pernah dilihat sebelumnya.

4.4. Pembuatan Model



Gambar 8. Pembentukan model algoritma

Setelah melalui tahap praproses, pembangunan model dilakukan menggunakan tiga algoritma klasifikasi populer, yaitu K-Nearest Neighbor (KNN), C4.5, dan Naïve Bayes. KNN mengklasifikasikan data berdasarkan kedekatan terhadap data yang telah diketahui kelasnya, C4.5 membentuk pohon keputusan berdasarkan kontribusi atribut dalam membedakan kelas, sedangkan NB menggunakan pendekatan probabilistik dengan asumsi independensi antar atribut. Penggunaan ketiga algoritma ini memungkinkan perbandingan akurasi untuk menentukan metode yang paling efektif dalam mengklasifikasikan data. Dapat di lihat pada Gambar 8.

4.5. Hasil pembetukan Model

4.5.1. KNN

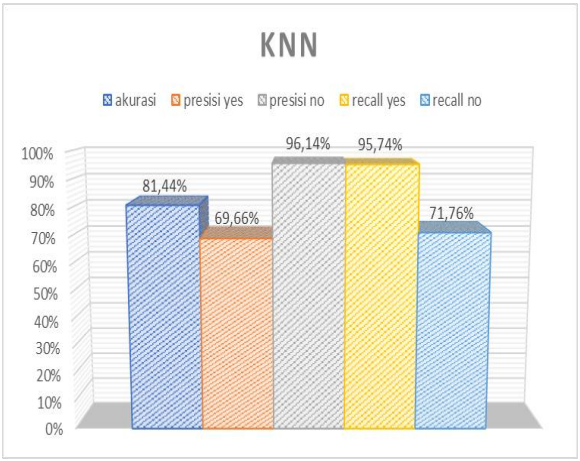
Evaluasi kinerja dari algoritma KNN yang digunakan dalam sebuah *software* rapidminner, yaitu antara kelas "YES" dan "NO". Evaluasi dilakukan menggunakan metrik *confusion matrix* dan indikator performa seperti akurasi, presisi, serta recall. Penilaian ini sangat penting untuk mengetahui seberapa baik model dalam membedakan kedua kelas tersebut. Hasil dari algoritma KNN dapat di lihat pada Gambar 9.

accuracy: 81.44% +/- 5.70% (micro average: 81.44%)

	true YES	true NO	class precision
pred. YES	225	98	69.66%
pred. NO	10	249	96.14%
class recall	95.74%	71.76%	

Gambar 9. Hasil KNN

Algoritma KNN memiliki akurasi sebesar 81.44%, dengan presisi tinggi untuk kelas NO (96.14%) namun lebih rendah untuk kelas YES (69.66%). Recall untuk kelas YES sangat tinggi (95.74%), menunjukkan bahwa model sangat baik dalam mengenali data positif (YES). Namun recall untuk kelas NO masih dapat ditingkatkan. Hal ini mengindikasikan bahwa model lebih sensitif terhadap kelas YES dibandingkan kelas NO. Visualisasi hasil dapat dilihat pada Gambar 10.



Gambar 10. Visual Hasil KNN

4.5.2. C4.5

Hasil evaluasi dari algoritma C4.5, salah satu algoritma decision tree populer, yang digunakan dalam klasifikasi biner antara kelas YES dan NO. Evaluasi dilakukan dengan metode K-Fold Cross Validation serta penerapan Feature Selection (FS) untuk

meningkatkan performa model. Hasilnya disajikan dalam bentuk confusion matrix, akurasi, serta metrik evaluasi lainnya seperti precision dan recall untuk masing-masing kelas. Dapat di lihat pada Gambar 11.

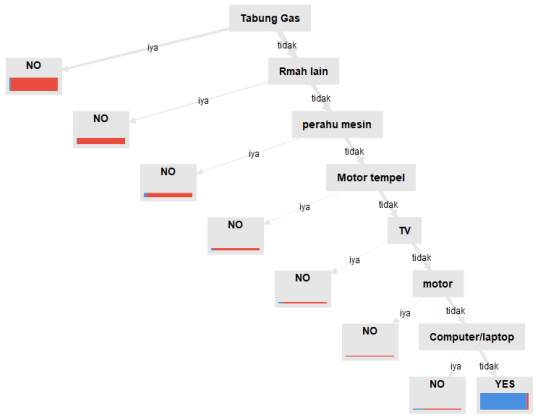
accuracy: 96.22% +/- 2.02% (micro average: 96.22%)

	true YES	true NO	class precision
pred. YES	223	10	95.71%
pred. NO	12	337	96.56%
class recall	94.89%	97.12%	

Gambar 11. Hasil C4.5

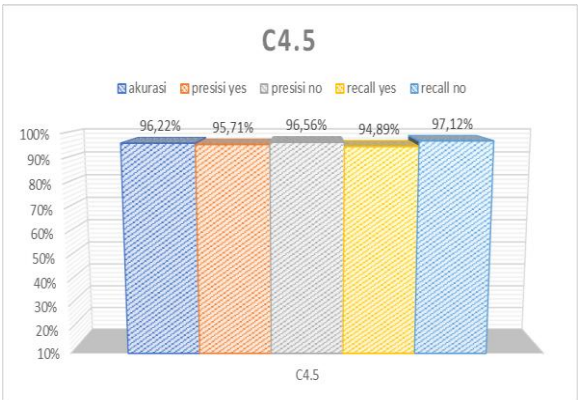
Hasil evaluasi model menunjukkan akurasi sebesar 96,22% ± 2,02%, dengan performa klasifikasi yang sangat baik. Sebanyak 223 data YES dan 337 data NO berhasil diklasifikasikan dengan benar, sementara kesalahan klasifikasi hanya terjadi pada 10 data NO dan 12 data YES. Precision untuk kelas YES dan NO masing-masing sebesar 95,71% dan 96,56%, sedangkan recall-nya mencapai 94,89% untuk YES dan 97,12% untuk NO, yang menunjukkan model mampu mengenali kedua kelas dengan cukup akurat dan seimbang.

Pohon Keputusan yang telah di hasilkan C4.5 dapat di lihat pada Gambar 12. Hasil visualisasi pohon keputusan yang telah melalui proses *feature selection*, yakni pemilihan atribut-atribut paling relevan untuk menentukan kelayakan penerima bantuan.



Gambar 12. Pohon keputusan C4.5

Pohon keputusan pada Gambar 12 menunjukkan bahwa proses pengambilan keputusan benar-benar mempertimbangkan tingkat kepemilikan aset rumah tangga sebagai indikator ekonomi. Semakin sedikit aset yang dimiliki, maka semakin besar kemungkinan seseorang dianggap layak untuk menerima bantuan. Visualisasi algoritma C4.5 terlihat jelas pada Gambar 13.



Gambar 13. Visual Hasil C4.5

4.5.3. NB

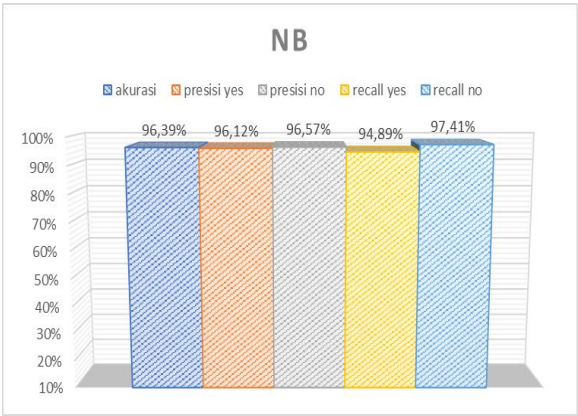
Performa dari algoritma Naive Bayes (NB) dalam menyelesaikan masalah klasifikasi biner (YES dan NO). Evaluasi dilakukan menggunakan confusion matrix, serta metrik performa utama seperti akurasi, presisi, dan recall. Hasil ini sangat penting untuk mengetahui seberapa baik model dalam melakukan klasifikasi terhadap data uji.

accuracy: 96.39% +/- 1.40% (micro average: 96.39%)

	true YES	true NO	class precision
pred. YES	223	9	96.12%
pred. NO	12	338	96.57%
class recall	94.89%	97.41%	

Gambar 14. Hasil NB

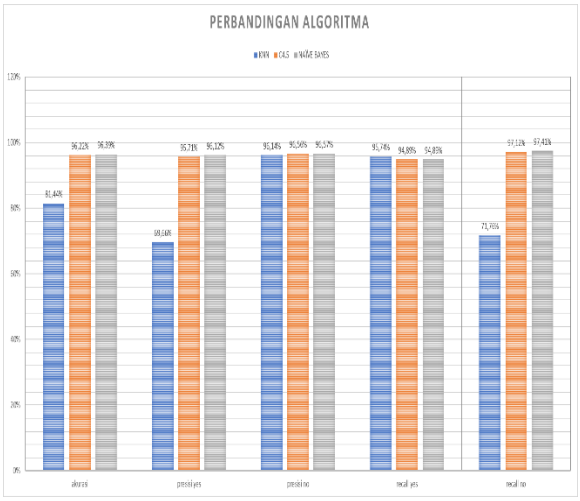
Berdasarkan hasil evaluasi pada Gambar 14, algoritma Naive Bayes menunjukkan performa yang lebih baik dengan akurasi sebesar 96.39%. Presisi untuk kelas YES adalah 96.12% dan recall-nya 94.89%, yang berarti model sangat andal dalam mengidentifikasi dan mengklasifikasikan data YES secara akurat. Untuk kelas NO, presisi 96.57% dan recall 97.41%, menandakan model sangat baik dalam menghindari kesalahan prediksi terhadap data negatif. Secara keseluruhan, model NB ini sangat efektif dalam klasifikasi biner dengan tingkat kesalahan yang sangat rendah. Visualisasi hasil terlihat pada Gambar 15.



Gambar 15. Visual Hasil NB

4.6. Perbandingan Hasil Algoritma

Berdasarkan hasil yang diperoleh, algoritma Naive Bayes menunjukkan performa paling unggul secara keseluruhan. Naive Bayes berhasil mencatatkan akurasi tertinggi sebesar 96,39%, di susul dengan Algoritma C4.5 sebesar 96,22%, dang algoritma KNN dengan nilai akurasi terendah dari dua algoritma lainnya yakni sebesar 81,44%. Visualisasi dari perbandingan dapat di lihat pada Gambar 16.



Gambar 16. Visual perbandingan

5. KESIMPULAN DAN SARAN

Penelitian ini membandingankan tiga algoritma yakni KNN, C4.5 dan NB dalam menganalisis data penerima PKH. Naive Bayes terbukti menjadi algoritma yang memberikan hasil terbaik dengan akurasi sebesar 96,39%. Algoritma C4.5 juga menunjukkan performa yang sangat kompetitif, dengan akurasi yang sama tinggi, yaitu 96,22%. Sedangkan algoritma KNN menghasilkan performa yang paling rendah dibandingkan dua algoritma lainnya. Dengan akurasi sebesar 81,44%. Sebagai saran, penelitian selanjutnya disarankan untuk mencoba menggunakan algoritma klasifikasi lainnya, seperti Random Forest, Support Vector Machine (SVM), atau metode berbasis deep learning, agar dapat memberikan wawasan lebih luas terkait kelebihan dan kekurangan masing-masing pendekatan serta menambahkan jumlah datasetnya agar hasil evaluasi semakin akurat dan bisa menggambarkan performa algoritma dalam situasi nyata yang lebih kompleks.

DAFTAR PUSTAKA

[1] A. Dina, I. Permana, F. Muttakin, and I. Maita, "Perbandingan Algoritma NBC, KNN, dan C4.5 Untuk Klasifikasi Penerima Bantuan Program Keluarga Harapan," *J. Media Inform. Budidarma*, vol. 7, no. 3, pp. 1079–1087, 2023, doi: 10.30865/mib.v7i3.6316.

[2] E. R. Meilaniwati and D. M. Fauzan, "Klasifikasi penduduk miskin penerima PKH menggunakan metode naïve bayes dan KNN Classification," *J. Kaji. dan Terap. Mat.*, vol. 8, no. 2, pp. 75–84, 2022, [Online]. Available: http://journal.student.uny.ac.id/ojs/index.php/jktm:

[3] S. Amaliyah, Jasmir, and S. Rianti, "Penerapan Data Mining Untuk Menentukan Kelompok Prioritas Penerima Bantuan PKH Menggunakan Metode Clustering K-Means Pada Desa Kuala Dendang," *J. Inform. Dan Rekayasa Komputer(JAKAKOM)*, vol. 3, no. 1, pp. 453–458, 2023, doi: 10.33998/jakakom.2023.3.1.802.

[4] S. Nuraeni, H. Harliana, and T. Prabowo, "ANALISIS AKURASI NAÏVE BAYES DAN KNN DALAM PENENTUAN PENERIMA PKH DI LOMBOK UTARA," *J. Inf. Syst. Manag.*, vol. 5, no. 2, pp. 121–126, 2024, doi: 10.24076/joism.2024v5i2.1205.

[5] E. Erwin *et al.*, *Transformasi Digital*, 1st ed., no. June. Kota Jambi: PT. Sonpedia Publishing Indonesia, 2023.

[6] L. O. Alyandi, L. Hadiani, and Irma, "Implementasi Naive Bayes dalam Analisis Sentimen Ulasan Game Honor of Kings di Playstore Implementation of Naive Bayes in Sentiment Analysis of Comments on the," *Bul. Sist. Inf. dan Teknol. Islam*, vol. 6, no. 1, pp. 29–37, 2025, doi: 10.33096/busiti.v6i1.2589.

[7] D. Muti and F. Y. Pamuji, "Analisis Perbandingan Metode Naïve Bayes dan K-NN dalam Klasifikasi Kelayakan Keluarga Terdaftar DTKS Penerimaan Bantuan Sosial di Desa Dubesi," *Semin. Nas. Sist. Inf. ...*, no. September, pp. 3753–3761, 2023, [Online]. Available: https://www.jurnalfti.unmer.ac.id/index.php/senasif/article/view/465%0Ahttps://www.jurnalfti.unmer.ac.id/index.php/senasif/article/download/465/413

[8] Ryanwar, "Penerapan Metode Algoritma C4.5 Untuk Memprediksi Loyalitas Karyawan Pada Pt.Xyz Berbasis Web," *Fak. Sains Dan Teknol. Univ. Buddhi Dharma*, p. 103, 2020.

[9] I. yusuf Akbar, "ANALISIS Perbandingan Metode K-Nn Dan Decision Tree Dalam Klasifikasi Kenyamanan Thermal Bangunan," Universitas Islam Negeri Maulana Malik Ibrahim Malang, 2021. [Online]. Available: http://etheses.uin-malang.ac.id/id/eprint/40087

[10] E. Fitriani, "Perbandingan Algoritma C4.5 Dan Naïve Bayes Untuk Menentukan Kelayakan Penerima Bantuan Program Keluarga Harapan," *Sistemasi*, vol. 9, no. 1, p. 103, 2020, doi: 10.32520/stmsi.v9i1.596.

[11] A. A. A. Arifin, W. Handoko, and Z. Efendi, "Implementasi Metode Naive Bayes Untuk Klasifikasi Penerima Program Keluarga Harapan," *J-Com (Journal Comput.*, vol. 2, no. 1, pp. 21–26, 2022, doi: 10.33330/j-com.v2i1.1577.

[12] A. Yani, Z. Azmi, and D. Suherdi, "Implementasi Data Mining Menganalisa Data Penjualan Menggunakan Algoritma K-Means Clustering,"

J. Sist. Inf. Triguna Dharma (JURSI TGD), vol. 2, no. 2, p. 315, Mar. 2023, doi: 10.53513/jursi.v2i2.6357.

[13] T. M. H. Hasibuan and R. M. A. , Syaiful Zuhri Harahap, "Analisis Perbandingan Algoritma C4.5 Dan Naive Bayes Dalam Menilai Kelayakan Bantuan Program Keluarga Harapan," *INFORMATIKA*, vol. 15, no. 1, pp. 72–86, 2024, doi: 10.25130/sc.24.1.6.

[14] A. Yudhana, S. Sunardi, and A. J. S. Hartanta, "Algoritma K-Nn Dengan Euclidean Distance Untuk Prediksi Hasil Penggergajian Kayu Sengon," *Transmisi*, vol. 22, no. 4, pp. 123–129, 2020, doi: 10.14710/transmisi.22.4.123-129.

[15] R. N. Sipahutar, I. R. Munthe, and A. P. Juledi, "Optimisasi Penilaian Kinerja Karyawan PT. Tolan Tiga Indonesia Estate Perlabian Dengan Algoritma C4.5," *Sport. Cult.*, vol. 15, no. 1, pp. 72–86, 2024, doi: 10.25130/sc.24.1.6.

[16] N. Ifada and R. Nayak, "A New Weighted-learning Approach for Exploiting Data Sparsity in Tag-based Item Recommendation Systems," *Int. J. Intell. Eng. Syst.*, vol. 14, no. 1, pp. 387–399, 2021, doi: 10.22266/IJIES2021.0228.36.