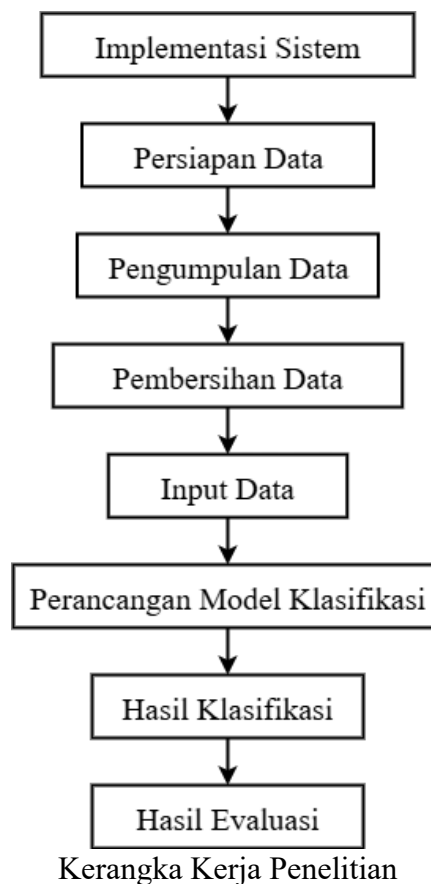


BAB III

ANALISIS DAN METODE

3.1. Algoritma Naïve Bayes

Algoritma Naïve Bayes dalam penelitian ini berperan sebagai inti sistem klasifikasi penerima bantuan BPJS di wilayah Rantauprapat, Labuhanbatu. Pemilihan metode ini didasarkan pada sifatnya yang sederhana, efisien, serta mampu memberikan hasil klasifikasi yang cukup akurat meskipun bekerja dengan data yang terbatas. Dengan penerapan algoritma ini, sistem yang dibangun mampu mengotomatisasi proses seleksi penerima bantuan sehingga lebih objektif dan mengurangi potensi subjektivitas yang sering muncul dalam metode manual.



1. Implementasi sistem dilakukan untuk menerapkan algoritma Naïve Bayes dalam proses klasifikasi penerima bantuan BPJS. Tahap ini menjadi inti penelitian karena seluruh proses dijalankan secara otomatis melalui sistem yang dirancang.
2. Persiapan data mencakup identifikasi atribut penting yang akan digunakan seperti usia, pekerjaan, pendapatan, jumlah tanggungan, dan status rumah. Langkah ini bertujuan agar data yang digunakan sesuai kebutuhan metode klasifikasi.
3. Pengumpulan data dilakukan dengan cara wawancara langsung dan dokumentasi data masyarakat di wilayah Rantauprapat. Data yang dikumpulkan harus relevan dan representatif agar hasil klasifikasi valid.
4. Pembersihan data dilakukan untuk menghilangkan data yang tidak lengkap, duplikat, atau tidak konsisten. Langkah ini penting agar data siap diproses dan tidak menurunkan kualitas model.
5. Input data dilakukan dengan memasukkan data yang telah dibersihkan ke dalam sistem atau aplikasi RapidMiner. Proses ini memungkinkan sistem untuk membaca dan mengolah data sesuai struktur yang dibutuhkan.
6. Perancangan model klasifikasi mencakup pemilihan algoritma Naïve Bayes dan pengaturan parameter model. Tujuannya adalah membangun model yang mampu memprediksi penerima bantuan secara akurat.
7. Hasil klasifikasi berupa prediksi status setiap data uji apakah termasuk kategori “Dapat” atau “Tidak Dapat”. Tahap ini menunjukkan kemampuan model dalam mengenali pola dari data training.

8. Hasil evaluasi diperoleh dengan mengukur performa model menggunakan metrik seperti akurasi, presisi, dan recall. Langkah ini digunakan untuk menilai seberapa baik model dalam melakukan klasifikasi.

Proses pengolahan data pada algoritma Naïve Bayes diawali dengan pengumpulan atribut masyarakat yang relevan, seperti pendapatan bulanan, jenis pekerjaan, jumlah tanggungan, dan status kepemilikan rumah. Data tersebut kemudian melalui tahap preprocessing agar bersih, terstruktur, dan sesuai untuk dianalisis. Pembersihan dan transformasi data dilakukan untuk menghindari ketidaksesuaian yang dapat memengaruhi akurasi hasil. Dengan data yang siap diproses, Naïve Bayes dapat bekerja lebih optimal dalam membangun pola klasifikasi berdasarkan distribusi probabilitas tiap atribut.

Selanjutnya, proses pelatihan model dilakukan menggunakan data training yang tersedia. Naïve Bayes menghitung probabilitas dari masing-masing variabel terhadap kategori yang menjadi target, yaitu "Dapat" atau "Tidak Dapat". Model yang dihasilkan kemudian diuji menggunakan data testing untuk melihat sejauh mana hasil klasifikasi dapat digeneralisasi ke data baru. Pada tahap ini, evaluasi dilakukan menggunakan metrik akurasi, presisi, dan recall yang mampu memberikan gambaran menyeluruh tentang performa model dalam mengklasifikasikan penerima bantuan secara tepat.

Hasil akhir dari penerapan algoritma Naïve Bayes berupa prediksi kategori penerima bantuan BPJS. Sistem ini menghasilkan keluaran yang jelas, yakni apakah seseorang berhak menerima bantuan atau tidak. Dengan keluaran yang akurat, cepat, dan mudah dipahami, sistem ini dapat dijadikan dasar dalam pengambilan

keputusan distribusi bantuan secara lebih transparan. Hal ini diharapkan mampu meningkatkan efektivitas penyaluran bantuan kesehatan, sehingga benar-benar sampai kepada masyarakat yang paling membutuhkan dan mengurangi potensi salah sasaran.

3.2. Pengumpulan Data

Proses pengumpulan data dalam penelitian ini dilakukan dengan menghimpun informasi dari 30 individu yang berasal dari wilayah Rantauprapat, Labuhanbatu. Data yang dikumpulkan mencakup variabel-variabel yang dianggap berpengaruh terhadap klasifikasi status penerimaan bantuan BPJS, yakni nama (kode), usia, jenis pekerjaan, pendapatan per bulan, jumlah tanggungan keluarga, status kepemilikan rumah, serta label klasifikasi akhir yaitu “Dapat” atau “Tidak Dapat”. Setiap entri data merepresentasikan satu individu dan karakteristik sosial ekonominya. Pengumpulan data ini dilakukan secara manual melalui observasi dan wawancara langsung terhadap warga, dengan mempertimbangkan kerahasiaan dan etika penelitian. Atribut-atribut tersebut dipilih karena secara logis mewakili kondisi yang menjadi dasar pertimbangan layak tidaknya seseorang menerima bantuan. Dari 30 data yang diperoleh, mayoritas menunjukkan karakteristik ekonomi menengah ke bawah seperti tidak bekerja, penghasilan di bawah satu juta rupiah, dan tinggal di rumah sewa atau menumpang. Data ini kemudian dijadikan sebagai dataset utama untuk melatih dan menguji model klasifikasi dengan algoritma Naive Bayes dalam proses pengambilan keputusan penerimaan bantuan.

Tabel 3. 1. Data Sampel

Nama	Usia	Pekerjaan	Pendapatan/Bulan	Jumlah Tanggungan	Status Rumah	Label
KP001	36	KP001	Rp 850,000.00	3	Sewa	Dapat
KP002	35	AS001	Rp 850,000.01	1	Milik Sendiri	Tidak Dapat

KP003	30	TBJ001	Rp 850,000.02	4	Menumpang	Dapat
KP004	28	P001	Rp 850,000.03	2	Sewa	Dapat
KP005	56	B001	Rp 500,000.00	4	Menumpang	Dapat
KP006	54	TBJ001	Rp 400,000.00	4	Menumpang	Dapat
KP007	62	TBJ001	Rp 350,000.00	3	Menumpang	Dapat
KP008	60	TBJ001	Rp 400,000.00	5	Menumpang	Dapat
KP009	21	B001	Rp 900,000.00	3	Sewa	Dapat
KP010	52	B001	Rp 500,000.00	4	Menumpang	Dapat
KP011	37	AN001	Rp 2,600,000.00	1	Milik Sendiri	Tidak Dapat
KP012	66	TB001	Rp 500,000.00	3	Sewa	Dapat
KP013	63	TB001	Rp 300,000.00	4	Menumpang	Dapat
KP014	33	B001	Rp 850,000.00	3	Menumpang	Dapat
KP015	32	KS001	Rp 2,200,000.00	0	Milik Sendiri	Tidak Dapat
KP016	70	TB001	Rp 250,000.00	4	Menumpang	Dapat
KP017	63	P001	Rp 700,000.00	4	Sewa	Dapat
KP018	35	PD001	Rp 2,000,000.00	1	Sewa	Dapat
KP019	50	TS001	Rp 900,000.00	3	Sewa	Dapat
KP020	35	TBJ001	Rp 350,000.00	3	Menumpang	Dapat

Tabel data di atas menyajikan informasi tentang 20 individu dengan berbagai latar belakang sosial ekonomi yang menjadi objek klasifikasi penerimaan bantuan BPJS di wilayah Rantauprapat, Labuhanbatu. Berdasarkan data tersebut, terlihat bahwa sebagian besar individu yang tergolong "Dapat" menerima bantuan adalah mereka yang tidak bekerja, memiliki pendapatan di bawah Rp 1.000.000 per bulan, memiliki tanggungan keluarga lebih dari dua orang, dan tinggal di rumah sewa atau menumpang. Sebaliknya, individu yang tergolong "Tidak Dapat" umumnya memiliki pekerjaan tetap seperti PNS atau karyawan swasta, berpenghasilan di atas Rp 2.000.000 per bulan, memiliki sedikit atau tanpa tanggungan, dan tinggal di rumah milik sendiri. Hal ini mengindikasikan bahwa atribut seperti pendapatan bulanan, status pekerjaan, jumlah tanggungan, dan status kepemilikan rumah memiliki pengaruh signifikan terhadap keputusan klasifikasi bantuan, dan menjadi dasar yang tepat untuk diterapkan dalam model klasifikasi menggunakan algoritma Naive Bayes.

3.3. Langkah-Langkah Pengolahan pada Metode Naive Bayes

1. Seleksi Data

Pada tahap seleksi data, proses ini bertujuan untuk membagi data mentah menjadi dua bagian utama, yaitu dataset training dan dataset testing, yang masing-masing digunakan untuk pelatihan dan pengujian model. Dari total 30 data yang tersedia, sebanyak 10 data pertama dipilih sebagai data training yang berfungsi untuk melatih model klasifikasi Naive Bayes agar dapat mengenali pola dan karakteristik dari setiap atribut terhadap label klasifikasi “Dapat” dan “Tidak Dapat”. Sementara itu, 20 data sisanya digunakan sebagai data testing untuk menguji seberapa baik model yang telah dilatih dapat mengklasifikasikan data baru secara akurat. Pembagian ini dilakukan secara manual dan berurutan untuk menjaga konsistensi dalam proses pelatihan dan evaluasi. Dengan proporsi pembagian ini, diharapkan model mampu menunjukkan performa yang optimal dalam memprediksi penerima bantuan secara tepat sasaran.

Tabel 3. 2. Data Training

Nam a	Usia	Pekerjaan	Pendapatan/Bulan	Jumlah Tanggungan	Status Rumah	Label
KP001	36	KP001	Rp 850,000.00	3	Sewa	Dapat
KP002	35	AS001	Rp 850,000.01	1	Milik Sendiri	Tidak Dapat
KP003	30	TBJ001	Rp 850,000.02	4	Menumpang	Dapat
KP004	28	P001	Rp 850,000.03	2	Sewa	Dapat
KP005	56	B001	Rp 500,000.00	4	Menumpang	Dapat
KP006	54	TBJ001	Rp 400,000.00	4	Menumpang	Dapat
KP007	62	TBJ001	Rp 350,000.00	3	Menumpang	Dapat
KP008	60	TBJ001	Rp 400,000.00	5	Menumpang	Dapat
KP009	21	B001	Rp 900,000.00	3	Sewa	Dapat
KP010	52	B001	Rp 500,000.00	4	Menumpang	Dapat

Tabel di atas merupakan data training yang digunakan dalam penelitian ini, terdiri dari 10 entri data calon penerima bantuan BPJS di Rantauprapat, Kabupaten Labuhanbatu. Setiap entri memuat informasi penting seperti nama responden, usia, jenis pekerjaan, jumlah pendapatan per bulan, jumlah tanggungan keluarga, status kepemilikan rumah, dan label klasifikasi apakah individu tersebut layak menerima bantuan (Dapat) atau tidak (Tidak Dapat). Data ini mencerminkan keberagaman kondisi sosial-ekonomi responden, mulai dari pekerjaan tidak tetap seperti buruh, kuli pasar, hingga PNS, serta variasi penghasilan yang mayoritas berada di bawah satu juta rupiah. Selain itu, kondisi tempat tinggal juga bervariasi, mulai dari menumpang, sewa, hingga milik sendiri. Data training ini berperan penting dalam proses pembelajaran algoritma Naive Bayes untuk mengenali pola hubungan antar atribut yang memengaruhi keputusan klasifikasi bantuan, sehingga model dapat digunakan untuk memprediksi secara akurat pada data testing yang belum dikenali sebelumnya.

Tabel 3. 3. Data Testing

Nam a	Usia	Pekerjaan	Pendapatan/Bulan	Jumlah Tanggungan	Status Rumah	Label
KP001	37	AN001	Rp 2,600,000.00	1	Milik Sendiri	Tidak Dapat
KP002	66	TB001	Rp 500,000.00	3	Sewa	Dapat
KP003	63	TB001	Rp 300,000.00	4	Menumpang	Dapat
KP004	33	B001	Rp 850,000.00	3	Menumpang	Dapat
KP005	32	KS001	Rp 2,200,000.00	0	Milik Sendiri	Tidak Dapat
KP006	70	TB001	Rp 250,000.00	4	Menumpang	Dapat
KP007	63	P001	Rp 700,000.00	4	Sewa	Dapat
KP008	35	PD001	Rp 2,000,000.00	1	Sewa	Dapat
KP009	50	TS001	Rp 900,000.00	3	Sewa	Dapat
KP010	35	TBJ001	Rp 350,000.00	3	Menumpang	Dapat

Tabel di atas menampilkan 10 entri data testing yang digunakan dalam penelitian ini untuk menguji kinerja model klasifikasi Naive Bayes dalam menentukan kelayakan penerimaan bantuan BPJS di Rantauprapat, Labuhanbatu. Data ini memuat atribut-atribut penting seperti nama responden, usia, pekerjaan, pendapatan bulanan, jumlah tanggungan, dan status rumah tinggal, yang diakhiri dengan label klasifikasi “Dapat” atau “Tidak Dapat” sebagai penanda apakah responden termasuk penerima bantuan atau bukan. Komposisi data testing menunjukkan variasi yang luas, baik dari segi pekerjaan (seperti PNS, petani, pedagang, hingga tidak bekerja), tingkat pendapatan yang mayoritas masih berada di bawah standar penghasilan minimum, serta kondisi tempat tinggal yang mencerminkan status ekonomi masing-masing responden. Dengan data testing ini, model yang telah dilatih pada data training akan diuji untuk melihat seberapa akurat dan efektif algoritma Naive Bayes dalam mengenali pola dan memprediksi kategori penerimaan bantuan secara realistis dan berdasarkan fakta lapangan.

2. Preprocessing Data

Sebelum digunakan dalam proses pelatihan model, data training terlebih dahulu melalui tahap pembersihan untuk memastikan kualitas dan konsistensi data yang optimal. Pembersihan data ini mencakup konversi format nilai pendapatan dari bentuk mata uang ke angka numerik yang dapat dikenali oleh sistem, penyederhanaan nilai atribut seperti pekerjaan dan status rumah menjadi format kategorikal yang sesuai dengan kebutuhan algoritma Naive Bayes, serta pemeriksaan terhadap adanya data kosong atau tidak valid yang dapat memengaruhi akurasi model. Proses ini penting agar data latih yang digunakan benar-benar

mencerminkan karakteristik yang dapat dianalisis oleh Naive Bayes, yang bekerja berdasarkan probabilitas dari kemunculan masing-masing atribut terhadap label klasifikasi. Dengan data yang telah dibersihkan dan disesuaikan, model Naive Bayes dapat dilatih secara optimal untuk mengenali pola-pola yang berkaitan dengan penerimaan bantuan (label “Dapat” dan “Tidak Dapat”), sehingga meningkatkan ketepatan dalam proses klasifikasi pada data testing. Untuk tahapan ini dilakukan dengan menggunakan aplikasi Microsoft excel yaitu sebagai berikut.

Tabel 3. 4. Pembersihan Data

Atribut	Partisi	Dapat	Tidak Dapat	Kategori	Nilai
Pekerjaan	KP001	0.1111	0	Dapat	0.9
	AS001	0	1		
	TBJ001	0.4444	0		
	P001	0.1111	0		
	B001	0.3333	0		
Pendapatan/Bulanan	850000	0.3333	1		
	500000	0.2222	0		
	400000	0.2222	0		
	350000	0.1111	0		
	400000	0.2222	0		
	900000	0.1111	0		
Jumlah Tanggungan	1	0	1	Tidak Dapat	0.1
	2	0.1111	0		
	3	0.3333	0		
	4	0.4444	0		
	5	0.1111	0		
Status Rumah	Sewa	0.3333	0		
	Milik Sendiri	0	1		
	Menumpang	0.6667	0		

Tabel di atas merupakan hasil dari proses pra-pemrosesan data latih yang telah dikonversi ke dalam bentuk probabilitas kemunculan masing-masing atribut terhadap label klasifikasi “Dapat” dan “Tidak Dapat”. Atribut-atribut seperti

Pekerjaan, Pendapatan Bulanan, Jumlah Tanggungan, dan Status Rumah dihitung frekuensinya berdasarkan distribusi dari 10 data training sebelumnya. Misalnya, pekerjaan “Tidak Bekerja” muncul sebanyak 4 kali dalam kategori “Dapat”, sehingga menghasilkan probabilitas 0.4444. Sementara pekerjaan seperti “PNS” hanya muncul dalam kategori “Tidak Dapat” dengan probabilitas 1.0. Hal yang sama berlaku pada atribut lainnya, seperti jumlah tanggungan “4” yang muncul 4 kali dalam kategori “Dapat” menghasilkan nilai 0.4444. Nilai-nilai probabilitas ini nantinya digunakan oleh algoritma Naive Bayes dalam menghitung kemungkinan kelas (label) dari data testing, berdasarkan kemunculan atribut-atribut pada setiap baris data. Dengan begitu, model dapat memperkirakan apakah seseorang berpeluang “Dapat” atau “Tidak Dapat” bantuan berdasarkan nilai-nilai input tersebut.

3. *Perhitungan Data*

Pada perhitungan yang akan dilakukan, pertama penulis akan menghitung data D1. Adapun perhitungannya sebagai berikut.

$$\begin{aligned}
 P(\text{Kategori}) &= P(\text{Pekerjaan}|\text{PNS}) \times P(\text{Pendapatan}|\text{Rp. 850.000}) \times P(\text{Jumlah} \\
 &\quad \text{Tanggungan}|1) \times P(\text{Status Rumah}) P(\text{Kategori}|\text{Dapat}) \\
 P(\text{Dapat}) &= P(\text{PNS}|\text{Dapat}) \times P(\text{Rp. 850.000}|\text{Dapat}) \times P(1|\text{Dapat}) \times P(\text{Milik} \\
 &\quad \text{Sendiri}|\text{Dapat}) \times P(\text{Kategori}|\text{Dapat}) \\
 &= 0 \times 0.3333 \times 0 \times 0 \times 0.9 \\
 &= 0,00137174 \text{ (Nilai Dapat)}
 \end{aligned}$$

$$\begin{aligned}
P(\text{Tidak Dapat}) &= P(\text{PNS}|\text{Tidak Dapat}) \times P(\text{Rp. 850.000}|\text{Tidak Dapat}) \times P(1|\text{Tidak Dapat}) \\
&\quad \times P(\text{Milik Sendiri}|\text{Tidak Dapat}) \times P(\text{Kategori}|\text{Tidak Dapat}) \\
&= 0 \times 1 \times 1 \times 0.1 \\
&= 0 \text{ (Nilai Tidak Dapat)}
\end{aligned}$$

Hasil yang diperoleh dari data D1 bahwasanya nilai Dapat yang diperoleh yaitu sebesar 0,00137174 dan untuk nilai tidak Dapat yang diperoleh yaitu sebesar 0. Jadi dari hasil yang sudah diperoleh bahwa D1 Dapat memperoleh BPJS.