



1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.
2. Penggunaan untuk kepentingan akademik, penelitian, dan pendidikan diperbolehkan dengan mencantumkan sumber.
3. Plagiarisme juga dilarang untuk kepentingan komersial atau pelanggaran hak cipta dapat dikenakan sanksi sesuai dengan UU Hak Cipta di Indonesia.
3. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.

BAB II

TINJAUAN PUSTAKA

Kajian pustaka memiliki peranan yang sangat penting dalam suatu penelitian ilmiah karena berfungsi sebagai landasan teoretis yang memperkuat dasar konseptual maupun metodologis dari penelitian yang dilakukan. Melalui kajian pustaka, peneliti dapat memperoleh pemahaman mendalam mengenai sejauh mana topik penelitian telah dikaji oleh penelitian-penelitian sebelumnya, mengidentifikasi kesenjangan penelitian (research gap), serta menentukan posisi penelitian yang sedang dilakukan dalam konteks keilmuan yang lebih luas. Dengan demikian, penyusunan landasan teori menjadi aspek yang esensial untuk memastikan bahwa penelitian memiliki arah yang terarah, relevan, serta memberikan kontribusi nyata terhadap pengembangan ilmu pengetahuan.

Dalam konteks penyaluran Bantuan Langsung Tunai (BLT), salah satu tantangan signifikan adalah penentuan penerima bantuan yang layak secara objektif dan akurat. Proses penentuan ini sering kali masih dilakukan secara manual dan subjektif, sehingga berpotensi menimbulkan kesalahan sasaran serta ketidakadilan dalam distribusi bantuan. Oleh karena itu, diperlukan pendekatan berbasis data yang dapat mendukung klasifikasi atau prediksi kelayakan penerima BLT.

Salah satu metode dalam bidang data mining yang dapat diterapkan untuk proses klasifikasi adalah algoritma Naive Bayes. Algoritma ini beroperasi berdasarkan prinsip probabilitas yang mengacu pada *Teorema Bayes*, serta memiliki kemampuan untuk memproses data berukuran kecil hingga menengah dengan cepat dan efisien. Algoritma Naive Bayes juga dikenal sebagai metode yang mudah diimplementasikan dan memberikan hasil yang cukup akurat, khususnya

dalam kasus-kasus yang melibatkan data kategorikal seperti data sosial dan ekonomi masyarakat desa.

Bab ini akan mengelaborasi teori-teori yang mendukung penelitian ini, mulai dari konsep dasar Bantuan Langsung Tunai, pengantar data mining dan proses klasifikasi, penjelasan mendalam mengenai algoritma Naive Bayes, serta metode evaluasi kinerja model klasifikasi. Selain itu, bab ini juga mengkaji beberapa penelitian terdahulu yang relevan sebagai bahan perbandingan dan landasan dalam pengembangan penelitian ini. Dengan tinjauan pustaka ini, diharapkan penelitian yang dilakukan memiliki arah yang jelas serta dapat memberikan kontribusi signifikan dalam upaya meningkatkan ketepatan sasaran penyaluran bantuan sosial di Desa Sungai Ambat.

2.1 Kemiskinan

Kemiskinan merupakan kondisi dimana seseorang atau sekelompok orang tidak mampu memenuhi kebutuhan dasar lain nya secara layak, seperti kebutuhan pangan, sandang, papan , pendidikan, dan kesehatan[7]. Menurut Badan Pusat Statistik (BPS), Kemiskinan diartikan sebagai ketidakmampuan dari sisi ekonomi untuk memenuhi kebutuhan dasar yang diukur dari sisi pengeluaran. Kemiskinan bukan hanya disebabkan oleh faktor rendahnya pendapatan, tetapi juga oleh rendahnya akses terhadap pendidikan, pekerjaan, serta sarana dan prasarana ekonomi. Oleh karena itu, penanganan kemiskinan memerlukan pendekatan yang menyeluruh, baik dari sisi kebijakan ekonomi, sosial, maupun teknologi.

Pemerintah indonesia telah melakukan berbagai upaya untuk menanggulangi kemiskinan, salah satunya dengan memberikan bantuan sosial kepada masyarakat miskin dalam bentuk bantuan langsung tunai (BLT). Program ini bertujuan





1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.
2. Penggunaan tanpa izin untuk kepentingan akademik, penelitian, dan pendidikan diperbolehkan dengan mencantumkan sumber. Plagiarisme juga dilarang dan dapat dikenakan sanksi.
3. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.

agar masyarakat miskin tetap dapat memenuhi kebutuhan hidup sehari-hari serta menjaga stabilitas ekonomi nasional ditengah kondisi kritis atau inflasi harga barang kebutuhan pokok.

2.2 Bantuan Langsung Tunai (BLT)

Bantuan Langsung Tunai (BLT) merupakan salah satu bentuk program kebijakan sosial yang dirancang oleh pemerintah untuk memberikan bantuan berupa uang tunai secara langsung kepada masyarakat berpenghasilan rendah atau kelompok rentan[8]. Program ini bertujuan untuk meningkatkan daya beli masyarakat serta mengurangi beban ekonomi yang timbul akibat berbagai kondisi yang memengaruhi kestabilan ekonomi, seperti krisis ekonomi, kenaikan harga bahan pokok, maupun bencana alam. Melalui penyaluran bantuan secara langsung, BLT diharapkan dapat memberikan dampak nyata terhadap kesejahteraan masyarakat dengan cara memperkuat daya tahan ekonomi rumah tangga miskin dan menjaga stabilitas sosial di tengah tekanan ekonomi nasional.

Bantuan Langsung Tunai (BLT) merupakan program bantuan sosial yang diselenggarakan oleh pemerintah melalui transfer langsung dana tunai kepada individu atau keluarga yang memenuhi kriteria spesifik[9]. Program ini dirancang untuk menyediakan dukungan finansial yang cepat dan adaptif bagi kelompok masyarakat yang terpengaruh oleh krisis ekonomi, termasuk inflasi, pandemi, atau peningkatan harga bahan pokok. Di Indonesia, BLT umumnya dihubungkan dengan inisiatif seperti BLT BBM (Bahan Bakar Minyak) atau BLT Desa, yang dikelola oleh Kementerian Sosial atau entitas terkait[10].

Dalam konteks jaminan sosial, BLT berfungsi sebagai bentuk perlindungan sosial yang memungkinkan pemerintah memenuhi tanggung jawabnya dalam



mendukung masyarakat yang membutuhkan. BLT tidak hanya memberikan bantuan finansial sementara, tetapi juga berperan sebagai instrumen dalam mengurangi kemiskinan dan meningkatkan kesejahteraan masyarakat, terutama pada masa krisis. Dengan demikian, program ini secara langsung berkontribusi pada upaya pencapaian Tujuan meningkatkan kesejahteraan masyarakat miskin. Mengurangi beban ekonomi keluarga penghasilan rendah, memastikan pemerataan bantuan sosial agar tepat sasaran.

2.3 Data Mining

Data mining adalah proses menemukan pola atau informasi menarik dari data dalam jumlah besar yang disimpan dalam database, data warehouse, atau tempat penyimpanan lainnya[11]. Data mining menggabungkan teknik dari berbagai bidang seperti statistik, pembelajaran mesin, kecerdasan buatan, dan basis data untuk mengekstraksi informasi yang bermanfaat dan relevan secara otomatis atau semiotomatis. Tujuannya adalah untuk menemukan pola tersembunyi yang dapat mendukung pengambilan keputusan strategis dan memberikan nilai tambah ekonomi. Proses ini merupakan inti dari *Knowledge Discovery in Databases (KDD)*, yang melibatkan tahapan seleksi data, pre-processing, transformasi, data mining, dan interpretasi hasil[12].

Data mining merupakan proses penemuan pola atau informasi yang berguna dari kumpulan data yang besar. Data mining adalah proses mengekstraksi pengetahuan yang sebelumnya tidak diketahui, namun berpotensi berguna, dari data yang tersimpan dalam basis data. Tujuan utama data mining adalah untuk mengidentifikasi pola tersembunyi dan hubungan antara variabel dalam suatu dataset. Beberapa tahapan dalam proses data mining meliputi:

1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.
2. Penggunaan tanpa izin untuk kepentingan akademik, penelitian, dan pendidikan diperbolehkan dengan mencantumkan sumber. Plagiarisme juga dilarang dan dapat dikenakan sanksi.
3. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.



1. Pembersihan data (data cleaning) : menghilangkan data yang tidak relevan atau mengandung kesalahan.
2. Integrasi data (data integration) : menggabungkan data dari berbagai sumber.
3. Seleksi data (data selection) : memilih data yang relevan untuk analisis.
4. Transformasi data (data transformation) : mengubah data menjadi format yang sesuai untuk pemrosesan.
5. Proses mining (data mining process) : penerapan algoritma untuk menemukan pola.
6. Evaluasi dan interpretasi (pattern evaluation) : menilai pola yang ditemukan agar bermanfaat bagi pengambilan keputusan.

Salah satu cabang dari data mining adalah klasifikasi, yaitu proses memprediksi katagori atau kelas dari suatu data berdasarkan atribut yang ada. Dalam kontes penelitian ini, klasifikasi digunakan untuk menentukan apakah seseorang layak atau tidak layak menerima BLT berdasarkan data sosial ekonominya.

2.4 Klasifikasi dalam Data Mining

Klasifikasi merupakan salah satu teknik utama dalam data mining yang digunakan untuk mengelompokkan atau memprediksi suatu data ke dalam kelas atau kategori tertentu berdasarkan pola yang telah dipelajari dari data sebelumnya[13]. Proses klasifikasi dilakukan dengan cara membangun model prediktif yang diperoleh dari data latih (training data), di mana setiap data memiliki atribut-atribut tertentu (fitur) dan label kelas yang telah diketahui. Model yang dihasilkan kemudian digunakan untuk menentukan kelas dari data baru yang belum diketahui kategorinya.



Klasifikasi adalah proses menemukan model atau fungsi yang menggambarkan dan membedakan kelas data atau konsep, dengan tujuan untuk dapat memprediksi label kelas dari objek yang belum diketahui kelasnya. Model klasifikasi ini dapat direpresentasikan dalam berbagai bentuk, seperti aturan klasifikasi (classification rules), pohon keputusan (decision tree), rumus matematis, atau jaringan probabilistik seperti Naive Bayes.

Dalam konteks data mining, proses klasifikasi terdiri atas dua tahap utama, yaitu:

1. Tahap pembelajaran (learning phase) : sistem mempelajari hubungan antara atribut-atribut dalam data dengan label kelas yang telah diketahui, untuk membangun model klasifikasi.
2. Tahap pengujian (testing phase) : model yang telah dibangun diuji menggunakan data uji (testing data) untuk mengukur akurasi dan kemampuan model dalam memprediksi kelas data baru.

Klasifikasi banyak digunakan dalam berbagai bidang, seperti prediksi kelayakan kredit, diagnosis medis, pengenalan pola, dan termasuk dalam penentuan kelayakan penerima Bantuan Langsung Tunai (BLT). Melalui klasifikasi, sistem dapat menganalisis faktor-faktor seperti pendapatan, pekerjaan, jumlah tanggungan, dan kondisi rumah, kemudian mengkategorikan setiap warga ke dalam kelompok layak atau tidak layak menerima bantuan sosial secara objektif dan terukur.

Dengan demikian, klasifikasi dalam data mining berperan penting dalam mendukung proses pengambilan keputusan berbasis data (data-driven decision making), karena mampu mengekstraksi pola tersembunyi dari data historis dan

menggunakannya untuk memprediksi keputusan di masa depan dengan tingkat akurasi yang dapat diukur.

2.5 Algoritma Naive Bayes

Algoritma Naive Bayes merupakan salah satu metode klasifikasi dalam data mining yang berlandaskan pada teori probabilitas Bayes. Metode ini digunakan untuk memprediksi kelas atau kategori dari suatu data berdasarkan nilai probabilitas tertinggi yang dihitung dari data pelatihan (training data). Disebut naive karena algoritma ini mengasumsikan bahwa setiap fitur atau atribut dalam data bersifat independen secara kondisional terhadap fitur lainnya, meskipun dalam kenyataannya hal tersebut tidak selalu benar[14]. Algoritma ini menggunakan *teorema Bayes* dengan asumsi bahwa setiap fitur (atribut) bersifat independen terhadap fitur lainnya.

Meskipun demikian, model ini tetap terbukti mampu memberikan performa klasifikasi yang baik dalam berbagai aplikasi nyata. Naive Bayes bekerja dengan menghitung probabilitas prior dari setiap kelas serta probabilitas likelihood dari masing-masing atribut terhadap kelas tersebut. Selanjutnya, probabilitas posterior dihitung menggunakan Teorema Bayes, dan kelas dengan nilai probabilitas terbesar akan dipilih sebagai hasil prediksi.

Pendekatan probabilistik ini menjadikan Naive Bayes sebagai algoritma yang sederhana namun sangat efisien, karena tidak memerlukan proses pelatihan yang kompleks. Selain itu, Naive Bayes mampu bekerja optimal pada data berukuran besar maupun data dengan fitur yang beragam, terutama ketika fitur-fitur tersebut berupa kategori. Kecepatan komputasi, ketahanan terhadap noise, dan kemampuan dalam menangani data berdimensi tinggi menjadikan algoritma ini





banyak digunakan dalam berbagai bidang seperti deteksi spam, text mining, analisis sentimen, serta klasifikasi data sosial ekonomi. Dengan karakteristik tersebut, Naive Bayes menjadi salah satu metode klasifikasi yang sering dipilih dalam penelitian maupun implementasi sistem pendukung keputusan yang memerlukan proses klasifikasi cepat dan hasil yang konsisten.

Kelebihan Naive Bayes:

1. Cepat dan efisien untuk dataset besar.
2. Dapat bekerja dengan baik pada data kategori.
3. Tidak memerlukan parameter kompleks.

Kelemahan Naive Bayes:

1. Asumsi independensi antar atribut sering tidak sesuai dengan kenyataan.
2. Tidak cocok untuk data dengan hubungan antar variabel yang kuat.

Algoritma Naive Bayes banyak digunakan karena memiliki kecepatan tinggi dalam proses pelatihan dan prediksi, serta kinerja yang baik pada data berukuran besar. Selain itu, metode ini efektif pada data dengan variabel kategori, seperti data sosial ekonomi untuk penentuan penerima bantuan sosial[15]. Meskipun demikian, algoritma ini memiliki kelemahan utama pada asumsi independensi antar atribut, yang dapat mengurangi akurasi jika antar variabel terdapat hubungan yang kuat. Oleh karena itu, dalam beberapa penelitian, Naive Bayes sering dikombinasikan dengan metode lain untuk meningkatkan performa, seperti feature selection atau ensemble learning[16].

2.6 Algoritma K-Nearest Neighbor(KNN)

K-Nearest Neighbor (K-NN) merupakan salah satu algoritma dalam bidang *machine learning* yang termasuk ke dalam metode supervised learning, yaitu



metode pembelajaran yang menggunakan data berlabel sebagai dasar untuk melakukan proses prediksi atau klasifikasi terhadap data baru[17]. Algoritma ini bekerja dengan prinsip bahwa suatu objek atau data yang memiliki karakteristik serupa cenderung berada pada kelompok atau kelas yang sama. Oleh karena itu, proses penentuan kelas suatu data dilakukan berdasarkan kedekatannya dengan sejumlah data tetangga terdekat yang terdapat pada data pelatihan (*training data*).

Dalam implementasinya, pengukuran kedekatan antar data umumnya menggunakan beberapa metode perhitungan jarak, seperti Euclidean Distance, Manhattan Distance, dan Minkowski Distance. Di antara metode tersebut, Euclidean Distance merupakan metode yang paling sering digunakan karena mampu mengukur jarak lurus antara dua titik dalam ruang multidimensi[18]. Semakin kecil nilai jarak yang dihasilkan, maka semakin tinggi tingkat kemiripan antara data yang dibandingkan. K-Nearest Neighbor termasuk algoritma yang sederhana namun efektif karena tidak memerlukan proses pembentukan model yang kompleks. Seluruh data pelatihan disimpan dan digunakan secara langsung pada saat proses prediksi dilakukan. Oleh karena itu, algoritma ini sering digunakan dalam berbagai bidang seperti klasifikasi data, pengenalan pola, sistem rekomendasi, data mining, dan analisis prediktif[19].

Keunggulan utama K-NN terletak pada kemudahan implementasi dan kemampuannya menghasilkan tingkat akurasi yang baik pada dataset yang sesuai. Namun, performa algoritma ini sangat dipengaruhi oleh pemilihan nilai k , ukuran dataset, serta metode pengukuran jarak yang digunakan. Algoritma K-NN dapat didefinisikan sebagai metode klasifikasi non-parametrik yang menentukan kategori suatu objek berdasarkan mayoritas kelas dari sejumlah tetangga terdekatnya dalam

1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.
2. Penggunaan tanpa izin untuk kepentingan akademik, penelitian, dan pendidikan diperbolehkan dengan mencantumkan sumber. Plagiarisme juga dilarang dan dapat dikenakan sanksi.
3. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.



ruang fitur (*feature space*). Karena kesederhanaan dan efektivitasnya, algoritma ini masih menjadi salah satu metode yang banyak digunakan sebagai dasar perbandingan dalam penelitian-penelitian di bidang *data mining*, *machine learning*, dan *artificial intelligence*.

2.7 Google Colab

Google Colab (Google Colaboratory) adalah sebuah layanan berbasis komputasi awan (*cloud computing*) yang dikembangkan oleh Google untuk menyediakan lingkungan pemrograman interaktif berbasis *Jupyter Notebook*[20]. Platform ini memungkinkan pengguna untuk menulis, menjalankan, dan mendokumentasikan kode program, khususnya bahasa Python, secara langsung melalui peramban (*browser*) tanpa memerlukan instalasi perangkat lunak tambahan pada komputer lokal[21]. Dengan memanfaatkan infrastruktur komputasi berbasis awan milik google, pengguna dapat menjalankan berbagai proses komputasi dengan lebih efisien serta mengakses proyek yang sedang dikerjakan dari berbagai perangkat yang memiliki koneksi internet.

Google Colab berfungsi sebagai sarana yang mendukung kegiatan pembelajaran, penelitian, analisis data, serta pengembangan Perangkat lunak, implementasi berbagai metode dalam bidang *machine learning*, *deep learning*, *data mining* dan *artificial intelligence* (AI)[22]. Platform ini menyediakan berbagai pustaka (*library*) Python yang umum digunakan dalam bidang ilmu data, seperti NumPy, Pandas, Matplotlib, TensorFlow, dan Scikit-learn, sehingga memudahkan pengguna dalam melakukan pengolahan data dan eksperimen komputasi. Salah satu keunggulan utama Google Colab adalah kemampuannya menyediakan akses ke sumber daya komputasi berperforma tinggi, seperti Graphics Processing Unit

1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.
2. Penggunaan tanpa izin untuk kepentingan akademik, penelitian, dan pendidikan diperbolehkan dengan mencantumkan sumber. Plagiarisme juga dilarang dan dapat dikenakan sanksi.
3. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.



1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.
2. Penggunaan untuk kepentingan akademik, penelitian, dan pendidikan diperbolehkan dengan mencantumkan sumber.
3. Plagiarisme juga dilarang untuk kepentingan komersial atau pelanggaran hak cipta dapat dikenakan sanksi sesuai dengan UU Hak Cipta di Indonesia.
3. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.

(GPU) dan Tensor Processing Unit (TPU), yang dapat digunakan untuk mempercepat proses komputasi tanpa memerlukan perangkat keras khusus[23]. Selain itu, Google Colab terintegrasi dengan Google Drive sehingga memudahkan pengguna dalam menyimpan, mengelola, dan berbagi dokumen notebook dan dataset secara daring. Dengan berbagai fitur yang dimilikinya, Google Colab menjadi platform yang efektif untuk mendukung kegiatan akademik maupun profesional, terutama dalam bidang pemrograman, data science, data mining, dan kecerdasan buatan. Kemudahan akses, fleksibilitas penggunaan, serta dukungan impor data dari berbagai sumber, seperti file lokal, Google Drive, GitHub, maupun repositori data lainnya. Google Colab sebagai salah satu alat yang banyak digunakan dalam lingkungan pendidikan dan penelitian modern[24].

2.8 Penelitian Terdahulu

Berdasarkan hasil review jurnal terdahulu, maka dapat perbandingan sebagaimana ditampilkan pada Tabel 2.1 di bawah ini:

Tabel 2.1 Riview Jurnal Terdahulu

No	Judul, Penulis, Dan Tahun	Hasil Penelitian
1	Sistem Klasifikasi Penyakit Diabetes Mellitus Menggunakan Metode K-Nearest Neighbor (K-NN). Fitri Yunita (2016)	Hasil dari penelitian ini yang telah dilakukan, algoritma K-Nearest Neighbor (K-NN) mampu digunakan untuk mengklasifikasikan data dengan baik berdasarkan beberapa variabel seperti usia, berat badan, tinggi badan, riwayat keluarga, gangguan destruksi sel beta, dan pola makan. Hasil pengujian menunjukkan bahwa tingkat akurasi yang diperoleh cukup tinggi, yaitu mencapai 96%, sehingga metode ini tergolong efektif dalam proses klasifikasi. Selain itu, jumlah data latih (training data) sangat berpengaruh terhadap tingkat akurasi yang dihasilkan. Semakin banyak data latih yang digunakan, maka semakin tinggi pula tingkat akurasi klasifikasi.

Tabel 2.1 Riview Jurnal Terdahulu Lanjutan

No	Judul, Penulis, Dan Tahun	Hasil Penelitian
		Dengan demikian, dapat disimpulkan bahwa metode K-Nearest Neighbor (K-NN) merupakan metode yang cukup baik dan layak digunakan dalam menyelesaikan permasalahan klasifikasi data.
2	Penentuan Kelayakan Masyarakat Miskin Penerima Bantuan Menggunakan Metode Naive Bayes (Studi Kasus: Kabupaten Penajam Paser Utara) Nur Madia, Anindita Septiarini, Heliza Rahmania Hatta, Hamdani Hamdani, Masna Wati (2023)	Implementasi dan hasil pengujian pada penelitian penentuan kelayakan masyarakat miskin penerima bantuan dapat diambil beberapa kesimpulan yaitu metode Naïve Bayes dapat diterapkan pada sistem penentuan kelayakan masyarakat miskin penerima bantuan karena mendapatkan accuracy yang cukup baik, dan berdasarkan hasil pengujian menunjukkan bahwa metode Naïve Bayes dapat menghasilkan accuracy tertinggi pada scenario ketiga yaitu 60% atau 328 data latih dan 40% atau 218 data uji dengan nilai accuracy sebanyak 77,98%. Pada penelitian lebih lanjut dapat dikembangkan dengan menggunakan Pada penelitian lebih lanjut dapat dikembangkan dengan menggunakan metode lain seperti Artificial Neural Network, Support Vector Machine, atau K-Nearest Neighbor untuk meningkatkan nilai akurasi. Kelebihan dari penelitian ini adalah pendekatan metodologis yang sistematis, penggunaan data resmi dan relevan, evaluasi performan yang lengkap, jumlah variabel yang komprehensif. Kekurangan dari penelitian ini adalah nilai akurasi relatif rendah, tidak dilakukan perbandingan dengan metode lain. minim visualisasi dan analisis hasil lebih lanjut, tidak ada uji signifikasi atau validasi silang (cross-validation).
3	Implementasi Algoritma Naive Bayes untuk Sistem Pendukung	Hasil dari penelitian ini yaitu berhasil mengimplementasikan algoritma Naive Bayes sebagai sistem pendukung keputusan untuk menentukan kelayakan penerima bantuan pendidikan. Algoritma Naive Bayes mampu mengklasifikasikan siswa berdasarkan kriteria.



Hak Cipta Dilindungi Undang-Undang

Universitas Islam Indragiri

1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.
 2. Penggunaan untuk kepentingan akademik, penelitian, dan pendidikan diperbolehkan dengan mencantumkan sumber.
 3. Penggunaan tanpa izin untuk kepentingan komersial atau pelanggaran hak cipta dapat dikenakan sanksi sesuai dengan UU Hak Cipta di Indonesia.
- Plagiarisme juga dilarang dan dapat dikenakan sanksi.
3. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.

Tabel 2.1 Riview Jurnal Terdahulu Lanjutan

No	Judul, Penulis, Dan Tahun	Hasil Penelitian
	Penerima Bantuan Pendidikan Chandra Nugroho, Tamriesfatno, Sulkifli(2025). Wisnu Sigrid dan	Algoritma ini dapat membantu pihak terkait dalam membuat keputusan yang lebih objektif dan terukur. Kelebihan dari penelitian ini adalah Tujuan penelitian jelas dan kontekstual, Pemanfaatan algoritma Naive Bayes yang sesuai, Data dan variabel penelitian relevan, Hasil penelitian mudah diinterpretasikan. Kekurangan dari penelitian ini adalah Jumlah data terbatas (hanya 60 data latih dan 10 data uji), Tidak dilakukan pengujian akurasi kuantitatif, Tidak ada perbandingan metode lain, Validasi model kurang komprehensif.
4	Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako Amat Damuri, Umbar Riyanto, Hengki Rusdianto, Mohammad Aminudin (2021).	Berdasarkan penelitian yang dilakukan oleh Amat Damuri, dkk (2021), diterapkan algoritma Naive Bayes untuk melakukan klasifikasi kelayakan penerima bantuan sembako. Penelitian ini menggunakan total 175 data, yang terdiri dari 135 data training dan 40 data testing. Hasil evaluasi menggunakan confusion matrix didapatkan akurasi yang dihasilkan untuk 135 data training dengan 40 data testing dan tujuh atribut yang digunakan menghasilkan akurasi sebesar 86%, recall 85%, dan presisi 88%. Akurasi dapat dipengaruhi oleh beberapa faktor, diantaranya: jumlah data training, data testing dan atribut yang digunakan. Kelebihan dari penelitian ini adalah Metode penelitian jelas dan sistematis, Menggunakan data riil dan Hasil evaluasi model cukup tinggi. Kekurangan dari penelitian ini adalah Ukuran dataset terbatas, Tidak ada perbandingan algoritma, Minim analisis faktor Menggunakan data riil dan Hasil evaluasi model cukup tinggi. Kekurangan dari penelitian ini adalah Ukuran dataset terbatas, Tidak ada perbandingan algoritma, Minim analisis faktor dominan, Kurangnya visualisasi hasil pengujian, Validasi model sederhana.



Hak Cipta Dilindungi Undang-Undang

Universitas Islam Indragiri

1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.
 2. Penggunaan untuk kepentingan akademik, penelitian, dan pendidikan diperbolehkan dengan mencantumkan sumber.
 3. Plagiarisme juga dilarang untuk kepentingan komersial atau pelanggaran hak cipta dapat dikenakan sanksi sesuai dengan UU Hak Cipta di Indonesia.
- Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.

Tabel 2.1 Riwiew Jurnal Terdahulu Lanjutan

No	Judul, Penulis, Dan Tahun	Hasil Penelitian
5	Optimalisasi Pengalokasian Bantuan Sosial di Kota Banjarmasin Menggunakan Pendekatan Naive Bayes Muhammad Iqbal Adha, Finki Dona Marleny, Ayu Ahadi Ningrum (2025)	Hasil penelitian menunjukkan bahwa model Naive Bayes mencapai akurasi 95% pada data pengujian dan seluruh data, dengan distribusi prioritas sebagai berikut: Prioritas Tinggi (46,2%), Prioritas Sedang (29,3%), dan Prioritas Rendah (24,5%). Kecamatan Banjarmasin Selatan, Banjarmasin Barat, dan Banjarmasin Utara diidentifikasi sebagai wilayah dengan jumlah prioritas tinggi terbanyak. Implikasi praktis dari penelitian ini adalah rekomendasi alokasi bantuan yang lebih terarah, strategi distribusi cepat, dan program pendukung seperti pelatihan keterampilan untuk kelompok prioritas sedang. Kesimpulan penelitian menegaskan bahwa pendekatan ini efektif untuk alokasi bantuan yang efisien, namun evaluasi berkala diperlukan untuk memastikan relevansi model di masa depan. Kelebihan dari penelitian ini adalah Pendekatan Metodologis yang Jelas dan Sistematis, Akurasi Model Sangat Tinggi (95%), Proses Preprocessing Data yang Lengkap, Visualisasi Data Informatif, Relevansi Sosial yang Kuat. Kekurangan dari penelitian ini adalah Ketergantungan pada Data Statis Tidak Menggunakan Penyeimbangan Data (SMOTE), Variabel Terbatas.
6	Perbandingan Algoritma Naïve Bayes, C4.5, dan K-Nearest Neighbor untuk Klasifikasi Kelayakan Program Keluarga Harapan Putri Ramadani1, Riszki Fadillah, Quratih Adawiyah, Suerni, Baginda	Hasil penelitian menunjukkan bahwa algoritma C4.5 memberikan kinerja terbaik dengan akurasi 91,8%, presisi 90,7%, recall 92,3%, dan AUC 0,944. Naive Bayes mencatat akurasi 87,2% dan recall 88,9%, sedangkan K-NN menghasilkan akurasi 89,9% dan recall 91,1%, namun memerlukan komputasi lebih tinggi. Temuan ini menunjukkan bahwa C4.5 lebih efektif dalam mengklasifikasikan kelayakan penerima PKH secara akurat dan efisien. Penelitian ini menegaskan potensi algoritma machine learning dalam mendukung pengambilan keputusan pada



Hak Cipta Dilindungi Undang-Undang

Universitas Islam Indragiri

1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.
2. Penggunaan untuk kepentingan akademik, penelitian, dan pendidikan diperbolehkan dengan mencantumkan sumber.
3. Penggunaan tanpa izin untuk kepentingan komersial atau pelanggaran hak cipta dapat dikenakan sanksi.
4. Plagiarisme juga dilarang dan dapat dikenakan sanksi.
5. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.

Tabel 2.1 Riview Jurnal Terdahulu Lanjutan

No	Judul, Penulis, Dan Tahun	Hasil Penelitian
	RestuAl Ghazali(2024)	program bantuan sosial. Studi lanjutan disarankan untuk memperluas cakupan data dan mengeksplorasi metode klasifikasi lainnya guna optimalisasi distribusi bantuan. Kelebihan dari penelitian ini adalah Menggunakan Data Lokal dan Aktual, Membandingkan Tiga Algoritma Populer, Evaluasi Model Lengkap, Visualisasi dan Implementasi Praktis, Metodologi Tersusun Baik. Kekurangan dari penelitian ini adalah Jumlah Data Terbatas (109 entri), Tidak Ada Seleksi Fitur Otomatis / Parameter Tuning, Konteks Wilayah Terbatas.
7	Implementasi Algoritma K-Nearest Neighbor Dalam Menentukan Penerima Bantuan Langsung Tunai. Anwar Pauji, Siti Aisyah, Agus Surip, Riko Saputra4, Irfan Ali(2020)	Penelitian yang dilakukan oleh Anwar Pauji dan rekan-rekan ini berfokus pada implementasi algoritma K-Nearest Neighbor (K-NN) dalam proses klasifikasi penerima Bantuan Langsung Tunai (BLT) di Desa Ciperna, Kabupaten Cirebon. Penelitian ini dilatarbelakangi oleh permasalahan yang sering muncul dalam penyaluran bantuan sosial, di mana banyak penerima yang tidak tepat sasaran. Hal tersebut terjadi akibat kurang optimalnya proses penetapan status keluarga miskin sebagai penerima manfaat. Oleh karena itu, peneliti berupaya mengembangkan model klasifikasi berbasis data mining dengan algoritma K-NN agar proses distribusi bantuan dapat lebih objektif, transparan, dan akurat. Kelebihan dari penelitian ini adalah Pendekatan sederhana dan mudah diimplementasikan, Dapat membantu pemerintah desa dalam pengambilan keputusan, Penggunaan atribut yang relevan. Kekurangan dari penelitian ini adalah Tingkat akurasi masih rendah (68,82%), Ketidakseimbangan data antar kelas, Keterbatasan variabel penelitian.



Hak Cipta Dilindungi Undang-Undang

Universitas Islam Indragiri

1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.
 2. Penggunaan untuk kepentingan akademik, penelitian, dan pendidikan diperbolehkan dengan mencantumkan sumber.
 3. Penggunaan tanpa izin untuk kepentingan komersial atau pelanggaran hak cipta dapat dikenakan sanksi sesuai dengan UU Hak Cipta di Indonesia.
- Plagiarisme juga dilarang dan dapat dikenakan sanksi.
3. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.

Tabel 2.1 Riview Jurnal Terdahulu Lanjutan

No	Judul, Penulis, Dan Tahun	Hasil Penelitian
8	Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Warga Penerima Bantuan Sosial Pajar Pahrudin, Kusno Harianto(2022)	Hasil dari penelitian ini adalah algoritma K-Nearest Neighbor (K-NN) terbukti efektif dalam mengklasifikasikan warga yang layak menerima bantuan sosial. Dengan memanfaatkan metode data mining, penelitian ini mengolah data warga berdasarkan atribut pendapatan, kondisi rumah, jumlah tanggungan, dan pekerjaan. Melalui perhitungan jarak menggunakan Euclidean Distance dan parameter $K = 5$, hasil menunjukkan bahwa data baru diklasifikasikan sebagai layak menerima bantuan sosial. Secara keseluruhan, metode K-NN dinilai sederhana namun akurat, serta berpotensi menjadi solusi untuk meningkatkan ketepatan penyaluran bantuan sosial di masyarakat. Kelebihan dari penelitian ini adalah Metode mudah diimplementasikan dan dipahami, Hasil klasifikasi bersifat objektif dan berbasis data, Fleksibel terhadap berbagai jenis atribut sosial ekonomi. Kekurangan dari penelitian ini adalah Belum dilakukan pengujian terhadap akurasi model secara kuantitatif, Ukuran dataset sangat terbatas, Sensitif terhadap nilai parameter K, Belum mempertimbangkan normalisasi dan pembobotan atribut, Tidak ada perbandingan dengan algoritma lain.
9	Klasifikasi Program Bantuan Sosial Menggunakan Algoritma K-Nearest Neighbor (K-NN) (Studi Kasus Kecamatan Malangbong Kabupaten Garut) Hamzah Nurrifqi Fakhri Fikrillah(2025)	Hasil dari penelitian yang dilakukan oleh Fikrillah (2025), penerapan algoritma K-Nearest Neighbor (K-NN) terbukti efektif dalam mengklasifikasikan keluarga penerima Program Bantuan Sosial (BANSOS) di Kecamatan Malangbong, Kabupaten Garut. Dengan menggunakan pendekatan CRISP-DM dan perangkat lunak RapidMiner, penelitian ini berhasil mengolah 19.943 data keluarga yang telah Melalui perhitungan jarak menggunakan Euclidean Distance dan parameter $K = 5$, hasil menunjukkan bahwa data baru diklasifikasikan sebagai layak menerima bantuan sosial. Secara



Hak Cipta Dilindungi Undang-Undang

Universitas Islam Indragiri

1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.
2. Penggunaan untuk kepentingan akademik, penelitian, dan pendidikan diperbolehkan dengan mencantumkan sumber. Plagiarisme juga dilarang untuk kepentingan komersial atau pelanggaran hak cipta dapat dikenakan sanksi sesuai dengan UU Hak Cipta di Indonesia.
3. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.

Tabel 2.1 Riview Jurnal Terdahulu Lanjutan

No	Judul, Penulis, Dan Tahun	Hasil Penelitian
		keseluruhan, metode K-NN dinilai sederhana namun akurat, serta berpotensi menjadi solusi untuk meningkatkan ketepatan penyaluran bantuan sosial di masyarakat. Kelebihan dari penelitian ini adalah Metode mudah diimplementasikan dan dipahami, Hasil klasifikasi bersifat objektif dan berbasis data, Fleksibel terhadap berbagai jenis atribut sosial ekonomi. Kekurangan dari Nilai recall sangat rendah, Tidak dijelaskan metode normalisasi atau standarisasi, Tidak ada perbandingan dengan algoritma lain, Tidak membahas kompleksitas komputasi, Ketidakseimbangan kelas tidak dijelaskan.
10	Sistem Pendukung Keputusan Penerima Bantuan Langsung Tunai Dana Desa Menggunakan Metode Algoritma K-Means Clustering Sarmila Sari, Joy Nashar Utamajaya (2022),	Berdasarkan hasil penelitian yang telah dilakukan secara manual dengan metode K-means clustering dan evaluasi hasil keseluruhan data menggunakan Software Rapidminer versi 5 dihasilkan Nama-Nama masyarakat yang layak menerima bantuan dengan hasil yang sama yaitu terdapat 3 cluster yang dihasilkan dalam mengelompokkan data penduduk kurang mampu penerima bantuan langsung tunai dana desa, dengan cluster 1 adalah kelompok dengan status penerima bantuan sosial tepat sasaran sebanyak 87 data Penduduk, cluster 2 dengan status penerima bantuan sosial dipertimbangkan sebanyak 193 Penduduk, dan cluster 3 dengan status penerima bantuan sosial tidak tepat sasaran sebanyak 56 Penduduk, dimana Pengolahan data yang dilakukan menghasilkan kelompok penerima bantuan yang telah dipertimbangkan menggunakan 5 variabel penelitian yaitu Nama, Pekerjaan, Tanggungan, Kondisi Rumah, dan penghasilan sehingga pihak desa bisa lebih mudah dalam menentukan penerima bantuan Langsung tunai dana desa dengan teknik Data Mining. Dari hasil analisis



Hak Cipta Dilindungi Undang-Undang

Universitas Islam Indragiri

1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.

2. Penggunaan untuk kepentingan akademik, penelitian, dan pendidikan diperbolehkan dengan mencantumkan sumber.

3. Penggunaan tanpa izin untuk kepentingan komersial atau pelanggaran hak cipta dapat dikenakan sanksi sesuai dengan UU Hak Cipta di Indonesia.

Plagiarisme juga dilarang dan dapat dikenakan sanksi. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.

Tabel 2.1 Riview Jurnal Terdahulu Lanjutan

No	Judul, Penulis, Dan Tahun	Hasil Penelitian
10		data dapat ditarik kesimpulan yaitu masyarakat yang menerima bantuan berada pada cluster 0 (cluster 1) yaitu masyarakat yang benar- benar membutuhkan bantuan dari pemerintah, dimana penerima bantuan bekerja sebagai buruh, Kondisi rumah terbuat dari papan, dan memiliki penghasilan di bawah Rp 500.000. Kelebihan dari penelitian ini adalah Pendekatan berbasis data yang objektif, Mempermudah pengambilan keputusan pemerintah desa, Pemanfaatan variabel sosial ekonomi yang relevan, Hasil konsisten antara perhitungan manual dan sistem (RapidMiner), Meningkatkan efisiensi proses penyaluran bantuan. Kekurangan dari penelitian ini adalah Jumlah variabel terbatas, Tidak dilakukan evaluasi model dengan metrik kuantitatif, Sensitif terhadap skala data dan outlier, Sensitif terhadap skala data dan outlier, Tidak ada perbandingan dengan algoritma lain.

Tabel 2.1 Berdasarkan tinjauan terhadap sepuluh jurnal terdahulu yang telah dianalisis, dapat disimpulkan bahwa penerapan algoritma machine learning dalam sistem pendukung keputusan penyaluran bantuan sosial telah menunjukkan efektivitas yang signifikan. Algoritma Naive Bayes menjadi metode yang paling banyak diteliti dengan tingkat akurasi yang bervariasi antara 77,98% hingga 95%, dimana penelitian Adha et al. (2025) mencapai akurasi tertinggi sebesar 95% dalam optimalisasi pengalokasian bantuan sosial di Kota Banjarmasin. Sementara itu, algoritma K-Nearest Neighbor (K-NN) juga menunjukkan performa yang memadai dengan rentang akurasi 68,82% hingga 89,9%, meskipun cenderung memerlukan komputasi yang lebih tinggi dan sensitif terhadap pemilihan parameter K.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.

2. Penggunaan untuk kepentingan akademik, penelitian, dan pendidikan diperbolehkan dengan mencantumkan sumber.

3. Plagiarisme juga dilarang dan dapat dikenakan sanksi.

4. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.

Universitas Islam Indragiri



1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.
2. Penggunaan tanpa izin untuk kepentingan akademik, penelitian, dan pendidikan diperbolehkan dengan mencantumkan sumber. Plagiarisme juga dilarang dan dapat dikenakan sanksi.
3. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.

Studi komparatif yang dilakukan oleh Ramadani et al. (2024) mengungkapkan bahwa algoritma C4.5 memberikan kinerja superior dengan akurasi 91,8%, presisi 90,7%, recall 92,3%, dan nilai AUC 0,944 dalam mengklasifikasikan kelayakan penerima Program Keluarga Harapan, mengungguli Naive Bayes (87,2%) dan K-NN (89,9%). Temuan ini mengindikasikan bahwa pemilihan algoritma yang tepat sangat bergantung pada karakteristik dataset dan konteks permasalahan yang dihadapi. Selain itu, pendekatan clustering menggunakan metode K-Means yang diterapkan oleh Sari dan Utamajaya (2022) mampu mengelompokkan penerima bantuan menjadi tiga kategori prioritas yaitu tepat sasaran, dipertimbangkan, dan tidak tepat sasaran, yang memfasilitasi pengambilan keputusan yang lebih terstruktur.

Meskipun demikian, terdapat beberapa keterbatasan metodologis yang konsisten ditemukan dalam sebagian besar penelitian, antara lain: pertama, ukuran dataset yang relatif terbatas dengan rentang 60 hingga 19.943 data yang berpotensi mengurangi generalisasi model; kedua, minimnya validasi model secara komprehensif seperti cross-validation atau uji signifikansi statistik; ketiga, kurangnya perbandingan sistematis antar algoritma untuk menentukan metode optimal; keempat, tidak adanya penanganan ketidakseimbangan kelas (class imbalance) yang dapat mempengaruhi performa klasifikasi; dan kelima, keterbatasan variabel penelitian yang belum mencakup faktor ekonomi dan sosial yang lebih kompleks. Oleh karena itu, penelitian lanjutan disarankan untuk mengintegrasikan multiple algorithms comparison dengan teknik ensemble learning, menerapkan *feature selection* dan *hyperparameter tuning* untuk

optimalisasi model, serta melakukan evaluasi berkala terhadap relevansi model dalam konteks dinamika sosial ekonomi masyarakat yang terus berkembang.

2.9 Variabel Penelitian

Variabel penelitian adalah sesuatu yang menjadi fokus pengamatan dalam suatu penelitian dan dapat diukur, dihitung, atau dikategorikan. Variabel digunakan untuk menjelaskan hubungan, pengaruh, atau perbedaan antara satu konsep dengan konsep lainnya. Dalam penelitian kuantitatif, variabel biasanya dibagi menjadi variabel yang memengaruhi (variabel independen) dan variabel yang dipengaruhi (variabel dependen). Dengan kata lain, variabel penelitian adalah unsur yang dijadikan acuan untuk mengumpulkan data sehingga peneliti dapat menjawab rumusan masalah.

2.9.1 Variabel Independen

Variabel independen (variabel bebas) merupakan variabel yang memengaruhi atau menjadi faktor penyebab perubahan pada variabel dependen, yaitu kelayakan penerima Bantuan Langsung Tunai (BLT). Dalam penelitian berjudul “Implementasi Algoritma Naive Bayes untuk Menentukan Kelayakan Penerima Bantuan Langsung Tunai di Desa Sungai Ambat”, variabel independen adalah faktor-faktor sosial ekonomi rumah tangga yang diduga berkontribusi terhadap penentuan status kelayakan penerima BLT. Variabel-variabel tersebut berkaitan langsung dengan kondisi kesejahteraan warga, sehingga menjadi masukan penting bagi model klasifikasi berbasis algoritma Naive Bayes. Adapun variabel independen yang digunakan dalam penelitian ini meliputi:

1. Usia Kepala Keluarga, yaitu umur kepala keluarga dalam satuan tahun;





Hak Cipta Dilindungi Undang-Undang

1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.

Penggunaan untuk kepentingan akademik, penelitian, dan pendidikan diperbolehkan dengan mencantumkan sumber.

2. Penggunaan tanpa izin untuk kepentingan komersial atau pelanggaran hak cipta dapat dikenakan sanksi sesuai dengan UU Hak Cipta di Indonesia.

Plagiarisme juga dilarang dan dapat dikenakan sanksi.

3. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.

Universitas Islam Indragiri

2. Jumlah Anggota Rumah Tangga, yakni total anggota yang tinggal satu rumah dan menjadi tanggungan ekonomi;
3. Jenis kelamin kepala keluarga, yaitu Jenis kelamin kepala keluarga yang tercatat sebagai penanggung jawab rumah tangga;
4. Pendidikan Kepala Keluarga berdasarkan tingkat pendidikan terakhir yang diselesaikan;
5. Pendapatan Bulanan, yaitu total penghasilan keluarga dalam satu bulan;
6. Status Pekerjaan Kepala Keluarga sebagai sumber nafkah utama;
7. Status Penerima Bantuan Sosial Lain, yaitu apakah keluarga menerima bantuan sosial lain dari pemerintah;
8. Kepemilikan Aset Dasar, seperti rumah, kendaraan, atau tanah;
9. Kelayakan Kondisi Rumah, yaitu kualitas fisik bangunan rumah;
10. Akses Air Bersih berdasarkan sumber air yang digunakan;
11. Akses Listrik sebagai ketersediaan fasilitas listrik di rumah tangga;
12. Keberadaan Lansia, yaitu apakah terdapat anggota keluarga berusia di atas 60 tahun; dan
13. Keberadaan Penyandang Disabilitas di dalam rumah tangga.

Hubungan variabel independen dengan penelitian ini sangat penting karena seluruh variabel tersebut menjadi input (fitur) bagi algoritma Naive Bayes untuk melakukan proses analisis dan klasifikasi. Melalui data yang dikumpulkan dari variabel-variabel sosial ekonomi tersebut, algoritma Naive Bayes akan mencari pola dan hubungan probabilistik yang dapat digunakan untuk memprediksi nilai variabel dependen, yaitu status kelayakan penerima BLT. Dengan demikian, kedua belas variabel independen tersebut berfungsi sebagai dasar bagi model untuk



menghasilkan prediksi yang objektif dan terukur, sekaligus mencerminkan kondisi sosial ekonomi masyarakat Desa Sungai Ambat sebagai dasar pengambilan keputusan oleh pemerintah desa.

2.9.2 Variabel Dependen

Dalam penelitian ini, variabel dependen yang digunakan adalah kelayakan penerima Bantuan Langsung Tunai (BLT) di Desa Sungai Ambat. Variabel ini menjadi fokus utama yang ingin diprediksi menggunakan algoritma Naive Bayes. Dengan kata lain, variabel dependen merupakan hasil atau keluaran dari proses analisis yang menunjukkan apakah seorang warga layak atau tidak layak menerima BLT berdasarkan kriteria yang berlaku dalam pendataan sosial ekonomi masyarakat desa.

Hubungan variabel dependen dengan penelitian ini sangat erat, karena tujuan utama dari penelitian adalah untuk memprediksi status kelayakan tersebut berdasarkan berbagai faktor yang memengaruhinya. Faktor-faktor tersebut, yang disebut sebagai variabel independen, meliputi antara lain jumlah pendapatan keluarga per bulan, jumlah tanggungan, jenis pekerjaan, kondisi fisik rumah, tingkat pendidikan kepala keluarga, akses terhadap fasilitas dasar, serta kepemilikan aset tertentu.

Melalui penerapan algoritma Naive Bayes, penelitian ini berupaya mengidentifikasi pola hubungan antara variabel independen dengan variabel dependen sehingga sistem mampu menghasilkan prediksi yang akurat mengenai status kelayakan penerima BLT. Dengan demikian, variabel dependen berperan penting sebagai tujuan akhir dari proses analisis dan prediksi dalam penelitian ini, sekaligus menjadi dasar bagi pemerintah desa dalam menentukan rekomendasi

1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.
2. Penggunaan tanpa izin untuk kepentingan akademik, penelitian, dan pendidikan diperbolehkan dengan mencantumkan sumber.
3. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.



pemberian BLT secara objektif dan tepat sasaran. Variabel kelayakan ini dikategorikan ke dalam dua kelas utama, yaitu:

1. Layak: warga memenuhi kriteria sosial ekonomi yang ditetapkan dan dianggap perlu menerima bantuan.
2. Tidak Layak: warga tidak memenuhi kriteria atau berada pada kondisi ekonomi yang lebih baik dibanding penerima prioritas.

Data mengenai kelayakan diperoleh melalui pendataan sosial ekonomi desa yang mencakup indikator kesejahteraan rumah tangga serta hasil klasifikasi model machine learning menggunakan algoritma Naive Bayes.

2.10 Rangkuman

Berdasarkan hasil kajian terhadap sepuluh penelitian terdahulu, dapat disimpulkan bahwa penerapan teknik machine learning dan data mining memiliki peran yang signifikan dalam mendukung proses pengambilan keputusan, khususnya pada penentuan kelayakan penerima bantuan sosial. Berbagai algoritma klasifikasi, seperti Naive Bayes, K-Nearest Neighbor (K-NN), C4.5, serta metode clustering K-Means, telah banyak diterapkan untuk mengolah data sosial ekonomi masyarakat sehingga menghasilkan keputusan yang lebih objektif, cepat, dan terukur. Secara umum, penelitian-penelitian tersebut menunjukkan bahwa pemanfaatan algoritma klasifikasi mampu meningkatkan ketepatan sasaran penyaluran bantuan dibandingkan proses seleksi yang dilakukan secara manual.

Hasil kajian menunjukkan bahwa algoritma Naive Bayes merupakan metode yang paling banyak digunakan dalam penelitian terkait klasifikasi penerima bantuan sosial. Hal ini disebabkan karena Naive Bayes memiliki proses komputasi yang sederhana, efisien, serta mampu menghasilkan tingkat akurasi yang cukup



tinggi pada data kategorikal. Berdasarkan penelitian yang dianalisis, tingkat akurasi Naive Bayes berada pada rentang 77,98% hingga 95%, bergantung pada karakteristik data, jumlah data latih, serta tahapan preprocessing yang diterapkan. Di sisi lain, algoritma K-Nearest Neighbor (K-NN) juga menunjukkan performa yang baik dengan tingkat akurasi yang kompetitif, namun kinerjanya sangat dipengaruhi oleh pemilihan nilai parameter K, metode pengukuran jarak, serta distribusi data yang digunakan. Selain kedua algoritma tersebut, beberapa penelitian juga membandingkan Naive Bayes dengan algoritma lain, seperti C4.5, yang pada beberapa kasus menghasilkan performa lebih tinggi. Temuan tersebut menunjukkan bahwa tidak terdapat satu algoritma yang selalu memberikan hasil terbaik pada seluruh dataset, melainkan performa model sangat dipengaruhi oleh karakteristik data, jumlah atribut, kualitas data, serta teknik pengolahan data yang diterapkan sebelum proses klasifikasi dilakukan. Oleh karena itu, proses evaluasi dan perbandingan beberapa algoritma menjadi langkah penting dalam menentukan model yang paling sesuai untuk suatu permasalahan.

Meskipun hasilnya menjanjikan, banyak penelitian memiliki keterbatasan, termasuk ukuran dataset yang kecil, minimnya validasi model, kurangnya perbandingan antar algoritma, tidak adanya penanganan ketidakseimbangan data, dan variabel penelitian yang belum komprehensif. Oleh sebab itu, penelitian selanjutnya perlu mengoptimalkan model dengan perbandingan multi-algoritma, teknik ensemble, feature selection, hyperparameter tuning, serta evaluasi berkala agar hasil lebih akurat dan relevan terhadap dinamika sosial ekonomi masyarakat.

1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.
2. Penggunaan tanpa izin untuk kepentingan akademik, penelitian, dan pendidikan diperbolehkan dengan mencantumkan sumber. Plagiarisme juga dilarang dan dapat dikenakan sanksi.
3. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.