



BAB III

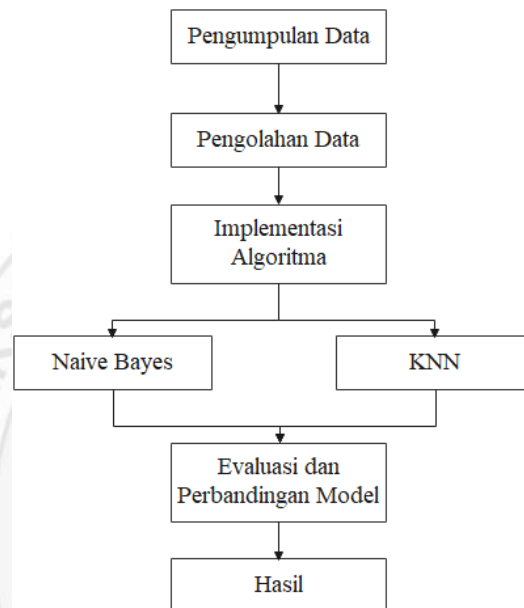
METODE PENELITIAN

Dalam bab metode penelitian ini dijelaskan langkah-langkah yang digunakan dalam proses menentukan kelayakan penerima Bantuan Langsung Tunai (BLT) di Desa Sungai Ambat. Metode penelitian disusun mulai dari tahap identifikasi masalah, pengumpulan data, pengolahan data, hingga proses klasifikasi dan evaluasi hasil. Penelitian ini menggunakan pendekatan kuantitatif, karena seluruh analisis dilakukan berdasarkan data sosial ekonomi masyarakat yang diolah secara sistematis dan terukur. Model yang digunakan dalam penelitian ini adalah algoritma Naive Bayes, yang berfungsi sebagai alat untuk melakukan menentukan kelayakan penerima BLT berdasarkan variabel independen yang telah ditentukan. Melalui penerapan algoritma tersebut, penelitian bertujuan menghasilkan kelayakan penerima bantuan yang objektif, akurat, dan dapat membantu pemerintah desa dalam menyalurkan bantuan secara tepat sasaran.

1.1 Kerangka Penelitian

Kerangka penelitian menggambarkan alur berpikir dan tahap kegiatan yang dilakukan penelitian dalam mencapai tujuan penelitian. Setiap langkah disusun secara sistematis agar proses pengumpulan, pengolahan, dan analisis data dapat berjalan dengan terarah dan menghasilkan kelayakan penerima Bantuan Langsung Tunai (BLT) yang valid. Proses dan langkah penelitian ini meliputi pengumpulan data warga Desa Sungai Ambat, pengolahan data pra-pemrosesan (preprocessing), implementasi algoritma Naive Bayes Sebagai Model utama, serta model *K-Nearest Neighbor* (KNN) sebagai pembanding. Evaluasi performa dilakukan menggunakan

metrik seperti *accuracy*, *precision*, *recall*, *F1-score* dan *confusion matrix* untuk menilai kualitas model. Selanjutnya, hasil penentuan dianalisis untuk menentukan efektifitas masing-masing algoritma serta mengidentifikasi variabel-variabel utama yang berpengaruh terhadap keputusan kelayakan penerima BLT di Desa Sungai Ambat.



Gambar 3.1 Alur Kerangka Penelitian

1.1.1 Pengumpulan data

Pengumpulan data yaitu mengumpulkan data warga Desa Sungai Ambat yang menjadi calon penerima BLT. Untuk memperoleh data yang akurat dan sesuai dengan kebutuhan penelitian, beberapa teknik pengumpulan data digunakan, yaitu:

1. Observasi

Teknik observasi digunakan untuk mengamati langsung kondisi masyarakat serta mengetahui proses administrasi penyaluran BLT di kantor



desa. Observasi membantu peneliti memahami situasi dan kondisi nyata yang relevan dengan penelitian.

2. Wawancara

Wawancara dilakukan secara terstruktur dengan perangkat desa guna mendapatkan penjelasan mengenai prosedur penentuan penerima BLT, kriteria-kriteria yang digunakan, dan permasalahan yang sering terjadi dalam proses seleksi. Wawancara ini memberikan informasi tambahan sebagai pendukung analisis data sekunder.

3. Dokumentasi

Dokumentasi dilakukan dengan mengumpulkan data yang sudah tersedia dalam arsip desa. Teknik ini digunakan untuk memperoleh dataset yang berisi atribut-atribut sosial ekonomi warga sebagai bahan utama dalam proses pemodelan Naive Bayes. Data yang diperoleh melalui dokumentasi merupakan data aktual dan telah digunakan dalam proses administrasi desa.

Data yang diperoleh terdiri dari tiga belas variabel input dan target sebanyak satu kelas. Adapun variabel input dalam penelitian ini dapat dilihat pada tabel 3.2 dibawah ini :

Tabel 3.2 Variabel Masukan

No	Variabel Independen	Definisi Operasional	Indikator	Jenis Data	Kategori / Skala
1	Usia Kepala Keluarga	Umur kepala keluarga yang dihitung dalam satuan tahun.	Usia berdasarkan tanggal lahir	Numerik	Tahun
2	Jumlah Anggota Rumah Tangga	Total anggota yang tinggal dalam satu rumah dan menjadi	Jumlah anggota keluarga	Numerik	Orang



Tabel 3.2 Variabel Masukan Lanjutan

No	Variabel Independen	Definisi Operasional	Indikator	Jenis Data	Kategori / Skala
		tanggungan ekonomi.			
3	Jenis kelamin kepala keluarga	Jenis kelamin kepala keluarga yang tercatat sebagai penanggung jawab rumah tangga	Laki-laki/ Perempuan	kategorikal	Biner
4	Pendidikan Kepala Keluarga	Tingkat pendidikan terakhir yang diselesaikan oleh kepala keluarga.	SD, SMP, SMA, >SMA	Kategorikal	Ordinal
5	Pendapatan Bulanan	Total penghasilan keluarga dalam satu bulan.	Upah, hasil usaha, hasil pertanian	Numerik / Kategori	Rupiah / Rendah–Sedang–Tinggi
6	Status Pekerjaan Kepala Keluarga	Jenis pekerjaan utama yang menjadi sumber nafkah keluarga.	Buruh, Petani, Wiraswasta, Tidak bekerja	Kategorikal	Nominal
7	Penerima Bantuan Sosial Lain	Status apakah rumah tangga menerima program bantuan lain.	Ya / Tidak	Kategorikal	Biner
8	Kepemilikan Aset Dasar	Kondisi kepemilikan aset seperti rumah, tanah, atau kendaraan.	Memiliki rumah, kendaraan, tanah	Kategorikal	Biner / Nominal



Tabel 3.2 Variabel Masuk

No	Variabel Independen	Definisi Operasional	Indikator	Jenis Data	Kategori / Skala
9	Kelayakan Kondisi Rumah	Kualitas bangunan rumah berdasarkan kondisi fisik.	Layak, Kurang Layak, Tidak Layak	Kategorikal	Ordinal
10	Akses Air Bersih	Jenis sumber air yang digunakan rumah tangga.	PDAM, Sumur, Sungai	Kategorikal	Nominal
11	Akses Listrik	Ketersediaan listrik di rumah tangga.	Ada / Tidak ada	Kategorikal	Biner
12	Keberadaan Lansia	Adanya anggota rumah tangga berusia >60 tahun.	Ya / Tidak	Kategorikal	Biner
13	Keberadaan Penyandang Disabilitas	Adanya anggota rumah tangga yang memiliki keterbatasan fisik atau mental.	Ya / Tidak	Kategorikal	Biner

Sedangkan sebagai unit output atau target yang akan dihasilkan dapat dilihat pada tabel 3.3 dibawah ini :

Tabel 3.3 Variabel Output

No	Variabel	Keterangan
1	0	Tidak Layak
2	1	Layak

3.1.2 Pengolahan data

Pengolahan data yaitu meliputi pembersihan data, penanganan nilai hilang, transformasi atribut, dan normalisasi data agar siap digunakan dalam



model. Pada tahap ini juga dilakukan pembagian dataset menjadi data latih dan data uji. Pengolahan data dilakukan agar data siap diproses ke dalam algoritma Naive Bayes dan KNN. Tahapan pengolahan data meliputi:

1. Pembersihan Data (Data Cleaning) Proses pembersihan data dilakukan untuk memastikan data yang digunakan benar-benar siap diproses oleh algoritma. menghapus data duplikat, menangani nilai hilang, dan memperbaiki kesalahan input
2. Transformasi Atribut (Data Transformation) Data yang telah dibersihkan selanjutnya ditransformasikan agar sesuai dengan kebutuhan model. Tahap tersebut yaitu : Diskritisasi atribut numerik, Pengkodean data kategorikal (label encoding atau one-hot encoding), Normalisasi data, Transformasi format, seperti mengubah tipe data (string → numerik).
3. Pemilihan Fitur (Feature Selection) Tahap ini bertujuan memilih atribut-atribut yang paling relevan dalam menentukan kelayakan penerima BLT. Pemilihan fitur dilakukan agar model lebih akurat dan tidak dipengaruhi oleh atribut yang tidak penting.
4. Pembagian Dataset, Data yang telah melewati tahap pembersihan dan transformasi selanjutnya dibagi menjadi dua yaitu Data Latih (Training Set) – 80% Digunakan untuk melatih model Naive Bayes dan KNN dengan mempelajari pola dari atribut-atribut yang ada. Data Uji (Testing Set) – 20% Digunakan untuk menguji kinerja model secara objektif dan mengukur kemampuan model dalam melakukan prediksi pada data baru.

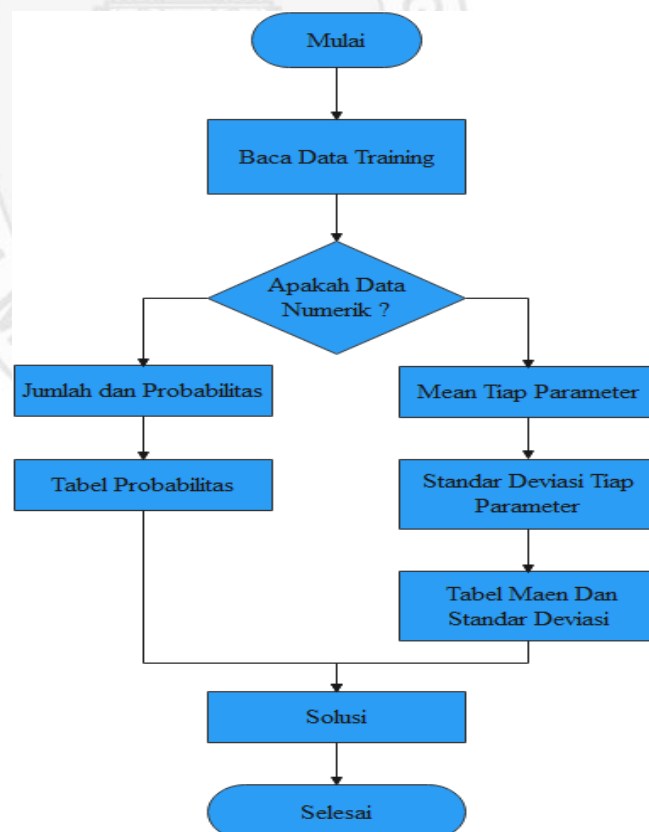


3.1.3 Implementasi Algoritma

Tahap implementasi algoritma merupakan proses memasukkan data yang telah melalui tahap pengolahan ke dalam model klasifikasi. Pada penelitian ini digunakan dua algoritma, yaitu Naive Bayes sebagai model utama dan K-Nearest Neighbor (KNN) sebagai model pembanding. Implementasi dilakukan menggunakan perangkat lunak Google Colab dengan langkah-langkah sebagai berikut:

1. Implementasi Algoritma Naive Bayes

Naive Bayes adalah algoritma klasifikasi berbasis probabilitas yang bekerja dengan menerapkan *Teorema Bayes* dan asumsi independensi antar atribut.



Gambar 3.2 Alur Kerja Naive Bayes



Teorema Bayes dirumuskan sebagai berikut :

$$P(H|X) = \frac{P(X | H) \times P(H)}{P(X)} \dots \dots \dots [1]$$

Adapun tahapan implementasinya dalam penelitian ini yaitu:

1. Memasukkan data latih ke model
2. Menghitung nilai prior dan likelihood untuk setiap kelas (Layak / Tidak Layak)
3. Menerapkan rumus probabilitas posterior Naive Bayes
4. Menguji model menggunakan data uji
5. Menghasilkan prediksi kelayakan penerima BLT

2. Implementasi Algoritma K-Nearest Neighbor(KNN)

KNN digunakan sebagai algoritma pembandingan untuk melihat perbedaan performa dengan model utama. KNN bekerja dengan mencari sejumlah tetangga terdekat berdasarkan jarak antar data. Adapun tahapan implementasinya dalam penelitian ini yaitu:

1. Melakukan normalisasi data karena KNN sensitif terhadap skala atribut
2. Menentukan nilai **K** terbaik (misalnya K = 3, 5, atau 7)
3. Menghitung jarak Euclidean antara data uji dan data latih

$$d(x, y) = \sqrt{\left(\sum (x_i - y_i)^2\right)} \dots \dots \dots [2]$$

4. Mengambil mayoritas kelas tetangga terdekat
5. Menghasilkan prediksi kelayakan penerima BLT



1.1. 4 Evaluasi Kinerja Model

Evaluasi dilakukan untuk mengetahui seberapa baik setiap algoritma melakukan klasifikasi. Ini dilakukan untuk melihat seberapa baik model yang digunakan dalam memberikan hasil. Proses evaluasi model ini adalah confusion matriks, akurasi, presisi, recall dan F1-score. Metrik yang digunakan adalah:

1. Accuracy (akurasi): mengukur peresentasi prediksi yang benar.

Rumus:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \dots \dots \dots [3]$$

2. Precision : Mengukur tingkat ketepatan ketika model memprediksi”layak”

Rumus :

$$Precision = \frac{TP}{TP + FP} \dots \dots \dots [4]$$

3. Recall : Kemampuan model mendeteksi warga yang benar-benar “layak”

Rumus:

$$Recall = \frac{TP}{TP + FN} \dots \dots \dots [5]$$

4. F1-Score : Gabungan harmonis antara precision dan recall

Rumus :

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \dots \dots \dots [6]$$

5. Confusion Matrix

Confusion matrix menggambarkan hubungan antara kelas aktual dan prediksi model.



Tabel 3.4 *Confusion Matrix*

Aktual / Prediksi	Layak	Tidak Layak
Layak	TP	FN
Tidak Layak	FP	TN

3.1.5 *Paired t-test* untuk Membandingkan Performa Algoritma

Paired t-test (uji t berpasangan) adalah uji statistik parametrik yang digunakan untuk menentukan apakah terdapat perbedaan yang signifikan antara dua kelompok nilai yang berpasangan atau menggunakan sampel yang sama. *Paired t-test* dilakukan melalui serangkaian tahapan untuk menentukan apakah terdapat perbedaan yang signifikan antara dua kelompok nilai berpasangan. Pada penelitian ini, nilai yang dibandingkan adalah akurasi algoritma Naive Bayes dan KNN pada setiap fold k-fold cross validation. Langkah-Langkah *Paired t-test* :

1. Mengumpulkan Pasangan Data (X_1 dan X_2)

Langkah pertama adalah mengumpulkan dua kelompok data yang saling berpasangan. Pada konteks penelitian, data tersebut adalah:

X_1 = akurasi Naive Bayes

X_2 = akurasi KNN

2. Menghitung Selisih Antar Pasangan (d_i)

Untuk setiap pasangan data, hitung selisihnya:

$$d_i = x_{(1,i)} - x_{(2,i)} \dots \dots \dots [7]$$

Hasilnya membentuk daftar selisih:

$$d_1, d_2, d_3, \dots, d_n$$

3. Menghitung Rata-rata Selisih ($\bar{d}$)

Rata-rata selisih dihitung menggunakan rumus:



$$\bar{d} = \frac{\sum_{i=1}^n d_i}{n} \dots \dots \dots [8]$$

Rata-rata ini menunjukkan kecenderungan selisih, apakah Naive Bayes cenderung lebih tinggi, lebih rendah, atau setara dengan KNN.

4. Menghitung Deviasi (Varians) Selisih (s_e^2)

Hitung varians dari selisih tersebut:

$$s_e^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-k} \dots \dots \dots [9]$$

Varians ini menggambarkan tingkat keragaman selisih antarpasangan.

5. Menghitung Standar Deviasi Selisih (s_e)

Standar deviasi diperoleh dari akar varians:

$$s_d = \sqrt{(s_d^2)} \dots \dots \dots [10]$$

Standar deviasi menunjukkan seberapa besar perbedaan data terhadap rata-rata selisih.

6. Menghitung Standard Error (SE)

Kesalahan standar dihitung dengan:

$$SE = \frac{sd}{\sqrt{n}} \dots \dots \dots [11]$$

Nilai ini digunakan untuk menormalkan rata-rata selisih dalam perhitungan uji t.

7. Menghitung Nilai t-hitung (t)

Nilai t-hitung diperoleh dengan rumus:

$$t = \frac{\bar{d}}{SE} = \frac{\bar{d}}{(\frac{s_d}{\sqrt{n}})} \dots \dots \dots [12]$$

Semakin besar nilai t-hitung, semakin besar kemungkinan terdapat perbedaan signifikan antara kedua algoritma.



8. Menentukan Derajat Kebebasan (df)

Derajat kebebasan untuk paired t-test adalah:

$$df = n - 1 \dots \dots \dots [13]$$

9. Menentukan Nilai p-value

Nilai p-value diperoleh dari tabel distribusi t atau perangkat lunak statistik Google Colab berdasarkan perhitungan nilai t-hitung dan derajat kebebasan.

10. Menarik kesimpulan

Bandingkan p-value dengan tingkat signifikansi ($\alpha = 0.05$):

Jika p-value < 0.05 : Terdapat perbedaan signifikan antara performa Naive Bayes dan KNN, Salah satu algoritma terbukti lebih unggul secara statistik.

Jika p-value ≥ 0.05 : Tidak terdapat perbedaan signifikan, Kedua algoritma memiliki performa yang setara secara statistik.

1.2 Alat dan perangkat penelitian

Alat dan perangkat penelitian merupakan komponen penting yang mendukung proses pengolahan data, analisis, dan penyusunan laporan. Pemilihan perangkat yang tepat bertujuan untuk memastikan bahwa seluruh tahapan penelitian dapat dilakukan secara efektif, efisien, dan akurat. Pada penelitian ini, alat dan perangkat yang digunakan meliputi perangkat keras (hardware) serta perangkat lunak (software) yang dijelaskan sebagai berikut:

1. Perangkat Keras (Hardware)

Perangkat keras digunakan untuk menjalankan aplikasi, mengolah data, dan melakukan pemodelan algoritma data mining. Adapun spesifikasi perangkat keras yang digunakan dalam penelitian ini adalah sebagai berikut:

1. Laptop/Komputer dengan spesifikasi minimum:





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. Prosesor: Intel Core i7 atau setara
3. RAM: Minimal 8 GB
4. Penyimpanan: Minimal 256 GB SSD
5. Sistem Operasi: Windows 11

Spesifikasi tersebut dipilih untuk memastikan proses pengolahan data dan pemodelan algoritma dapat berjalan tanpa hambatan, terutama saat memproses dataset berukuran sedang hingga besar.

2. Perangkat Lunak (Software)

Perangkat lunak digunakan untuk memfasilitasi pengolahan data, analisis, visualisasi, dan penyusunan dokumen penelitian. Perangkat lunak yang digunakan dalam penelitian ini meliputi:

1. Microsoft Excel, digunakan untuk pengelolaan data awal, pembersihan data sederhana, serta visualisasi dasar.
 2. Google Colab, digunakan untuk proses pemodelan data mining seperti algoritma Naive Bayes, pengujian model, serta evaluasi hasil.
 3. Microsoft Word, digunakan dalam proses penyusunan laporan penelitian, dokumentasi hasil, serta pembuatan grafik dan interpretasi data.
- ## 3. Data Penelitian

Selain perangkat keras dan lunak, penelitian ini juga memanfaatkan data sebagai objek utama analisis. Data yang digunakan adalah data sosial ekonomi warga Desa Sungai Ambat yang terdiri dari 13 variabel input dan 1 variabel output yang kemudian diolah dan dianalisis menggunakan metode data mining. Data ini merupakan komponen utama dalam proses pemodelan dan menjadi dasar dalam perumusan hasil peneliti