

**IMPLEMENTASI DATA *MINING* DENGAN  
ALGORITMA *NAIVE BAYES* UNTUK KLASIFIKASI  
KELAYAKAN PENERIMA BANTUAN SEMBAKO**

**SKRIPSI**



**Oleh:  
Ayu Entini Lumban Raja  
190210031**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNIK DAN KOMPUTER  
UNIVERSITAS PUTERA BATAM  
TAHUN 2023**

**IMPLEMENTASI DATA *MINING* DENGAN  
ALGORITMA *NAIVE BAYES* UNTUK KLASIFIKASI  
KELAYAKAN PENERIMA BANTUAN SEMBAKO**

**SKRIPSI**

**Untuk memenuhi salah satu syarat  
Memperoleh gelar Sarjana**



**Oleh:  
Ayu Entini Lumban Raja  
190210031**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNIK DAN KOMPUTER  
UNIVERSITAS PUTERA BATAM  
TAHUN 2023**

## SURAT PERNYATAAN ORISINALITAS

Yang bertanda tangan di bawah ini saya:

Nama : Ayu Entini Lumban Raja

Npm : 190210031

Fakultas : Teknik dan Komputer

Program Studi : Teknik Informatika

Menyatakan Bahwa “Skripsi” yang saya buat dengan judul:

**IMPLEMENTASI DATA *MINING* DENGAN ALGORITMA *NAIVE BAYES*  
UNTUK KLASIFIKASI KELAYAKAN PENERIMA BANTUAN  
SEMBAKO**

Adalah hasil karya Pribadi dan bukan “duplikasi” dari karya orang lain. Sepengetahuan saya, di dalam naskah skripsi ini tidak terdapat karya ilmiah atau pendapat yang pernah ditulis atau di terbitkan oleh orang lain, kecuali yang secara tertulis dikutip di dalam naskah ini dan disebutkan dalam sumber kutipan dan daftar pustaka.

Apabila ternyata di dalam naskah skripsi ini dapat dibuktikan terdapat unsur-unsur PLAGIASI, saya bersedia naskah skripsi ini digugurkan dan gelar yang saya peroleh dibatalkan, serta diproses sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian pernyataan ini saya buat dengan sebenarnya tanpa ada paksaan dari siapapun.

Batam, 02 Agustus 2023



Ayu Entini Lumban Raja

190210031

**IMPLEMENTASI DATA *MINING* DENGAN  
ALGORITMA *NAIVE BAYES* UNTUK KLASIFIKASI  
KELAYAKAN PENERIMA BANTUAN SEMBAKO**

**SKRIPSI**

**Untuk memenuhi salah satu syarat  
Memperoleh gelar Sarjana**

**Oleh:  
Ayu Entini Lumban Raja  
190210031**

**Telah disetujui oleh Pembimbing pada tanggal  
Seperti tertera di bawah ini**

**Batam, 28 Juli 2023**

  
**Koko Handoko, S.Kom., M. Kom  
Pembimbing**

## ABSTRAK

Batam memiliki jumlah penduduk sebanyak 1.196.396 jiwa perseptember 2020, namun terdapat penurunan laju pertumbuhan penduduk dalam kurun waktu 2010 hingga 2020 sebesar 2,32%. Dari jumlah keseluruhan penduduk Batam presentase jumlah penduduk penerima bantuan sebanyak 5,19 % angka ini meningkat dibandingkan maret 2022 sementara itu garis penerima bantuan tercatat tahun 2022 sebesar 783.730 jiwa perkapita perbulan. Jumlah rumah tangga miskin di Kecamatan Batu Aji dengan kelurahan Bukit Tempayang penduduk 15.857 jiwa, ruta miskin 300 jiwa, Kelurahan Buliang penduduk 37.531 jiwa, ruta miskin 289 jiwa, Kelurahan Kibling penduduk 28.693 jiwa, ruta miskin 1.057 jiwa, Kelurahan Tanjung Uncang penduduk 28.037 jiwa, ruta miskin 1.039 jiwa. Penerima bantuan merupakan suatu ketidakmampuan dalam memenuhi kebutuhan dasar termasuk makanan, pakaian, pendidikan dan perumahan. Dengan demikian di perlukan dilakukan sebuah strategi dalam penanggulangan tingkat penerima bantuan yaitu dengan menyediakan data penerima bantuan yang akurat dan tepat sasaran. Data *mining* dapat diartikan sebagai sekumpulan proses yang digunakan untuk mengeksplorasi dan mencari nilai berupa informasi terhadap relasi-relasi kompleks yang selama ini sudah tersimpan dalam basis data. *Naive Bayes* adalah algoritma yang ada pada data *mining* dan bagian dari teknnik klasifikasi data *mining* dengan menggunakan teknik probabilitas dan statistik untuk memperkirakan ataupun memprediksi peluang-peluang yang akan terjadi berdasarkan peluang sebelumnya yaitu terdapat dua kelas layak dan tidak layak. Hasil dari klasifikasi yang akan dilakukan nantinya akan membantu dalam pengolahan bantuan untuk membantu mengambil keputusan terkait klasifikasi penentuan penerima sembako. Berdasarkan data item set diatas dengan 25 data record maka di dapat diketahui bahwa terdapat 71 % dengan klasifikasi bantuan layak dan 100 % tidak layak dengan kepala keluarga laki-laki. Terdapat 29 % dengan klasifikasi bantuan sembako layak dan 0 % tidak layak dengan kepala keluarga perempuan. Berikut dibawahn ini probabilitas kriteria Kepala Keluarga. Dan dengan melakukan pengujian perhitungan secara manual dan menggunakan software rapid minner mendapatkan nilai acuracy 100%

**Kata kunci:** Data *mining*;Metode Naive Bayes;Klasifikasi;Sembako

## **ABSTRACT**

*Batam has a population of 1,196,396 as of September 2020, but there has been a decline in the population growth rate from 2010 to 2020 by 2.32%. Of the total population of Batam, the percentage of the population receiving assistance was 5.19%, this figure increased compared to March 2022, while the line of aid recipients was recorded in 2022 of 783,730 people per capita per month. The number of poor households in Batu Aji District is Bukit Tempayang Village with 15,857 inhabitants, 300 poor households, 37,531 inhabitants in Buliang Village, 289 poor households, 28,693 inhabitants of Kibling Village, 1,057 poor households, 28,037 inhabitants of Tanjung Uncang Village, 1,039 poor households. The beneficiary of assistance is an inability to meet basic needs including food, clothing, education and housing. Thus it is necessary to carry out a strategy in dealing with the level of beneficiaries, namely by providing accurate and targeted data on beneficiaries. Data mining can be interpreted as a set of processes that are used to explore and find value in the form of information on complex relationships that have been stored in databases so far. Naive Bayes is an algorithm that exists in data mining and is part of the data mining classification technique by using probability and statistical techniques to estimate or predict opportunities that will occur based on previous opportunities, namely there are two feasible and inappropriate classes. The results of the classification that will be carried out later will help in the processing of assistance to help make decisions regarding the classification of determining basic food recipients. Based on the data item set above with 25 data records, it can be seen that there are 71% classified as eligible assistance and 100% are not eligible with male family heads. There are 29% classified as proper basic food assistance and 0% are not eligible with female family heads. The following below is the probability criteria for the head of the family. And by *testing* calculations manually and using rapid minner software, you get an accuracy value of 100%.*

**Keywords:** *Classification; Data mining; Groceries; Naive Bayes method;*

## KATA PENGANTAR

Atas berkat dan rahmat Tuhan Yang Maha Esa yang telah melimpahkan segala rahmat dan karuniaNya, sehingga penulis dapat menyelesaikan laporan tugas akhir yang merupakan salah satu persyaratan untuk menyelesaikan program studi strata satu (S1) pada Program Studi Teknik Informatika Universitas Putera Batam.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Karena itu, kritik dan saran akan senantiasa penulis terima dengan senang hati. Dengan segala keterbatasan, penulis menyadari pula bahwa skripsi ini takkan terwujud tanpa bantuan, bimbingan dan dorongan dari berbagai pihak. Untuk itu, dengan segala kerendahan hati, penulis menyampaikan ucapan terimakasih kepada:

1. Ibu DR.Nur Elfi Husda,S.Kom.,M.Si Selaku Rektor Universitas Putera Batam;
2. Bapak Welly Sugiyanto,S.T.,M.M Selaku Dekan Fakultas Teknik dan Komputer;
3. Bapak Andi Maslan,S.T.,M.Si Selaku Ketua Program Studi Teknik Informatika;
4. Bapak Rahmat Fauzi, S.Kom., M.Kom Selaku Pembimbing Akademik pada Program Studi Teknik Informatika Universitas Putera Batam;
5. Bapak Koko Handoko, S.Kom., M.Kom Selaku Pembimbing Skripsi pada Program Studi Teknik Informatika Universitas Putera Batam;
6. Dosen dan Staff Universitas Putera Batam;
7. Selaku Kedua orang tua penulis yang selalu mendoakan dan memberikan dukungan sehingga skripsi ini dapat terselesaikan;
8. Keluarga penulis yang selalu memberikan dukungan dan motivasi kepada penulis;
9. Teman-teman seperjuangan yang bersedia membagi ilmu dan sharing pendapat;
10. Semua pihak yang telah bersedia meluangkan waktu, tenaga dan pikiran dalam memberikan data dan informasi selama penulis membuat skripsi ini yang tidak dapat disebutkan satu persatu.

Semoga Tuhan Yang Maha Esa membalas kebaikan dan selalu mencurahkan berkatNya, Amin.

Batam, 2 Agustus 2023

Ayu Entini Lumban Raja

## DAFTAR ISI

<b>HALAMAN SAMPUL</b> .....	i
<b>HALAMAN JUDUL</b> .....	ii
<b>SURAT PERNYATAAN ORISINALITAS</b> .....	iii
<b>HALAMAN PENGESAHAN</b> .....	iv
<b>ABSTRAK</b> .....	v
<b>ABSTRACT</b> .....	vi
<b>KATA PENGANTAR</b> .....	vii
<b>DAFTAR ISI</b> .....	viii
<b>DAFTAR GAMBAR</b> .....	x
<b>DAFTAR TABEL</b> .....	xi
<b>BAB I PENDAHULUAN</b> .....	1
1.1 Latar Belakang.....	1
1.2 Identifikasi Masalah.....	3
1.3 Batasan Masalah.....	4
1.4 Rumusan Masalah.....	4
1.5 Tujuan Penelitian.....	5
1.6 Manfaat Penelitian.....	5
1.6.1 Manfaat Teoritis.....	5
1.6.2 Manfaat Praktis.....	6
<b>BAB II TINJAUAN PUSTAKA</b> .....	7
2.1 Teori Dasar.....	7
2.2 KDD ( <i>Knowledge Discovery In Database</i> ).....	7
2.3 <i>Data Mining</i> .....	8
2.3.1 Langkah-langkah data <i>mining</i> .....	10
2.3.2 Metode Naive Bayes.....	11
2.3.3 Klasifikasi.....	13
2.4 Penerima bantuan.....	14
2.5 Bantuan Sembako.....	14
2.6 Objek Penelitian.....	15
2.7 <i>Software</i> Pendukung.....	16
2.7.1 <i>RapidMiner</i> .....	16
2.7.2 <i>Microsoft Excel</i> .....	17
2.8 Penelitian Terdahulu.....	18
2.9 Kerangka Pemikiran.....	21
<b>BAB III METODELOGI PENELITIAN</b> .....	22
3.1 Desain Penelitian.....	22
3.2 Teknik Pengumpulan Data.....	23
3.3 Operasional Variabel.....	24



3.4	Teknik Analisis Data .....	25
3.4.1	Data Preprocessing.....	26
3.4.2	Seleksi Data .....	30
3.4.3	Random Data .....	37
3.4.4	Klasifikasi Data .....	38
3.4.4.1	Naive Bayes.....	40
3.4.4.2	Accuracy Algoritma Naïve Bayes .....	40
3.5	Lokasi dan Jadwal Penelitian .....	41
3.4.1	Lokasi Penelitian.....	41
3.4.2	Jadwal Penelitian .....	41
<b>BAB IV HASIL DAN PEMBAHASAN.....</b>		<b>43</b>
4.1	Hasil Penelitian.....	43
4.1.1	Data Set.....	43
4.1.2	Seleksi Data.....	49
4.2	Analisa Proses Algoritama .....	52
4.2.1	Probabilitas Kepala Keluarga P (K) .....	54
4.2.1	Probabilitas Kondisi Rumah P (R) .....	54
4.2.1	Probabilitas Jumlah Penghasilan P (J) .....	55
4.2.1	Probabilitas Status Pemilik Rumah P (P).....	56
4.3	Hasil Pengujian.....	56
4.3	Implementasi <i>RapidMiner</i> .....	60
<b>BAB V KESIMPULAN DAN SARAN .....</b>		<b>68</b>
5.1	Kesimpulan.....	68
5.2	Saran .....	68
<b>DAFTAR PUSTAKA .....</b>		<b>70</b>
<b>LAMPIRAN .....</b>		<b>74</b>
1.	Pendukung Penelitian.....	74
2.	Daftar Riwayat Hidup .....	76
3.	Surat Keterangan Penelitian .....	77

## DAFTAR GAMBAR

<b>Gambar 2. 1</b> KDD(Kowladge Discovery in Database).....	8
<b>Gambar 2. 2</b> RapidMiner .....	17
<b>Gambar 2. 3</b> Microsoft excel.....	18
<b>Gambar 2. 4</b> Kerangka Pemikiran .....	21
<b>Gambar 3. 1</b> Desain Penelitian .....	22
<b>Gambar 3. 2</b> Lokasi Penelitian .....	41
<b>Gambar 4. 1</b> Nilai Akurasi .....	58
<b>Gambar 4. 2</b> Import Data Training .....	60
<b>Gambar 4. 3</b> Data Import Step 3.....	61
<b>Gambar 4. 4</b> Data Import Step 4.....	62
<b>Gambar 4. 5</b> Data Testing .....	62
<b>Gambar 4. 6</b> Data Import Step Akhir.....	63
<b>Gambar 4. 7</b> Menghubungkan Operator .....	64
<b>Gambar 4. 8</b> Icon Run.....	64
<b>Gambar 4. 9</b> Perhitungan Hasil .....	65
<b>Gambar 4. 10</b> Grafik Korelasi .....	66

## DAFTAR TABEL

<b>Tabel 3. 1</b>	Data Preprocessing.....	26
<b>Tabel 3. 2</b>	Golongan Kepala Rumah Tangga .....	29
<b>Tabel 3. 3</b>	Golongan Kondisi Rumah .....	29
<b>Tabel 3. 4</b>	Golongan Pekerjaan .....	29
<b>Tabel 3. 5</b>	Jumlah Penghasilan.....	29
<b>Tabel 3. 6</b>	Pemilik Rumah .....	29
<b>Tabel 3. 7</b>	Seleksi Data .....	30
<b>Tabel 3. 8</b>	Random/Pengacakan Data.....	37
<b>Tabel 3. 9</b>	Data Training .....	38
<b>Tabel 3. 10</b>	Data Testing.....	39
<b>Tabel 3. 12</b>	Jadwal Penelitian .....	42
<b>Tabel 4. 1</b>	Data item set .....	43
<b>Tabel 4. 2</b>	Keterangan data .....	45
<b>Tabel 4. 3</b>	Seleksi Data .....	50
<b>Tabel 4. 4</b>	Probabilitas Kelas .....	52
<b>Tabel 4. 5</b>	Data Training .....	52
<b>Tabel 4. 6</b>	Data Testing.....	53
<b>Tabel 4. 7</b>	Probabilitas Kriteria Kepala Keluarga .....	54
<b>Tabel 4. 8</b>	Probabilitas Kondisi Rumah.....	55
<b>Tabel 4. 9</b>	Probabilitas Jumlah Penghasilan.....	55
<b>Tabel 4. 10</b>	Probabilitas Pemilik Rumah .....	56
<b>Tabel 4. 11</b>	Hasil data testing.....	57
<b>Tabel 4. 12</b>	Class prediksi.....	57
<b>Tabel 4. 13</b>	Accuracy.....	58
<b>Tabel 4. 14</b>	Attribute Correlation .....	66

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Batam memiliki Jumlah penduduk perseptember 2020 adalah 1.196.396 jiwa namun rata-rata laju pertumbuhan penduduk pada periode 2010 - 2020 adalah 2,32%. Dari jumlah keseluruhan penduduk Batam presentase jumlah penduduk penerima bantuan sebanyak 5,19 % angka ini meningkat dibandingkan maret 2022 sementara itu garis penerima bantuan tercatat tahun 2022 sebesar 783.730 jiwa perkapita perbulan. Sementara penerima manfaat di Batam menjadi masalah yang cukup besar, Pulau Batam memiliki tingkat pertumbuhan penduduk yang cukup tinggi.

Menurut hasil sensus di Kelurahan Batu Aji, sebagian besar penduduk pendatang memiliki keahlian di berbagai bidang. Wilayah Batu Aji berkembang dengan industri merupakan area layanan komersial. Selain itu, wilayah kelurahan yang berkembang terbatas pada perdagangan barang kebutuhan sehari-hari yang dikemas di toko dan pedagang kecil lainnya. Jumlah rumah tangga miskin di Kecamatan Batu Aji dengan kelurahan Bukit Tempayang penduduk 15.857 jiwa, ruta miskin 300 jiwa, Kelurahan Buliang penduduk 37.531 jiwa, ruta miskin 289 jiwa, Kelurahan Kibling penduduk 28.693 jiwa, ruta miskin 1.057 jiwa, Kelurahan Tanjung Uncang penduduk 28.037 jiwa, ruta miskin 1.039 jiwa.

Penerima bantuan salah satu persoalan yang paling mendasar sehingga harus menjadi pusat perhatian pemerintah dari suatu negara sehingga menemukan

kesejahteraan. Penerima bantuan merupakan suatu ketidakmampuan dalam memenuhi kebutuhan dasar termasuk makanan, pakaian, pendidikan dan perumahan. Dengan demikian di perlukan dilakukan sebuah strategi dalam penanggulangan tingkat penerima bantuan yaitu dengan menyediakan data penerima bantuan yang akurat dan tepat sasaran (Damuri et al., 2021). Dengan data yang akurat tersebut pemerintah berupaya agar tingkat penerima bantuan di suatu daerah dapat teratasi, dengan cara memberikan dorongan berupa program bantuan pangan yang di tujukan terhadap masyarakat yang mengalami penerima bantuan.

Pemerintah memberikan sembako (sembilan bahan pokok), Pemanfaatan teknik data mining untuk menentukan kelayakan penerima bantuan sembako merupakan ide yang menarik dan berpotensi untuk meningkatkan efisiensi dan ketepatan sasaran dari program bantuan tersebut. Data mining adalah proses eksplorasi dan analisis data yang digunakan untuk menemukan pola, keterhubungan, dan informasi yang berharga dari kumpulan data yang besar dan kompleks.

Data mining merupakan serangkaian proses analisis dan eksplorasi data yang digunakan untuk menemukan pola, relasi, dan informasi berharga yang tersembunyi dalam basis data yang besar dan kompleks. Proses ini melibatkan penggunaan teknik statistik, matematika, dan kecerdasan buatan untuk menggali pengetahuan baru dari data yang ada. (Fajrin & Handoko, 2018). Dengan cara menggali informasi tersebut sehingga berbentuk pola yang menarik dari sumber data besar sehingga mendapatkan data informasi berupa pengetahuan statistik, matematika yang di inginkan dalam tindakan mengambil keputusan sehingga *mining* data dapat

membantu tahap perencanaan dengan memberikan informasi yang akurat untuk membuat prediksi berdasarkan pengalaman masa lalu untuk kondisi masa kini.

Fungsi prediksi dalam menentukan pola-pola tertentu dari sebuah data dan dari pola itu dapat di ketahui berbagai macam variabel. Sebuah metode yang dapat digunakan dalam memprediksi adalah dengan menggunakan algoritma *Naive Bayes*. Salah satu algoritma dalam data mining yang digunakan untuk teknik klasifikasi dengan menggunakan teknik probabilitas dan statistik adalah Naive Bayes. Algoritma Naive Bayes berdasarkan pada teorema probabilitas Bayes dan dapat digunakan untuk melakukan klasifikasi pada data dengan mengestimasi probabilitas dari kelas tertentu berdasarkan fitur-fitur atau atribut yang ada dalam data. yaitu terdapat dua kelas layak dan tidak layak. Hasil dari klasifikasi yang akan dilakukan nantinya akan membantu dalam pengolahan bantuan untuk membantu membuat keputusan mengenai klasifikasi yang mengidentifikasi penerima kebutuhan pokok.

## **1.2 Identifikasi Masalah**

Dari latar belakang di atas, maka dapat ditentukan masalah sebagai berikut:

1. Tingkat penerima bantuan di Batam mengalami kenaikan sehingga harus menjadi pusat perhatian pemerintah daerah.
2. Strategi dipimpin oleh pemerintah daerah untuk berupaya mengurangi penerima bantuan adalah dengan menyediakan data penerima bantuan yang akurat dan tepat sasaran namun hal ini belum dapat terlaksana dengan baik.

3. Program bantuan yang di buat oleh pemerintah untuk membantu dan mendorong perekonomian berupa sembako tidak tepat sasaran pada saat di lapangan.

### 1.3 Batasan Masalah

Agar penelitian tidak lari dari objek permasalahan maka peneliti perlu membatasi masalah dengan jelas dan tepat:

1. Data yang diambil diperoleh dari kantor kelurahan Buliang Batu aji - Batam Kepulauan Riau, dan data yang di peroleh berdasarkan data pemberian sembako tahun 2021-2022.
2. Data penelitian akan diubah menjadi informasi akurat dengan memanfaatkan metode algoritma *Naive Bayes*.
3. Penelitian ini menggunakan *data mining* untuk prediksi layak atau tidak layak terima sembako dan menggunakan *rapid minner* untuk proses pengujian data.

### 1.4 Rumusan Masalah

Rumusan masalah yang teridentifikasi dalam penelitian ini adalah sebagai berikut:

1. Bagaimana cara implementasi dari prediksi kelayakan terima bantuan sembako pada kelurahan batu aji batam?
2. Bagaimana cara agar data yang telah di peroleh dari daerah dapat menghasilkan pengetahuan yang baru untuk klasifikasi kelayakan bantuan sembako?

3. Bagaimana cara agar hasil perhitungan untuk klasifikasi kelayakan terima sembako tepat sasaran?

### **1.5 Tujuan Penelitian**

Adapun tujuan yang akan di capai oleh peneliti adalah sebagai berikut:

1. Untuk penerapan implementasi menggunakan metode algoritma *Naive Bayes* sehingga diharapkan dapat memperoleh data yang valid.
2. Untuk memperoleh data baru dalam menentukan klasifikasi kelayakan bantuan sembako menggunakan data *mining* sebagai metode pengambilan pendukung keputusan berdasarkan data tahun 2021-2022.
3. Untuk hasil implementasi menggunakan data *mining* dalam prediksi layak atau tidak layak terima sembako dan menggunakan *rapid minner* untuk proses pengujian data.

### **1.6 Manfaat Penelitian**

Adapun manfaat dari penelitian ini yaitu:

#### **1.6.1 Manfaat Teoritis**

1. Bagi Peneliti

Menambah pengetahuan dan memperkuat teori tentang penggunaan data mining dan algoritma *Naive Bayes* untuk menentukan kelayakan penerima sembako.



## 2. Bagi Pembaca

Berbagi informasi sebagai pengetahuan kepada semua pembaca mengenai peranan penting dalam menggunakan data *mining*.

## 3. Bagi Akademisi

Untuk memberikan gambaran kepada peneliti selanjutnya yang akan meneliti lebih dalam mengenai data *mining* khususnya menggunakan algoritma Naive Bayes.

### **1.6.2 Manfaat Praktis**

1. Diharapkan dengan dilakukan Penelitian ini dapat membantu para pemerintahan daerah untuk mendapatkan data yang valid sehingga program-program yang di lakukan dapat terlaksanakan dengan baik.
2. Diharapkan dengan penelitian ini menambah ilmu pengetahuan lebih lagi bagi peeliti mengenai *data mining* dan penggunaan algoritma *Naive Bayes*.
3. Diharapkan dengan penelitian ini pembagian sembako kepada masyarakat yang membutuhkan tepat sasaran.

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 Teori Dasar**

Teori dasar adalah teori-teori yang bersumber dari buku jurnal-jurnal yang berisi informasi dan dijadikan sebagai bahan referensi yang relevan untuk mendukung penelitian. Berikut di bawah ini adalah Teori yang digunakan dalam penelitian ini:

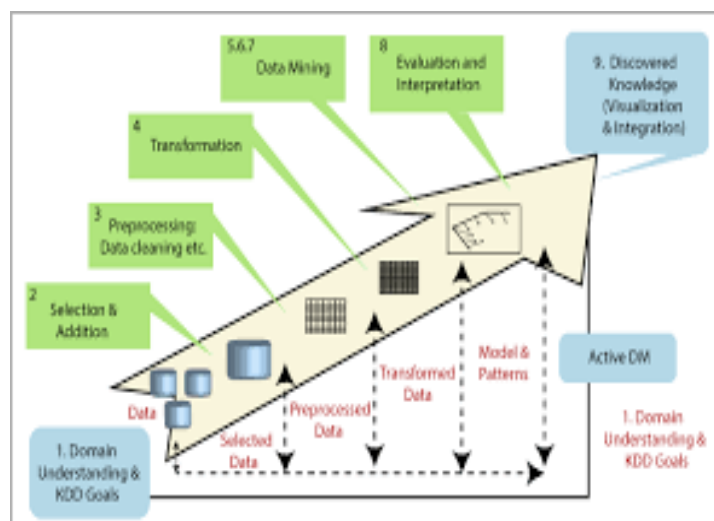
#### **2.2 KDD (*Knowledge Discovery In Database*)**

KDD adalah bagian dari sains yang dimana mencakup pada penyimpanan atau dan di jabarkan sebagai data proses guna untuk dapat mengenal pola-pola sebagai pengetahuan baru dari sekelompok data besar dan menghasilkan suatu informasi yang lebih mudah untuk dipahami. Berikut di bawah ini merupakan proses-proses yang dihasilkan dari pemurnian data, integrasi data dan presentasi pengetahuan (Wijaya & Dwiasnati, 2020).

Menurut (Indra Borman & Wati, 2020) KDD adalah bentuk proses yang memiliki beberapa rangkaian proses interaksi yang terurut salah satunya ialah data *mining*. Urutan langkah KDD di deskripsikan sebagai berikut:

1. Pembersihan data, adalah suatu tindakan yang dilakukan untuk menghilangkan data dan noise yang tidak konsisten.
2. Integrasi data adalah proses penggabungan data dalam jumlah besar dari berbagai sumber.

3. Seleksi data, merupakan pengambilan data dengan tujuan untuk di gunakan sebagai proses analisa dalam *data mining*.
4. Transformasi data, merupakan tahap transformasi data kedalam bentuk yang tepat.
5. *Data mining*, adalah proses pengaplikasian suatu cara untuk dapat mengekstrak pola-pola pada data.
6. Evaluasi pola, merupakan tahap pengidentifikasian beberapa pola yang menarik yang akan dijadikan presentasi pengetahuan.
7. Presentasi pengetahuan, bertujuan untuk dapat mempresentasikan kepada pengguna pengetahuan yang ditemukan dengan memvisualisasikan pengetahuan tersebut.



**Gambar 2. 1** KDD(Kowladge Discovery in Database)  
**Sumber:** (Wijaya & Dwiasnati, 2020)

### 2.3 Data Mining

*data mining* adalah metode yang penting dalam ilmu komputer yang digunakan untuk mencari pola-pola atau informasi berharga dari data yang besar

dan kompleks. Proses data mining mencakup berbagai teknik dan algoritma yang digunakan untuk menggali pengetahuan dari data yang ada. Dengan cara menggali informasi tersebut sehingga berbentuk pola yang menarik dari sumber data besar sehingga mendapatkan data informasi berupa pengetahuan statistik, matematika yang di inginkan dalam tindakan mengambil keputusan sehingga *data mining* dapat membantu tahap perencanaan dengan memberikan informasi yang akurat untuk membuat prediksi berdasarkan pengalaman masa lalu untuk kondisi saat ini.

Data mining adalah proses atau kegiatan yang melibatkan pengumpulan data yang berukuran besar dan kemudian melakukan analisis untuk mengekstraksi pola, hubungan, atau informasi yang berharga dari data tersebut. Hasil analisis data mining ini dapat digunakan untuk mendapatkan wawasan baru, membuat keputusan yang lebih informasi, atau untuk tujuan lain yang bermanfaat (Tyas et al., 2021). Dari beberapa pengertian diatas dapat disimpulkan Proses data mining melibatkan langkah-langkah analisis data yang kompleks, termasuk preprocessing data, eksplorasi data, penerapan algoritma data mining, dan evaluasi hasil. Data mining memanfaatkan teknik-teknik dari berbagai disiplin ilmu seperti statistika, kecerdasan buatan, machine learning, dan analisis data untuk mencari informasi yang berharga dari data yang ada.

*Data mining* ialah proses menganalisis dan memeriksa kumpulan informasi untuk menjalin koneksi yang tidak terduga dan menyajikannya dengan cara yang jelas dan bermanfaat bagi pemilik data. Hasil dari data mining dapat digunakan dalam berbagai bidang dan aplikasi, termasuk bisnis, keuangan, ilmu pengetahuan, pemasaran, kesehatan, dan banyak lagi. Dengan menggali pengetahuan dari data

yang besar dan kompleks, data mining membantu dalam mengidentifikasi pola-pola dan hubungan yang tidak selalu terlihat secara langsung, sehingga memungkinkan pengambilan keputusan yang lebih baik dan lebih efisien. Data *mining* adalah seperangkat prosedur yang digunakan untuk mengekstraksi pengetahuan yang sebelumnya belum diketahui secara manual dari kumpulan informasi (Surahman et al., 2023).

### 2.3.1 Langkah-langkah data *mining*

Dalam prosesnya data *mining* dibagi atas beberapa langkah. Adapun langkah tersebut sebagai berikut: (Riani et al., 2019) :

a. *Data Preprocessing*

Merupakan langkah yang pertama pada data pengecekan data set. Yang berfungsi sebagai proses pembersihan data dari suata. Membagi data base menjadi beberapa bagian data untuk membatasi ukuran data yang terlalu besar dan mempercepat proses analisis data saat menggunakan algoritma data mining.

b. Seleksi Data

Merupakan proses dari memilih data yang akan digunakan berdasarkan pada kelas-kelas data yang dibutuhkan

c. Pengacakan Data

Merupakan proses yang digunakan untuk mengacak data. Karena proses ini adalah proses pengacakan data selama algoritma berlangsung maka pengacakan data berdampak pada data lain.

d. **Klasifikasi Data**

Merupakan bagian dari suatu kegiatan pada sistem yang membantu beberapa kelompok atau golongan yang sering digunakan untuk memperkirakan kelas suatu objek yang levelnya tidak diketahui.

e. **Evaluasi Performa**

merupakan tahap dalam melihat keakuratan atau akurasi nilai pada metode yang digunakan.

### **2.3.2 Metode Naive Bayes**

*Naive Bayes* adalah algoritma yang ada pada data *mining* dan bagian dari teknik klasifikasi data *mining* dengan menggunakan teknik probabilitas dan statistika untuk memperkirakan ataupun memprediksi peluang-peluang yang akan terjadi berdasarkan peluang sebelumnya yaitu terdapat dua kelas layak dan tidak layak (Wijaya & Dwiasnati, 2020).

Berikut adalah alur lengkap dari metode Naive Bayes:

1. **Baca Data Training:** Langkah pertama adalah membaca data training, yaitu data yang digunakan untuk melatih model Naive Bayes. Data training terdiri dari contoh-contoh data dengan label kelas yang sudah diketahui.
2. **Hitung Jumlah dan Probabilitas:** Setelah membaca data training, langkah selanjutnya adalah menghitung jumlah data pada masing-masing kategori atau kelas dan menghitung probabilitas dari masing-masing kategori. Probabilitas ini menunjukkan seberapa sering data muncul dalam kategori tertentu dalam data training.

3. Data Numerik: Jika data yang digunakan adalah data numerik, maka langkah selanjutnya adalah mencari nilai mean (rata-rata) dan standar deviasi (deviasi standar) dari setiap parameter yang merupakan data numerik

Menurut (Surahman et al., 2023) Metode *Naive Bayes*, adalah prosedur statistik untuk menghasilkan produk probabilistik yang kuat, digunakan untuk menentukan kemungkinan atau preferensi seseorang dengan menggabungkan dan data *sampling* frekuensi dari kumpulan data yang tersedia. Algoritma yang menggunakan prinsip Bayes dapat meringkas setiap atribut secara mandiri atau tanpa mempengaruhi secara signifikan konsekuensi yang diberikan oleh nilai pada variabel kelas. Ketika digunakan untuk database dengan jumlah big data yang besar.

Pengklasifikasi *Naive Bayes*, juga dikenal sebagai pengklasifikasi *Bayes*, adalah pengklasifikasi statistik yang dapat digunakan untuk memprediksi probabilitas menjadi anggota suatu kelas. Teorema *Bayes*, yang menjadi dasar *Naive Bayes Classifier*, memberikan kemampuan klasifikasi yang seragam dengan *decision tree* serta *neural network*. Selain itu, *Naive Bayes Classifier* yang telah dicoba dan benar bekerja dengan sangat baik dan cepat saat digunakan dengan *database* yang berisi banyak data. Selain itu, *Naive Bayes Classifier* yang telah dicoba dan benar bekerja dengan sangat baik dan cepat saat digunakan dengan *database* yang berisi banyak data (Putri et al., 2021).

$$P(T|X) = P(X|T)(P(T0/P(X))$$

Rumus : Algoritma Naive Bayes

X = Data dengan class yang belum diketahui

$T$  = Hipotesis data  $X$  pada class spesifik

$P(T|X)$  = Probabilitas T berdasarkan kondisi

$P(X)$  = Probabilitas  $X$  berdasarkan kondisi hipotesis  $T$

$P(X)$  = Probabilitas dari  $T$

### 2.3.3 Klasifikasi

Klasifikasi merupakan proses di mana kita mencari model atau fungsi yang dapat membedakan dan menjelaskan kelas-kelas data atau objek berdasarkan fitur-fitur atau atribut yang dimilikinya. Model atau fungsi tersebut kemudian dapat digunakan untuk memprediksi kelas atau label dari objek yang label kelasnya tidak diketahui. Klasifikasi pada data *mining* adalah penempatan objek ke dalam kategori atau kelas yang telah ditetapkan sebelumnya (Wijaya & Dwiasnati, 2020). Klasifikasi sering digunakan untuk memprediksi kelas. Komponen-komponen utama dari proses klasifikasi yaitu:

1. Kelas, khususnya variabel dependen klasifikasi.
2. *Predictors*, yaitu variabel bebas dan sifat dari data yang dikategorikan.
3. *Training* data set yaitu *Predictors*, yaitu variabel bebas dan sifat dari data yang dikategorikan.
4. *Testing* data set yaitu data baru yang akan dikelompokkan untuk mengetahui akurasi dari model yang digunakan untuk proses klasifikasi.



## **2.4 Penerima bantuan**

Penerima bantuan adalah salah satu persoalan yang paling mendasar sehingga harus menjadi pusat perhatian pemerintah dari suatu negara sehingga menemukan kesejahteraan. penerima bantuan merupakan suatu ketidakmampuan dalam memenuhi kebutuhan dasar termasuk makanan, pakaian, pendidikan dan perumahan. Dengan demikian diperlukan dilakukan sebuah strategi dalam penanggulangan tingkat penerima bantuan yaitu dengan menyediakan data penerima bantuan yang akurat dan tepat sasaran (Sosial et al., 2023).

Dengan data yang akurat tersebut pemerintah berupaya agar tingkat penerima bantuan di suatu daerah dapat teratasi, dengan cara memberikan dorongan berupa program bantuan pangan yang di tujukan terhadap masyarakat yang mengalami penerima bantuan. Pemerintah memberikan sembako (sembilan bahan pokok) termasuk makanan dan minuman yang digunakan untuk kebutuhan sehari-hari setiap bulannya melalaui mekanisme perbankan.

## **2.5 Bantuan Sembako**

Bantuan sembako merupakan program perlindungan sosial pemerintah untuk individu yang kurang mampu, tidak mampu, dan perbatasan termasuk bantuan sosial. Karena tidak termasuk kontribusi atau premi dari penerima, bantuan sosial adalah barang publik yang ditawarkan kepada individu dan rumah tangga yang sangat miskin. Metode utama penyediaannya adalah dengan menerapkan ide solidaritas vertikal (Damuri et al., 2021).

Bantuan sembako atau disebut dengan bantuan sosial adalah bantuan yang dirancang untuk memberi manfaat bagi rumah tangga termiskin guna membantu mereka yang tidak mampu, mengurangi isolasi sosial, dan menghilangkan disinsentif tenaga kerja. Manfaat diberikan dalam bentuk program jaminan sosial, namun tidak ada program asuransi sosial dan tidak ada pembayaran atau sumbangan kepada penerima manfaat. Bantuan sosial diberikan tanpa persyaratan atau batasan yang rumit bagi penerimanya. Sama halnya dengan bantuan tenaga kerja, kebutuhan penerima hanya karena mereka sudah dipekerjakan oleh BPJS Ketenagakerjaan (Surahman et al., 2023).

bantuan sosial adalah bentuk bantuan yang diberikan oleh pemerintah, organisasi nirlaba, atau lembaga sosial kepada individu, keluarga, atau masyarakat yang membutuhkan lainnya yang membantu masyarakat dari kemiskinan untuk mencapai standar gizi minimal, standar hidup minimum, atau keduanya. Menurut gagasan FAO, dukungan sosial bagi karyawan dimaksudkan untuk memenuhi persyaratan agar pekerja dapat menjalani kehidupan yang terhormat. Pekerja harus dapat memenuhi kebutuhan dasarnya, termasuk kebutuhan gizi dan gizi untuk kesehatan keluarganya (Sosial et al., 2023).

## **2.6 Objek Penelitian**

Menurut hasil sensus di Kecamatan Batu Aji, sebagian besar warga pendatang yang melamar adalah yang ahli di berbagai bidang. Wilayah Batu Aji berkembang dengan industri merupakan area layanan komersial. Selain itu, kawasan kelurahan berkembang hanya diperbolehkan berdagang barang sehari-hari dalam kemasan di

toko-toko dan pedagang kecil lainnya. Jumlah rumah tangga miskin di Kecamatan Batu Aji dengan kelurahan Bukit Tempayang penduduk 15.857 jiwa, ruta miskin 300 jiwa, Kelurahan Buliang penduduk 37.531 jiwa, ruta miskin 289 jiwa, Kelurahan Kibling penduduk 28.693 jiwa, ruta miskin 1.057 jiwa, Kelurahan Tanjung Uncang penduduk 28.037 jiwa, ruta miskin 1.039 jiwa.

## **2.7 Software Pendukung**

Merupakan perangkat lunak yang digunakan pada penelitian ini yang mendukung proses penyelesaian penelitian. Berikut di bawah ini beberapa perangkat lunak yang digunakan:

### **2.7.1 RapidMiner**

*RapidMiner* adalah suatu perangkat lunak yang digunakan untuk mendukung suatu perhitungan dengan data *mining* karena memiliki berbagai sistem operasi. Perangkat lunak ini bersifat terbuka sehingga dapat digunakan oleh siapapun. *Rapidminer* pertama kali rilis dikenalkan dengan nama Yale yang memiliki kepanjangan *Yet another learning environment* pada tahun 2001 oleh Ralf Klinkenbert (Putri et al., 2021).

Menurut (Fajrin & Handoko, 2018) *Rapidminer* merupakan suatu perangkat lunak yang paling atas atau paling di pilih saat melakukan proses data *mining*. *Rapidminer* juga menyediakan GUI yang dapat mendefinisikan metode yang harus digunakan pengguna untuk memproses data.



**Gambar 2. 2** *RapidMiner*  
**Sumber:** (Putri et al., 2021)

### **2.7.2** *Microsoft Excel*

*Microsoft Excel* adalah bagian dari *Microsoft Office* yang berfungsi untuk dapat mengolah data berupa angka dan melakukan perhitungan. *Excel* digunakan untuk melakukan berbagai tugas analisis data, perhitungan matematika, dan menyajikan data dalam bentuk tabel, grafik, atau laporan. terdapat beberapa fungsi lain dari *microsoft excel* yaitu dapat membuat laporan keuangan lebih mudah, karena pada *microsoft excel* tersedia rumus yang dapat digunakan. Selain itu grafik data pada *microsoft excel* dapat memudahkan dalam memproses pembuatan tabel ataupun grafik (Putri et al., 2021).

*Microsoft Excel* merupakan salah satu program atau aplikasi yang merupakan bagian dari paket instalasi *Microsoft Office*, yang memiliki fungsi sebagai pengolahan bentuk angka menggunakan *spreadsheet* yang terdiri dari baris dan kolom untuk dapat mengeksekusi perintah. *Microsoft excel* telah didistribusikan sebagai *software* pengolahan bentuk angka yang mendunia dan multi platform.

Beberapa kelebihan *microsoft excel* sebagai berikut:

1. *User Interface* yang dapat dengan mudah dipahami
2. Kontapabilitas dengan berbagai platform atau sistem operasi.
3. Dapat dengan mudah dipelajari oleh pengguna awam.
4. Tersedia banyak lisensi.
5. Mempunyai ekstensi (.xls) terpopuler untuk *software spreadsheet*.
6. Dapat membaca ekstensi standart spreadsheet
7. Mendukung *visual basic*.



**Gambar 2. 3** *Microsoft excel*  
**Sumber:** (Putri et al., 2021)

## **2.8 Penelitian Terdahulu**

Penelitian terdahulu merupakan penelitian yang telah dilakukan oleh peneliti sebelumnya dan kemudian dijadikan kutipan pada penelitian ini.

1. Dari jurnal penelitian (Utomo & Purba, 2019) dengan judul “Penerapan Data *Mining* Pada Data Gempa Bumi Terhadap Potensi Tsunami Di Indonesia ” Prosiding Seminar Nasional *Riset Information Science* (SENARIS) ISSN 2686-0260 Setember 2019 memberikan kesimpulan bahwasanya Dengan menerapkan algoritma klasifikasi *Naive Bayes* pada data seismik kemungkinan tsunami di Indonesia, seseorang dapat mendeteksi

kemungkinan dampak gempa bumi. Dengan data eksperimen yang digunakan, efek yang dihasilkan adalah kemungkinan terjadinya tsunami.

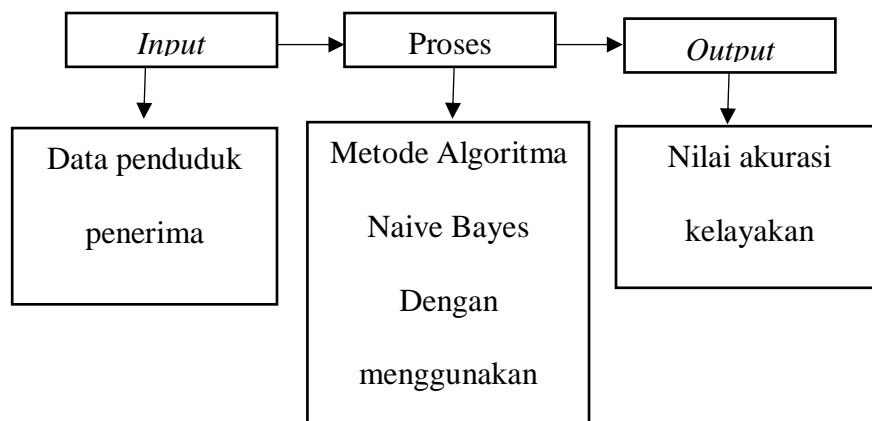
2. Dari jurnal penelitian (Pandie, 2019) dengan judul “Implementasi Algoritma Data Mining *Naive Bayes* Pada Koperasi” J-ICON vol 6 no 1 maret 2018 memberikan kesimpulan bahwa Algoritma Naive Bayes dapat digunakan untuk mengklasifikasikan kategori kredit nasabah yang ingin mengajukan pinjaman koperasi karena memiliki tingkat akurasi klasifikasi sebesar 94,77%.
3. Dari jurnal penelitian (Putri et al., 2021) dengan judul “Implementasi Data Mining Untuk Prediksi Penjualan Pestisida Pada CV Mitra Artha Sejati Menggunakan Algoritma *Naive Bayes*” eProsiding Teknik Informatika (PROTEKTIF) vol 1 no 1 november 2020 memberikan kesimpulan bahwa hasil prediksi menggunakan algoritma Naive Bayes dapat memberikan nilai tingkat akurasi sebesar 94.59% dengan akurasi kelas yaitu “YES” 100.00%, “NO” 94.44% untuk Class recall yaitu “YES” 33.33% sedangkan “NO” 100.00 % jumlah data training adalah 443 data dan 111 data untuk diuji. Sedangkan untuk perhitungan kedua menggunakan cross-validation menghasilkan tingkat akurasi sebesar 92,06%.
4. Dari jurnal penelitian (Wijaya & Dwiasnati, 2020) dengan judul “Implementasi Data Mining Dengan Algoritma *Naive Bayes* Pada Penjualan Obat” Jurnal Informatika Vol 7 no 1 april 2020 ISSN 2355-6579 memberikan kesimpulan bahwa Uji *recall* data penjualan obat dengan menggunakan algoritma *Naive Bayes* menghasilkan tingkat akurasi sebesar 88,00% dimana pada saat pengujian model data digunakan full dataset sebagai data test.
5. Dari jurnal penelitian (Sosial et al., 2023) dengan judul “Implementasi Algoritma C4.5 Untuk Menentukan Penerima Bantuan Sosial” JATI (Jurnal

ahasiswa Teknik Informatia Vol 7 No 1 Februari 2023 memberikan kesimpulan bahwa penerapan algoritma C4.5 dalam klasifikasi penerima bantuan sembako nontunai penerima sembako tunai dengan hasil pengujian akurasi memperoleh 58.18%.

6. Dari jurnal penelitian (Indra Borman & Wati, 2020) dengan judul “Penerapan Data Mining Dalam Klasifikasi Data Anggota Kopdit Sejahtera Bandar Lampung Dengan Algoritma *Naive Bayes*” Jurnal Ilmiah Fakultas Ilmu Komputer ISSN 1979-5254 memberikan kesimpulan bahwa dari pengujian menggunakan WEKA dan *RapidMiner* menggunakan 1064 record data training dan 300 record data test menghasilkan tingkat akurasi 70,33%, recall 70,33% dan akurasi 100%.
7. Dari jurnal penelitian (Fajrin & Handoko, 2018) dengan judul “Penerapan Data Mining Untuk Mengolah Tata Letak Buku Dengan Metode Association rule” Jurnal Ilmiah Informatika (JIF) ISSN 2337-8379 memberikan kesimpulan bahwa penerapan algoritma FP-Growth berdasarkan pada nilai *support* dan *confidence* pada data yang diolah dalam jumlah yang sangat besar, maka hal demikian akan lebih sulit sehingga perlu pemahaman tingkat *support* dan *confidence* dari data agar mendapatkan hasil yang maksimal.

## 2.9 Kerangka Pemikiran

Merupakan gambaran penelitian dalam bentuk kerangka yang telah di buat oleh peneliti dari awal hingga akhir proses penelitian sehingga dapat membantu peneliti dalam menyelesaikan penelitian tepat sasaran.



**Gambar 2. 4** Kerangka Pemikiran

**Sumber:** Data Penelitian 2023

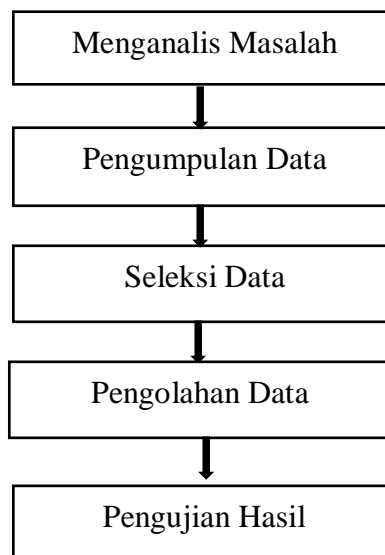


## BAB III

### METODELOGI PENELITIAN

#### 3.1 Desain Penelitian

Desain penelitian adalah rencana atau strategi yang digunakan oleh peneliti untuk menjawab pertanyaan penelitian dan mencapai tujuan penelitian.. Gambar 3.1 merupakan desain penelitian pada penelitian ini:



**Gambar 3. 1** Desain Penelitian  
**Sumber:** Data Penelitian 2023

Penjelasan gambar desain penelitian di atas sebagai berikut:

1. Menganalisis Masalah

Masalah yang telah ditemukan pada Kelurahan Buliang Batu Aji adalah Tingkat penerima bantuan di Batam mengalami kenaikan sehingga harus menjadi pusat perhatian pemerintah daerah.

## 2. Pengumpulan Data

Data yang akan terkumpul merupakan data penerima sembako, Data yang diambil berasal dari kantor Kelurahan Buliang Batu aji - Batam Kepulauan Riau, dan data yang di peroleh berdasarkan data pemberian sembako tahun 2021-2022.

## 3. Seleksi Data

Data yang telah dilakukan pengumpulan kemudian di seleksi sesuai kebutuhan penelitian yaitu, data nama, pekerjaan, tanggungan tahun 2021-2022.

## 4. Pengolahan Data

Data yang telah di seleksi akan dilakukan pengolahan dengan cara perhitungan hasil menggunakan data *mining* metode *Naive Bayes*.

## 5. Pengujian Hasil

Setelah data semua telah dikelola dengan baik maka langkah selanjutnya dilakukan pengujian apakah sudah sesuai dengan tujuan penelitian yaitu ketetapan sasaran kelayakan penerima sembako dengan menggunakan software pendukung yaitu *RapidMiner*.

### **3.2 Teknik Pengumpulan Data**

Merupakan cara yang digunakan peneliti untuk memperoleh data atau informasi mengenai data penelitian. Berikut adalah teknik pengumpulan data yang digunakan oleh peneliti:

1. Wawancara

Merupakan teknik yang dilakukan peneliti dengan langsung bertanya kepada narasumber yaitu Bapak kelurahan batu aji dan meminta izin data penerima sembako di jadikan data penelitian

2. Observasi

Merupakan pengumpulan data dibuat oleh peneliti dengan cara terjun langsung ke lapangan yaitu kantor kelurahan batu aji, dan melihat kondisi yang ada didalam.

3. Studi Pustaka

Merupakan cara yang dilakukan peneliti dengan mencari informasi melibatkan penelitian dari jurnal dan buku.

### **3.3 Operasional Variabel**

Variabel eksekutif merupakan cara untuk menentukan variabel dan indikator mana yang memiliki hubungan yang berbeda satu sama lain dalam suatu penelitian. Berikut adalah variabel-variabel yang digunakan dalam penelitian ini:

1. Nama

Merupakan panggilan yang digunakan untuk pengenalan diri.

2. Kepala Rumah Tangga

Orang yang mencari nafkah yang bertanggung jawab atas kebutuhan dan keperluan pada keluarga tersebut baik laki-laki atau perempuan.

3. Kondisi Rumah

Merupakan suatu bentuk kediaman yang ditempatin oleh keluarga tersebut dengan kategori batu permanen dan batu gamping.

4. Pekerjaan

Merupakan usaha dari perorangan untuk mendapatkan ubah atau gaji dari usaha tersebut, dalam kategori ini adalah karyawan, pedagang, guru honor dan wiraswata.

5. Jumlah Penghasilan

Merupakan suatu pendapatan yang diperoleh dari bekerja dengan kategori rendah, sedang dan tinggi.

6. Status Pemilik Rumah

Merupakan hak kepemilikan atas kediaman tersebut dalam kategori milik Pribadi atau sewa.

7. Status Kelayakan

Merupakan status yang menjadi patokan untuk seseorang mendapatkan bantuan sembako dan biasanya dilihat dari beberapa faktor yang mempengaruhi seperti keterangan di atas.

### **3.4 Teknik Analisis Data**

Teknik analisis data digunakan untuk dapat memahami alur algoritma pada naive bayes. Adapun tahapan tersebut bersifat interaktif, pengguna terlibat langsung atau melalui perantaraan base. Tahapan-tahapan tersebut antara lain sebagai berikut:

### 3.4.1 Data Preprocessing

Langkah yang pertama kali dilakukan adalah dengan melakukan pengecekan terhadap dataset penerima sembako yaitu melakukan tahapan preprocessing. Data preprocessing adalah proses membersihkan data dari noise teknik ini dilakukan untuk mengatasi ukuran database yang besar dengan membagi database menjadi beberapa bagian sehingga akan mempercepat proses scanning data saat algoritma data mining diterapkan.

**Tabel 3. 1** *Data Preprocessing*

No	Nama	Kepala Rumah Tangga	Kondisi Rumah	Pekerjaan	Jumlah Penghasilan	Status pemilik rumah
1	Adi Sudrajat	Laki-laki	Batu permanen	Karyawan	Rendah	Milik Pribadi
2	Alil	Laki-laki	Batu permanen	Karyawan	Sedang	Milik Pribadi
3	Amir	Laki-laki	Batu permanen	Karyawan	Sedang	Sewa
4	Andre	Laki-laki	Gamping	Pedagang	Rendah	Sewa
5	Anang Sitinjak	Laki-laki	Gamping	Guru (Honor)	Sedang	Sewa
6	Agus Saputra	Laki-laki	Batu permanen	Karyawan	Sedang	Milik Pribadi
7	Agus Salim	Laki-laki	Gamping	Karyawan	Sedang	Milik Pribadi
8	Agung Sudrajat	Laki-laki	Batu permanen	Karyawan	Tinggi	Milik Pribadi
9	Arslan	Laki-laki	Batu permanen	Pedagang	Tinggi	Milik Pribadi
10	Ardam	Laki-laki	Batu permanen	Pedagang	Tinggi	Milik Pribadi
11	Arman	Laki-laki	Gamping	Pedagang	Tinggi	Milik Pribadi
12	Arish	Laki-laki	Gamping	Karyawan	Sedang	Milik Pribadi
13	Arifin	Laki-laki	Gamping	Karyawan	Sedang	Milik Pribadi
14	Astari putri	Perempuan	Batu permanen	Guru (Honor)	Sedang	Sewa
15	Arsalan	Laki-laki	Batu permanen	Karyawan	Rendah	Milik Pribadi
16	Burhan ramadan	Laki-laki	Batu permanen	Karyawan	Rendah	Milik Pribadi
17	Basid ahlan	Laki-laki	Gamping	Guru (Honor)	Sedang	Milik Pribadi
18	Bania laia	Perempuan	Gamping	Guru (Honor)	Sedang	Sewa
19	Bahir samsul	Laki-laki	Batu permanen	Karyawan	Sedang	Milik Pribadi
20	Binyamin	Laki-laki	Batu permanen	Karyawan	Rendah	Milik Pribadi
21	Badi ananda	Laki-laki	Gamping	Karyawan	Tinggi	Milik Pribadi

22	Basharat	Laki-laki	Gamping	Karyawan	Tinggi	Milik Pribadi
23	Calvin	Laki-laki	Gamping	Pedagang	Tinggi	Milik Pribadi
24	Charles alwi	Laki-laki	Batu permanen	Pedagang	Rendah	Sewa
25	Charlos	Laki-laki	Batu permanen	Karyawan	Rendah	Sewa
26	Casildo	Laki-laki	Batu permanen	Karyawan	Rendah	Milik Pribadi
27	Caspar	Laki-laki	Batu permanen	Karyawan	Sedang	Sewa
28	Ceilo	Laki-laki	Gamping	Wiraswasta	Sedang	Sewa
29	Chairil	Laki-laki	Gamping	Wiraswasta	Rendah	Sewa
30	Chiko	Laki-laki	Gamping	Wiraswasta	Sedang	Sewa
31	Chistoper	Laki-laki	Batu permanen	Wiraswasta	Sedang	Milik Pribadi
32	Cival	Laki-laki	Batu permanen	Pedagang	Sedang	Milik Pribadi
33	Conrad	Laki-laki	Batu permanen	Karyawan	Tinggi	Milik Pribadi
34	Cullen	Laki-laki	Gamping	Karyawan	Tinggi	Milik Pribadi
35	Devan	Laki-laki	Batu permanen	Guru (Honor)	Sedang	Sewa
36	Dani	Laki-laki	Batu permanen	Karyawan	Rendah	Milik Pribadi
37	Daniel sitorus	Laki-laki	Gamping	Karyawan	Sedang	Milik Pribadi
38	Defan	Laki-laki	Gamping	Guru (Honor)	Rendah	Milik Pribadi
39	Delvin	Laki-laki	Gamping	Guru (Honor)	Sedang	Milik Pribadi
40	Danish	Laki-laki	Batu permanen	Karyawan	Sedang	Milik Pribadi
41	Dalil	Laki-laki	Batu permanen	Karyawan	Sedang	Milik Pribadi
42	Dananjaya	Laki-laki	Batu permanen	Karyawan	Sedang	Sewa
43	Darib	Laki-laki	Gamping	Karyawan	Tinggi	Milik Pribadi
44	Daris	Laki-laki	Gamping	Wiraswasta	Tinggi	Milik Pribadi
45	Dary	Laki-laki	Batu permanen	Wiraswasta	Tinggi	Milik Pribadi
46	Eric	Laki-laki	Gamping	Wiraswasta	Rendah	Sewa
47	Emil	Laki-laki	Gamping	Pedagang	Sedang	Sewa
48	Elvis	Laki-laki	Gamping	Pedagang	Sedang	Sewa
49	Eliot	Laki-laki	Batu permanen	Wiraswasta	Rendah	Milik Pribadi
50	Ewander	Laki-laki	Batu permanen	Wiraswasta	Sedang	Sewa
51	Elkana	Perempuan	Batu permanen	Wiraswasta	Sedang	Sewa
52	Edwin	Laki-laki	Gamping	Wiraswasta	Sedang	Sewa
53	Emanuel	Laki-laki	Gamping	Pedagang	Rendah	Milik Pribadi
54	Elian	Laki-laki	Batu permanen	Pedagang	Sedang	Milik Pribadi
55	Evan	Laki-laki	Batu permanen	Pedagang	Sedang	Sewa
56	Elvin	Laki-laki	Gamping	Karyawan	Rendah	Milik Pribadi
57	Fadhillah	Perempuan	Gamping	Karyawan	Sedang	Milik Pribadi

58	Fatur	Laki-laki	Gamping	Pedagang	Sedang	Milik Pribadi
59	Fatlan	Laki-laki	Batu permanen	Pedagang	Sedang	Sewa
60	Feronika	Perempuan	Batu permanen	Karyawan	Tinggi	Sewa
61	Fernanda	Laki-laki	Batu permanen	Karyawan	Tinggi	Milik Pribadi
62	Ferdinan	Laki-laki	Gamping	Pedagang	Tinggi	Milik Pribadi
63	Feni	Perempuan	Batu permanen	Wiraswasta	Rendah	Sewa
64	Farid	Laki-laki	Batu permanen	Wiraswasta	Sedang	Milik Pribadi
65	Fanya	Laki-laki	Gamping	Wiraswasta	Sedang	Milik Pribadi
66	Fatimah	Perempuan	Gamping	Guru (Honor)	Rendah	Milik Pribadi
67	Ganimi	Laki-laki	Gamping	Karyawan	Sedang	Sewa
68	Gania	Laki-laki	Batu permanen	Karyawan	Sedang	Sewa
69	Gareth	Laki-laki	Batu permanen	Karyawan	Sedang	Milik Pribadi
70	Gregory	Laki-laki	Batu permanen	Karyawan	Tinggi	Milik Pribadi
71	Hamdan	Laki-laki	Gamping	Karyawan	Rendah	Sewa
72	Hafizah	Laki-laki	Batu permanen	Karyawan	Sedang	Milik Pribadi
73	Hamal	Laki-laki	Batu permanen	Karyawan	Sedang	Milik Pribadi
74	Herna	Perempuan	Gamping	Karyawan	Rendah	Milik Pribadi
75	Haris	Laki-laki	Gamping	Karyawan	Sedang	Milik Pribadi
76	Ibnu abas	Laki-laki	Gamping	Karyawan	Sedang	Sewa
77	Ifaaz	Laki-laki	Batu permanen	Wiraswasta	Sedang	Milik Pribadi
78	Ilham	Laki-laki	Batu permanen	Wiraswasta	Tinggi	Milik Pribadi
79	Irwanda	Laki-laki	Batu permanen	Wiraswasta	Tinggi	Milik Pribadi
80	Irhakim	Laki-laki	Gamping	Karyawan	Rendah	Milik Pribadi
81	Ilma	Laki-laki	Batu permanen	Pedagang	Sedang	Milik Pribadi
82	Isna wati	Perempuan	Batu permanen	Pedagang	Sedang	Sewa
83	Iskandar	Laki-laki	Gamping	Karyawan	Rendah	Milik Pribadi
84	Julia	Perempuan	Gamping	Karyawan	Sedang	Milik Pribadi
85	Jeno	Laki-laki	Gamping	Pedagang	Sedang	Milik Pribadi
86	Juneidi	Laki-laki	Batu permanen	Karyawan	Sedang	Milik Pribadi
87	Jamal	Laki-laki	Batu permanen	Karyawan	Tinggi	Milik Pribadi
88	Jekson	Laki-laki	Batu permanen	Pedagang	Tinggi	Sewa
89	Jainudin	Laki-laki	Gamping	Pedagang	Rendah	Milik Pribadi
90	James	Laki-laki	Batu permanen	Karyawan	Sedang	Milik Pribadi
91	Karen	Perempuan	Batu permanen	Pedagang	Sedang	Milik Pribadi
92	Kartini	Perempuan	Gamping	Pedagang	Rendah	Sewa
93	Kania	Perempuan	Gamping	Karyawan	Sedang	Sewa
94	Kamal	Laki-laki	Gamping	Wiraswasta	Sedang	Milik Pribadi

95	Lukman	Laki-laki	Batu permanen	Wiraswasta	Rendah	Sewa
96	Lex ardiantara	Laki-laki	Batu permanen	Karyawan	Sedang	Milik Pribadi
97	Miswan	Laki-laki	Batu permanen	Pedagang	Sedang	Milik Pribadi
98	Michael	Laki-laki	Gamping	Pedagang	Rendah	Milik Pribadi
99	Messi	Laki-laki	Gamping	Karyawan	Sedang	Sewa
100	Mia	Perempuan	Gamping	Karyawan	Sedang	Sewa

**Sumber:** Data Penelitian 2023

Semua data dikonversi kedalam bentuk numerik dengan data sebagai berikut:

**Tabel 3. 2** Golongan Kepala Rumah Tangga

Feature Kepala Rumah Tangga	Nilai Numerik
Laki-laki	1
Perempuan	2

**Sumber:** Data Penelitian 2023

**Tabel 3. 3** Golongan Kondisi Rumah

Feature Kondisi Rumah	Nilai Numerik
Batu permanen	1
Gamping	2

**Sumber:** Data Penelitian 2023

**Tabel 3. 4** Golongan Pekerjaan

Feature Pekerjaan	Nilai Numerik
Guru (honorar)	1
Karyawan	2
Pedagang	3
Wiraswasta	4

**Sumber:** Data Penelitian 2023

**Tabel 3. 5** Jumlah Penghasilan

Feature Jumlah Penghasilan	Nilai Numerik
Tinggi	1
Sedang	2
Rendah	3

**Sumber:** Data Penelitian 2023

**Tabel 3. 6** Pemilik Rumah

Feature Pemilik rumah	Nilai Numerik
Milik Pribadi	1
Sewa	2

**Sumber:** Data Penelitian 2023



Sehingga penentuan layak menerima bantuan sembako, Jika jumlah tanggungan  $\leq 1$ , dan jika kepala rumah tangga laki-laki dengan kondisi rumah batu permanen, dan pekerjaan sebagai karyawan, dan memiliki penghasilan tinggi, dan status kepemilikan rumah milik sendiri maka layak mendapatkan bantuan sembako sebaliknya tidak layak.

### 3.4.2 Seleksi Data

Untuk mendapatkan tingkat akurasi yang maksimal maka dilakukan seleksi *feature* dengan *cross validation*. Metode yang digunakan adalah *weight by correlation*. Proses ini merupakan *preprocessing* data untuk dijadikan sebagai data uji yang relevan terhadap pembagian data sembako yaitu dengan 100 data record dan dengan kelas kepala keluarga, kondisi rumah, Jumlah penghasilan dan Status Pemilik. Berikut merupakan tabel hasil seleksi data:

**Tabel 3. 7** Seleksi Data

Jumah_Tanggungan	Kepala_Rumah_Tangga	Kondisi_Rumah	Pekerjaan	Jumlah_Penghasilan	Status_pemilik_rumah	Jumlah_Golongan	Status Kelayakan
0	1	1	3	1	1	7	Layak
2	1	1	3	2	1	10	Tidak Layak
1	1	1	3	2	0	8	Tidak Layak
2	1	0	2	1	0	6	Layak
2	1	0	0	2	0	5	Layak
2	1	1	3	2	1	10	Tidak Layak
1	1	0	3	2	1	8	Tidak Layak
2	1	1	3	3	1	11	Tidak Layak
2	1	1	2	3	1	10	Tidak Layak
2	1	1	2	3	1	10	Tidak Layak

2	1	0	2	3	1	9	Tidak Layak
2	1	0	3	2	1	9	Tidak Layak
2	1	0	3	2	1	9	Tidak Layak
1	0	1	0	2	0	4	Layak
2	1	1	3	1	1	9	Tidak Layak
2	1	1	3	1	1	9	Tidak Layak
2	1	0	0	2	1	6	Layak
2	0	0	0	2	0	4	Layak
2	1	1	3	2	1	10	Tidak Layak
1	1	1	3	1	1	8	Tidak Layak
2	1	0	3	3	1	10	Tidak Layak
0	1	0	3	3	1	8	Tidak Layak
1	1	0	2	3	1	8	Tidak Layak
1	1	1	2	1	0	6	Layak
2	1	1	3	1	0	8	Tidak Layak
2	1	1	3	1	1	9	Tidak Layak
2	1	1	3	2	0	9	Tidak Layak
2	1	0	1	2	0	6	Layak
2	1	0	1	1	0	5	Layak
2	1	0	1	2	0	6	Layak
2	1	1	1	2	1	8	Tidak Layak
2	1	1	2	2	1	9	Tidak Layak
2	1	1	3	3	1	11	Tidak Layak
1	1	0	3	3	1	9	Tidak Layak
2	1	1	0	2	0	6	Layak
2	1	1	3	1	1	9	Tidak Layak
1	1	0	3	2	1	8	Tidak Layak
2	1	0	0	1	1	5	Layak
2	1	0	0	2	1	6	Layak
2	1	1	3	2	1	10	Tidak Layak

2	1	1	3	2	1	10	Tidak Layak
2	1	1	3	2	0	9	Tidak Layak
2	1	0	3	3	1	10	Tidak Layak
1	1	0	1	3	1	7	Layak
2	1	1	1	3	1	9	Tidak Layak
2	1	0	1	1	0	5	Layak
2	1	0	2	2	0	7	Layak
2	1	0	2	2	0	7	Layak
2	1	1	1	1	1	7	Layak
1	1	1	1	2	0	6	Layak
2	0	1	1	2	0	6	Layak
2	1	0	1	2	0	6	Layak
2	1	0	2	1	1	7	Layak
2	1	1	2	2	1	9	Tidak Layak
2	1	1	2	2	0	8	Tidak Layak
2	1	0	3	1	1	8	Tidak Layak
2	0	0	3	2	1	8	Tidak Layak
2	1	0	2	2	1	8	Tidak Layak
2	1	1	2	2	0	8	Tidak Layak
2	0	1	3	3	0	9	Tidak Layak
2	1	1	3	3	1	11	Tidak Layak
1	1	0	2	3	1	8	Tidak Layak
2	0	1	1	1	0	5	Layak
2	1	1	1	2	1	8	Tidak Layak
2	1	0	1	2	1	7	Layak
2	0	0	0	1	1	4	Layak
2	1	0	3	2	0	8	Tidak Layak
2	1	1	3	2	0	9	Tidak Layak
2	1	1	3	2	1	10	Tidak Layak
1	1	1	3	3	1	10	Tidak Layak

2	1	0	3	1	0	7	Layak
2	1	1	3	2	1	10	Tidak Layak
2	1	1	3	2	1	10	Tidak Layak
2	0	0	3	1	1	7	Layak
2	1	0	3	2	1	9	Tidak Layak
2	1	0	3	2	0	8	Tidak Layak
2	1	1	1	2	1	8	Tidak Layak
1	1	1	1	3	1	8	Tidak Layak
2	1	1	1	3	1	9	Tidak Layak
2	1	0	3	1	1	8	Tidak Layak
2	1	1	2	2	1	9	Tidak Layak
2	0	1	2	2	0	7	Layak
2	1	0	3	1	1	8	Tidak Layak
2	0	0	3	2	1	8	Tidak Layak
1	1	0	2	2	1	7	Layak
2	1	1	3	2	1	10	Tidak Layak
2	1	1	3	3	1	11	Tidak Layak
2	1	1	2	3	0	9	Tidak Layak
2	1	0	2	1	1	7	Layak
1	1	1	3	2	1	9	Tidak Layak
2	0	1	2	2	1	8	Tidak Layak
2	0	0	2	1	0	5	Layak
2	0	0	3	2	0	7	Layak
2	1	0	1	2	1	7	Layak
2	1	1	1	1	0	6	Layak
1	1	1	3	2	1	9	Tidak Layak
2	1	1	2	2	1	9	Tidak Layak
2	1	0	2	1	1	7	Layak
2	1	0	3	2	0	8	Tidak Layak
2	0	0	3	2	0	7	Layak

1	1	1	3	2	0	8	Tidak Layak
2	1	0	2	1	0	6	Layak
2	1	0	0	2	0	5	Layak
2	1	1	3	2	1	10	Tidak Layak
1	1	0	3	2	1	8	Tidak Layak
2	1	1	3	3	1	11	Tidak Layak
2	1	1	2	3	1	10	Tidak Layak
2	1	1	2	3	1	10	Tidak Layak
2	1	0	2	3	1	9	Tidak Layak
2	1	0	3	2	1	9	Tidak Layak
2	1	0	3	2	1	9	Tidak Layak
1	0	1	0	2	0	4	Layak
2	1	1	3	1	1	9	Tidak Layak
2	1	1	1	2	1	8	Tidak Layak
1	1	1	1	3	1	8	Tidak Layak
2	1	1	1	3	1	9	Tidak Layak
2	1	0	1	1	1	6	Layak
2	1	1	1	2	1	8	Tidak Layak
2	0	1	1	2	0	6	Layak
2	1	0	3	1	1	8	Tidak Layak
2	0	0	3	2	1	8	Tidak Layak
1	1	0	0	2	1	5	Layak
2	1	1	3	2	1	10	Tidak Layak
2	1	1	3	3	1	11	Tidak Layak
2	1	1	0	3	0	7	Layak
2	1	0	0	1	1	5	Layak
1	1	1	3	2	1	9	Tidak Layak
2	0	1	3	2	1	9	Tidak Layak
2	0	0	0	1	0	3	Layak
2	0	0	3	2	0	7	Layak

1	1	1	1	3	1	8	Tidak Layak
2	1	1	1	3	1	9	Tidak Layak
2	1	0	3	1	1	8	Tidak Layak
2	1	1	3	2	1	10	Tidak Layak
2	0	1	3	2	0	8	Tidak Layak
2	1	0	3	1	1	8	Tidak Layak
2	0	0	3	2	1	8	Tidak Layak
1	1	0	3	2	1	8	Tidak Layak
2	1	1	3	2	1	10	Tidak Layak
2	1	1	3	3	1	11	Tidak Layak
2	1	1	3	3	0	10	Tidak Layak
2	1	0	3	1	1	8	Tidak Layak
1	1	1	3	2	1	9	Tidak Layak
2	0	1	3	2	1	9	Tidak Layak
2	0	0	0	1	0	3	Layak
2	0	0	3	2	0	7	Layak
2	1	0	1	2	1	7	Layak
2	1	1	1	1	0	6	Layak
1	1	1	3	2	1	9	Tidak Layak
2	1	1	0	2	1	7	Layak
2	1	0	0	1	1	5	Layak
2	1	0	3	2	0	8	Tidak Layak
2	0	0	3	2	0	7	Layak
1	1	1	3	2	0	8	Tidak Layak
2	1	0	2	1	0	6	Layak
2	1	0	0	2	0	5	Layak
2	1	1	3	2	1	10	Tidak Layak
1	1	0	3	2	1	8	Tidak Layak
2	1	1	3	3	1	11	Tidak Layak
2	1	1	2	3	1	10	Tidak Layak

2	1	1	2	3	1	10	Tidak Layak
2	1	0	2	3	1	9	Tidak Layak
2	1	0	3	2	1	9	Tidak Layak
2	1	0	3	2	1	9	Tidak Layak
1	0	1	0	2	0	4	Layak
2	1	1	3	1	1	9	Tidak Layak
1	1	1	3	2	0	8	Tidak Layak
2	1	0	2	1	0	6	Layak
2	1	0	0	2	0	5	Layak
2	1	1	3	2	1	10	Tidak Layak
1	1	0	3	2	1	8	Tidak Layak
2	1	1	3	3	1	11	Tidak Layak
2	1	1	2	3	1	10	Tidak Layak
2	1	1	2	3	1	10	Tidak Layak
2	1	0	2	3	1	9	Tidak Layak
2	1	0	3	2	1	9	Tidak Layak
2	1	0	3	2	1	9	Tidak Layak
2	1	0	3	1	1	8	Tidak Layak
2	1	1	0	2	1	7	Layak
2	0	1	0	2	0	5	Layak
2	1	0	3	1	1	8	Tidak Layak
2	0	0	3	2	1	8	Tidak Layak
1	1	0	0	2	1	5	Layak
2	1	1	3	2	1	10	Tidak Layak
2	1	1	3	3	1	11	Tidak Layak
2	1	1	0	3	0	7	Layak
2	1	0	0	1	1	5	Layak
1	1	1	3	2	1	9	Tidak Layak
2	0	1	3	2	1	9	Tidak Layak
2	0	0	0	1	0	3	Layak

2	0	0	3	2	0	7	Layak
2	1	0	1	2	1	7	Layak
2	1	1	1	1	0	6	Layak
1	1	1	3	2	1	9	Tidak Layak
2	1	1	0	2	1	7	Layak
1	1	0	3	2	1	8	Tidak Layak
2	1	1	3	2	1	10	Tidak Layak
2	1	1	3	3	1	11	Tidak Layak
2	1	1	3	3	0	10	Tidak Layak
2	1	0	3	1	1	8	Tidak Layak

**Sumber:** Data Penelitian 2023

### 3.4.3 Random Data

Merupakan proses pengacakan data pada dataset penerima sembako yang mempengaruhi performa dari status kelayakan hal tersebut disebabkan karena adanya model klasifikasi yang dilakukan selama pelatihan data. Penerapan pengacakan data terhadap data akan dilakukan setelah hasil preprocessing dan seleksi. Maka dihasilkan 25 data record. Berikut ini tabel pengacakan atau random data:

**Tabel 3. 8** Random/Pengacakan Data

Kepala Rumah Tangga	Kondisi Rumah	Jumlah Penghasilan	Status pemilik rumah
Laki-laki	Batu permanen	Rendah	Milik Pribadi
Laki-laki	Batu permanen	Sedang	Milik Pribadi
Laki-laki	Batu permanen	Sedang	Sewa
Laki-laki	Gamping	Rendah	Sewa
Laki-laki	Gamping	Sedang	Sewa
Laki-laki	Batu permanen	Sedang	Milik Pribadi
Laki-laki	Gamping	Sedang	Milik Pribadi
Laki-laki	Batu permanen	Tinggi	Milik Pribadi



Laki-laki	Batu permanen	Tinggi	Milik Pribadi
Laki-laki	Batu permanen	Tinggi	Milik Pribadi
Laki-laki	Gamping	Tinggi	Milik Pribadi
Laki-laki	Gamping	Sedang	Milik Pribadi
Laki-laki	Gamping	Sedang	Milik Pribadi
Perempuan	Batu permanen	Sedang	Sewa
Laki-laki	Batu permanen	Rendah	Milik Pribadi
Laki-laki	Batu permanen	Rendah	Milik Pribadi
Laki-laki	Gamping	Sedang	Milik Pribadi
Perempuan	Gamping	Sedang	Sewa
Laki-laki	Batu permanen	Sedang	Milik Pribadi
Laki-laki	Batu permanen	Rendah	Milik Pribadi
Laki-laki	Gamping	Tinggi	Milik Pribadi
Laki-laki	Gamping	Tinggi	Milik Pribadi
Laki-laki	Gamping	Tinggi	Milik Pribadi
Laki-laki	Batu permanen	Rendah	Sewa
Laki-laki	Batu permanen	Rendah	Sewa

Pengacakan data bertujuan untuk menguji konsistensi tingkat akurasi pada algoritma *naïve bayes*.

#### 3.4.4 Klasifikasi Data

Klasifikasi terhadap dataset menggunakan algoritma *naive bayes* untuk penyelesaian. Performa klasifikasi yang disajikan dengan melakukan validasi pada dataset dari hasil pengacakan yaitu berdasarkan pemerolehan dari *software rapidminer* dengan 25 sebagai data training dan 5 data testing .

**Tabel 3. 9 Data Training**

Kepala Rumah Tangga	Kondisi Rumah	Jumlah Penghasilan	Status pemilik rumah	Status Kelayakan
---------------------	---------------	--------------------	----------------------	------------------

Laki-laki	Batu permanen	Rendah	Milik Pribadi	Layak
Laki-laki	Batu permanen	Sedang	Milik Pribadi	Tidak Layak
Laki-laki	Batu permanen	Sedang	Sewa	Tidak Layak
Laki-laki	Gamping	Rendah	Sewa	Layak
Laki-laki	Gamping	Sedang	Sewa	Layak
Laki-laki	Batu permanen	Sedang	Milik Pribadi	Tidak Layak
Laki-laki	Gamping	Sedang	Milik Pribadi	Tidak Layak
Laki-laki	Batu permanen	Tinggi	Milik Pribadi	Tidak Layak
Laki-laki	Batu permanen	Tinggi	Milik Pribadi	Tidak Layak
Laki-laki	Batu permanen	Tinggi	Milik Pribadi	Tidak Layak
Laki-laki	Gamping	Tinggi	Milik Pribadi	Tidak Layak
Laki-laki	Gamping	Sedang	Milik Pribadi	Tidak Layak
Laki-laki	Gamping	Sedang	Milik Pribadi	Tidak Layak
Perempuan	Batu permanen	Sedang	Sewa	Layak
Laki-laki	Batu permanen	Rendah	Milik Pribadi	Tidak Layak
Laki-laki	Batu permanen	Rendah	Milik Pribadi	Tidak Layak
Laki-laki	Gamping	Sedang	Milik Pribadi	Layak
Perempuan	Gamping	Sedang	Sewa	Layak
Laki-laki	Batu permanen	Sedang	Milik Pribadi	Tidak Layak
Laki-laki	Batu permanen	Rendah	Milik Pribadi	Tidak Layak
Laki-laki	Gamping	Tinggi	Milik Pribadi	Tidak Layak
Laki-laki	Gamping	Tinggi	Milik Pribadi	Tidak Layak
Laki-laki	Gamping	Tinggi	Milik Pribadi	Tidak Layak
Laki-laki	Batu permanen	Rendah	Sewa	Layak
Laki-laki	Batu permanen	Rendah	Sewa	Tidak Layak

**Sumber:** Data Penelitian 2023

**Tabel 3. 10** Data Testing

Kepala Rumah Tangga	Kondisi Rumah	Jumlah Penghasilan	Status pemilik rumah	Status Kelayakan
Laki-laki	Gamping	Rendah	Sewa	Layak
Laki-laki	Batu permanen	Sedang	Milik Pribadi	Tidak Layak

Laki-laki	Gamping	Sedang	Sewa	Layak
Laki-laki	Gamping	Sedang	Milik Pribadi	Tidak Layak
Laki-laki	Gamping	Tinggi	Milik Pribadi	Tidak Layak

**Sumber:** Data Penelitian 2023

#### 3.4.4.1 Naive Bayes

Berikut penggunaan rumus naive bayes:

$$P(T|X) = P(X|T)(P(T)/P(X))$$

Rumus: Algoritma Naive Bayes

X=Data dengan class yang belum diketahui

T=Hipotesis data X pada class spesifik

P(T|X)=Probabilitas T berdasarkan kondisi

P(X)=Probabilitas X berdasarkan kondisi hipotesis T

P(X)=Probabilitas dari T

#### 3.4.4.2 Accuracy Algoritma Naive Bayes

Tahap ini dilakukan untuk melakukan evaluasi *performa* dengan melihat nilai akurasi dan metode yang akan digunakan dengan rumus akurasi sebagai berikut:

$$Akurasi = \frac{TP + TN}{TP + FN + FS + TN}$$

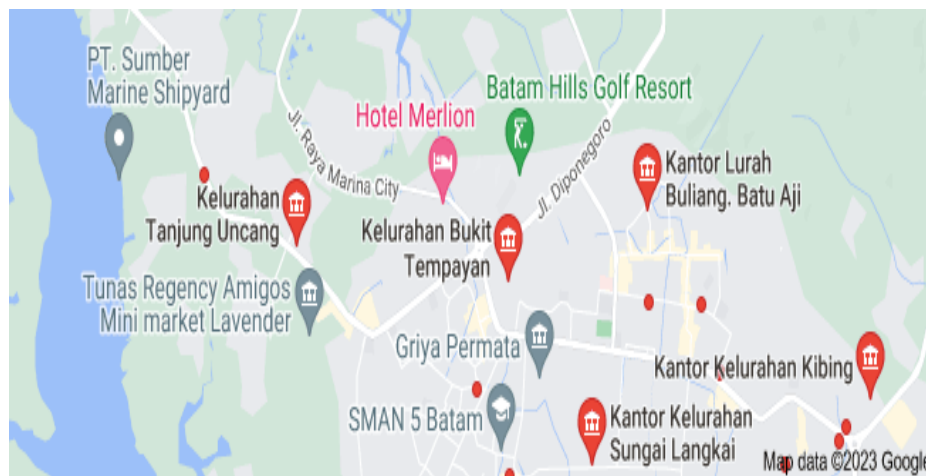
Keterangan :

- a. Akurasi adalah jumlah prediksi yang benar yang menyatakan puas.
- b. TP(*True Positive*) adalah jumlah responden yang diklasifikasikan sebagai positif oleh *classifier*.

### 3.5 Lokasi dan Jadwal Penelitian

#### 3.4.1 Lokasi Penelitian

Lokasi merupakan tempat dimana peneliti mendapatkan data penelitian. Berikut merupakan lokasi atau tempat data di peroleh yaitu dari kantor kelurahan Batu aji - Batam Kepulauan Riau.



**Gambar 3. 2** Lokasi Penelitian  
**Sumber:** Data Penelitian 2023

#### 3.4.2 Jadwal Penelitian

Merupakan jadwal yang diperlukan oleh peneliti dalam menyelesaikan penelitian sehingga tepat waktu:

**Tabel 3. 11** Jadwal Penelitian

No	Kegiatan	Tahun 2023																		
		Maret 2023			April 2023				Mai 2023				Jun 2023				Juli 2023			
		1	2	3	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	Pengajuan Surat penelitian	■	■	■																
2	Nyusun Bab I			■	■	■	■													
3	Nyusun Bab II					■	■	■	■											
4	Nyusun Bab III							■	■	■	■									
5	Nyusun Bab IV										■	■	■	■	■	■				
6	Nyusun Bab V, Daftar Pustaka, Lampiran															■	■	■	■	