

**PENERAPAN DATA MINING DENGAN *NAÏVE BAYES*
UNTUK KLASIFIKASI PENERIMA BANTUAN
SOSIAL (Studi Kasus: Kabupaten Sumenep)**

SKRIPSI



Oleh :

PUTU EKA FRISKA PERMATASARI
NIM. 190441100132

Pembimbing 1	:	Achmad Jauhari, S.T., M.Kom.	19810109 200604 1 003
Pembimbing 2	:	Doni Abdul Fatah, S.Kom., M.Kom	19870520 201903 1 013

**PROGRAM STUDI SISTEM INFORMASI
JURUSAN TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS TRUNOJOYO MADURA
2023**

“PENERAPAN DATA MINING DENGAN *NAÏVE BAYES* UNTUK KLASIFIKASI PENERIMA BANTUAN SOSIAL

(Studi Kasus: Kabupaten Sumenep)”

Putu Eka Friska Permatasari

(190441100132)

Pembimbing I : Achmad Jauhari, S.T., M.Kom.

Pembimbing II : Doni Abdul Fatah, S.Kom., M.Kom.

ABSTRAK

Indonesia dapat dikatakan sebagai negara yang memiliki tingkat kemiskinan yang tergolong tinggi. Permasalahan ini tidak pernah terselesaikan hingga saat ini. tidak hanya aspek ekonomi saja, tetapi aspek sosial adalah salah satu yang memiliki nilai penting dalam menilai kemiskinan. Kemiskinan juga terjadi salah satunya di Kabupaten Sumenep. Di Bappeda Sumenep keluarga yang telah didaftarkan ini tergolong sebagai keluarga yang kurang mampu. Data ini, memiliki beberapa parameter yang diduga masuk kedalam kategori sebagai keluarga yang tepat sasaran untuk menerima bantuan, namun data ini belum valid, karena tidak di proses dengan metode yang ada dan cukup menyita waktu yang lama. Dari masalah tersebut, dibutuhkan sistem yang dapat memproses data dengan efektif dan efisien untuk menghemat waktu dan meminimalisir terjadinya kesalahan yang ada. Pada penelitian ini, metode yang digunakan oleh peneliti adalah metode *naïve bayes* kemudian data yang digunakan diperoleh dari Badan Penelitian dan Pengembangan Daerah (Bappeda) Kabupaten Sumenep. Dalam pengujian, peneliti menggunakan pengujian *Confusion Matrix* yang digunakan untuk mencari nilai akurasi yang nantinya dihasilkan pada perhitungan *naïve bayes*. Pada skripsi ini, diharapkan dapat mengetahui hasil klasifikasi keluarga yang layak menerima bantuan sosial

dan untuk mengetahui tingkat akurasi penggunaan metode *naïve bayes* menggunakan *confusion matrix*.

Kata Kunci : Klasifikasi, Kemiskinan, *Naïve Bayes*, Data Mining, *Confusion Matrix*.

DAFTAR ISI

ABSTRAK	2
DAFTAR GAMBAR	6
BAB I PENDAHULUAN	8
1.1 Latar Belakang	8
1.1.1 Permasalahan.....	11
1.1.2 Usulan Solusi: Metode	11
1.2.3 Pertanyaan Penelitian	11
1.2 Tujuan dan Manfaat.....	11
1.2.1 Tujuan	11
1.2.2 Manfaat	12
1.3 Batasan Masalah	12
1.4 Metodologi Penulisan Skripsi	13
1.4.1 Objek Penelitian	13
1.4.2 Metode Pengumpulan Data	13
1.5 Metodologi Penulisan Skripsi.....	13
1.6 Jadwal Pelaksanaan Penelitian	14
BAB II TINJAUAN PUSTAKA.....	16
2.1 Kemiskinan	16
2.2 <i>Naïve Bayes</i>	16
2.3 Metode <i>Naïve</i>	18
2.4 Klasifikasi	18
2.5 <i>Confusion Matrix</i>	19
2.6 Data Mining	19
2.7 Penelitian Terkait	21
BAB III METODE USULAN.....	25
3.1 Identifikasi Masalah	25
3.2 Analisa Sistem.....	25

3.2.1 Analisa Kebutuhan	25
a. Kebutuhan Antarmuka	25
b. Kebutuhan Data	26
c. Kebutuhan Fungsional	26
d. Kebutuhan Non Fungsional	26
3.3 Desain Rancang Sistem	27
3.3.1 Flowchart	28
3.2.1.1 Flowchart <i>Naïve Bayes</i>	28
3.2.1.2 Flowchart Sistem Algoritma <i>Naïve Bayes</i>	29
3.3 Perhitungan Manual	30
3.3.1 Perhitungan manual menggunakan Algoritma <i>Naïve Bayes</i>	31
3.4 Pengujian Sistem	35
3.4.1 Pengujian Akurasi	35
REFERENSI	37

DAFTAR GAMBAR

Gambar 2. 1 Alur Algoritma Naive Bayes.....	17
Gambar 2. 2 Alur Informasi Data Mining.....	20
Gambar 3. 1 Tahapan Metode Usulan.....	25
Gambar 3. 2 Arsitektur Sistem.....	27
Gambar 3. 3 Flowchart Naive Bayes	28
Gambar 3. 4 Flowchart Sistem Naive Bayes	29
Gambar 3. 5 Data Sampel Perhitungan Manual.....	31
Gambar 3. 6 Data Latih Perhitungan Manual	31
Gambar 3. 7 Data Uji Perhitungan Manual.....	32

DAFTAR TABEL

Tabel 2.1 Confusion Matrix	19
Tabel 2. 2 Penelitian Terkait	22
Tabel 3. 1 Confusion Matrix Kelas Prediksi	36

BAB I

PENDAHULUAN

1.1 Latar Belakang

Indonesia dapat dikatakan sebagai negara yang memiliki tingkat kemiskinan yang tergolong tinggi. Permasalahan ini tidak pernah terselesaikan hingga saat ini. tidak hanya aspek ekonomi saja, tetapi aspek sosial adalah salah satu yang memiliki nilai penting dalam menilai kemiskinan. Di perkotaan, tuntutan mengenai kebutuhan hidup yang tidak dapat dikesampingkan membuat masyarakat harus berfikir dan bekerja keras demi memenuhi hidup sehari-hari. Dimana kadang pengeluaran bisa saja lebih besar daripada pemasukan yang ada.

Dipicu dengan kenaikan harga kebutuhan hidup dan pudarnya kebiasaan sosial lokal yang berfungsi untuk menekan tekanan kemiskinan, membuat masyarakat miskin yang berada di kota harus membayar fasilitas publik seperti kebutuhan air bersih dan lainnya [1]. Pemerintah membuat program yang dapat membantu masyarakat miskin di Indonesia dengan memberi bantuan sosial dengan tujuan untuk memberantas kemiskinan. Meskipun tidak dapat diberantas secara keseluruhan, tetapi pemerintah berusaha untuk memenuhi kebutuhan hidup sehari-hari. Berbagai bantuan yang disalurkan pemerintah kepada masyarakat miskin ini disayangkan tidak sesuai dengan yang diharapkan. Pada saat dilapangan seringkali bantuan diberikan tidak sesuai sasaran.

Di Bappeda Kabupaten Sumenep, ada beberapa parameter yang masuk kedalam kategori penerima bantuan kemiskinan untuk masyarakat di daerah tersebut. Namun karena cara yang dilakukan masih menggunakan cara manual dan terkadang rentannya terjadi kesalahan pada saat proses data yang ada, membuat masyarakat yang seharusnya layak untuk diberi bantuan menjadi tidak layak untuk diberi bantuan. Hal ini merugikan beberapa pihak, dimana program pemerintah yang dikhususkan untuk masyarakat yang tidak mampu menjadi salah sasaran dan masyarakat yang layak diberi bantuan tidak mendapatkan program tersebut yang akhirnya masyarakat ini menjadi kesusahan.

Klasifikasi sendiri memiliki arti sebagai gambaran dan perbedaan dari kelas data yang memiliki tujuan untuk dapat memprediksi kelas dari objek yang label kelasnya belum ditemukan [2]. Klasifikasi dilakukan agar data yang diolah dengan cepat dan tepat. Apabila data telah diklasifikasi biasanya dalam proses analisa menjadi lebih mudah. Ini dapat mempermudah pihak Bappeda untuk mengkategorikan masyarakat yang cocok untuk program pemerintah. Penelitian ini juga mengklasifikasikan beberapa parameter dengan menggunakan metode *naïve bayes*, dimana metode *naïve bayes* sendiri adalah algoritma yang digunakan untuk klasifikasi menggunakan teorema bayes yang berasumsi bahwa nilai antar variabel independensi (bebas).

Pada penelitian terkait tentang “Perbedaan Metode *Naïve Bayes* dan C4.5 Untuk Mengelompokkan Status Gizi Balita” menghasilkan nilai akurasi yang dimana pada metode *naïve bayes* sebesar 94,20% dan C4.5 sebesar 85,80%. Hasil akhir yang dihasilkan oleh penelitian tersebut adalah metode *naïve bayes* memiliki tingkat akurasi yang baik dibandingkan metode *decision tree*[3]

Pada penelitian mengenai “Penerapan Menggunakan *Naïve Bayes* Untuk Mengelompokkan Penerima Bantuan di Desa Pelangsian” memiliki tujuan untuk menganalisa bantuan yang akan disalurkan kepada masyarakat yang tidak mampu di desa tersebut. Pada penelitian ini, peneliti menggunakan metode *naïve bayes* dengan menggunakan *software* rapidminer dalam menghitung akurasi data. Akurasi yang dihasilkan adalah 83,67%, selain itu *recall* yang dihasilkan sebesar 74,29% dan *precision* 78,79% [4].

Pada penelitian mengenai “Implementasi Data Mining Dalam Mempekirakan Penerima Bantuan Pangan Non Tunai (BPNT) di Desa Wanacala Memakai *Naïve Bayes*” menggunakan sebelas kriteria dalam menyeleksi penerima bantuan BPNT ini. pengujian ini dibantu oleh RapidMiner dengan menghasilkan nilai akurasi sebesar 96%. Dalam penelitian ini, kelas yang dibutuhkan hanya ada dua yaitu kelas layak dan tidak layak. Penelitian ini menggunakan metode *naïve bayes* yang dimana metode tersebut merupakan metode klasifikasi dengan memanfaatkan teori probabilitas yang mempunyai tingkat akurasi tinggi [5]

Penelitian mengenai “SPK Menggunakan *Naïve Bayes* Dalam Penentuan PKH (Studi Kasus : Kelurahan Karanganyar Gunung Semarang)” menggunakan dua label yaitu label layak dan tidak layak, dimana penelitian ini menggunakan 5 variabel dalam menyeleksi masyarakat yang layak untuk dibantu. Pada penelitian ini, peneliti menggunakan framework PHP dalam pembuatan sistemnya. Penelitian ini menghasilkan nilai akurasi sebesar 93,33%, nilai presisi 87,50% dan *recall* sebesar 100% [6].

Penelitian mengenai “Implementasi Data Mining Menggunakan Algoritma C4.5 Untuk Penentuan Penerima Bansos (Studi Kasus: Kantor Kelurahan Martoba)” menggunakan pengujian Algoritma C4.5 kedalam RapidMiner dan memiliki hasil nilai validasi yang sepadan. Akurasi yang dihasilkan dari penelitian ini yaitu sebesar 100%. Dalam perhitungan menggunakan metode tersebut, terdapat faktor yang berpengaruh yaitu penghasilan dengan nilai gain sebesar 0,8474359 [7].

Berdasarkan penelitian tersebut, peneliti memilih metode *naïve bayes* yang dimana metode tersebut memiliki tingkat akurasi yang tinggi dalam pengklasifikasian. Pengujian ini dilakukan dengan menambahkan metode *confusion matrix*. Dimana *Confusion matrix* sendiri merupakan tabel yang mendeskripsikan performa dari sebuah model atau algoritma secara spesifik [8]

Berdasarkan penelitian mengenai “Penerapan *Naïve Bayes* Untuk Mengelompokkan Bantuan PKH di Desa Minggiran Kediri” yang dimana pada penelitian tersebut hanya menggunakan satu parameter saja [9]. Sedangkan dalam penelitian ini nantinya akan menentukan masyarakat yang layak diberi bantuan dengan menggunakan parameter sebanyak 16.

Berdasarkan uraian diatas, penulis mengambil sebuah judul yaitu “PENERAPAN DATA MINING DENGAN *NAÏVE BAYES* UNTUK KLASIFIKASI PENERIMA BANTUAN SOSIAL (Studi Kasus : Kabupaten Sumenep) dengan harapan dan tujuan agar penelitian ini dapat bermanfaat bagi Bappeda Kabupaten Sumenep dalam pengklasifikasian penerima bantuan di daerah tersebut.

1.1.1 Permasalahan

Pada dataset yang diperoleh dari Bappeda Kabupaten Sumenep ini tidak didukung dengan perhitungan yang valid. Data yang diambil juga memerlukan waktu yang lama dalam pemrosesan dan bisa saja rentan terjadinya kesalahan dalam pengklasifikasian untuk mengetahui keluarga yang layak untuk menerima bantuan sosial.

1.1.2 Usulan Solusi: Metode

Setelah melakukan pengumpulan data yang diperoleh dari Badan Perencanaan Pembangunan, Penelitian dan Pengembangan Daerah (Bappeda) Kabupaten Sumenep mengalami kepadatan keluarga yang mengalami kemiskinan, metode yang diusulkan adalah metode *naïve bayes*. Pada penelitian ini diusulkan solusi untuk membuat sistem yang bertujuan untuk mengklasifikasikan keluarga yang layak untuk menerima bantuan sosial yang ada di Kabupaten Sumenep.

1.2.3 Pertanyaan Penelitian

1. Bagaimana pengklasifikasian masyarakat yang layak untuk menerima bantuan sosial menggunakan algoritma *naïve bayes*
2. Berapa tingkat akurasi dengan *confusion matrix* yang didapat menggunakan algoritma *naïve bayes* ?

1.2 Tujuan dan Manfaat

1.2.1 Tujuan

Adapun tujuan dari penelitian ini adalah sebagai berikut :

1. Untuk mengetahui klasifikasi penerima bantuan sosial menggunakan algoritma *naïve bayes*.
2. Untuk mengetahui tingkat akurasi *confusion matrix* yang didapat menggunakan algoritma *naïve bayes*.

1.2.2 Manfaat

Adapun Manfaat yang didapat dari penelitian ini adalah :

1). Bagi Penulis

Sarana bagi penulis dalam mengimplementasikan ilmu yang didapat selama masa perkuliahan. Mendapat wawasan yang lebih luas dan pengalaman dalam penelitian berlangsung.

2). Bagi Universitas Trunojoyo Madura

Sebagai bahan referensi mahasiswa lainnya yang akan menyusun Skripsi mengenai penerapan data mining untuk klasifikasi penerima bantuan sosial menggunakan algoritma *naïve bayes*.

3). Bagi Pengguna

Membantu pengguna untuk mengetahui klasifikasi masyarakat yang layak diberi bantuan sosial di daerah Kabupaten Sumenep.

1.3 Batasan Masalah

Batasan masalah pada tugas akhir ini adalah:

- a. Metode yang digunakan adalah metode *naïve bayes*.
- b. Parameter yang digunakan adalah pendidikan, kepemilikan rumah, memiliki simpanan uang/perhiasan/ternak/lainnya, jenis atap, dinding, lantai, sumber penerangan, bahan bakar memasak, sumber air minum, memiliki fasilitas BAB, penerima BPNT, BPUM, BST, PKH, dan penerima sembako.
- c. Data yang digunakan diperoleh dari Bappeda Kabupaten Sumenep tahun 2022 dengan jumlah data 155.223.
- d. Sistem yang dirancang adalah sistem klasifikasi untuk mengetahui penerima bantuan sosial.
- e. Target pada sistem yang akan dibangun adalah sangat layak, layak dan tidak layak.
- f. Sistem yang dibangun menggunakan bahasa Python.

1.4 Metodologi Penulisan Skripsi

1.4.1 Objek Penelitian

Objek penelitian yang digunakan dalam penulisan laporan proposal ini adalah data yang keluarga individu Desil 1-3 di Kabupaten Sumenep yang terdaftar pada Bappeda Kabupaten Sumenep.

1.4.2 Metode Pengumpulan Data

1. Observasi

Melakukan penelitian langsung ke instansi dan pihak terkait dalam pengumpulan data dan beberapa informasi yang dibutuhkan untuk penelitian dan sesuai dengan kasus permasalahan yang terjadi.

2. Studi Pustaka

Melakukan pencarian dan mempelajari buku-buku yang sesuai dengan kasus permasalahan hingga fakta yang sudah ada dari internet.

3. Wawancara

Melakukan tanya jawab dengan pihak terkait yang ada hubungannya dengan permasalahan yang terjadi sesuai kebutuhan dalam penelitian ini.

1.5 Metodologi Penulisan Skripsi

BAB I PENDAHULUAN

Pada bab ini berisikan latar belakang masalah, rumusan masalah, tujuan dan manfaat, batasan masalah, sistematika penulisan dan jadwal pelaksanaan yang akan dilakukan oleh peneliti.

BAB II TINJAUAN PUSTAKA

Pada bab ini berisikan tinjauan pustaka yang diambil oleh peneliti yang relevan dengan penelitian serta kajian teori dan sistem yang akan digunakan.

BAB III METODE USULAN

Pada Bab ini berisi mengenai beberapa tahapan peneliti dalam mengumpulkan data, pemecahan masalah dan perancangan sistem yang akan digunakan.

REFERENSI

Pada daftar pustaka berisi referensi yang digunakan oleh peneliti untuk menunjang penelitian yang dilakukan.

1.6 Jadwal Pelaksanaan Penelitian

Berikut adalah estimasi jadwal pelaksanaan penelitian yang akan dilakukan. Estimasi jadwal penelitian dapat dilihat pada table 1.1.

Tabel 1. 1 Jadwal Pelaksanaan Penelitian

NO	KEGIATAN	BULAN 1				BULAN 2				BULAN 3				BULAN 4				BULAN 5				BULAN 6			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	Mempelajari Metode Naïve Bayes																								
2	Mencari Referensi jurnal																								
3	Pengumpulan Data																								
4	Uji Coba Data																								
5	Penyusunan Proposal																								
6	Membangun Program																								
7	Uji Coba Sistem																								
8	Penyusunan Laporan																								
9	Sidang Akhir																								

BAB II

TINJAUAN PUSTAKA

Pada sub bab ini akan dideskripsikan mengenai teori yang digunakan dalam penelitian. Landasan yang dijabarkan dapat tentang istilah pada topik yang akan diteliti. Teori yang diperoleh dalam penelitian ini bisa dari kutipan para penulis buku ataupun penelitian terdahulu.

2.1 Kemiskinan

Kemiskinan menjadi salah satu permasalahan yang ada di Indonesia. Badan Pusat Statistik (BPS) mengkategorikan masyarakat miskin menjadi dua yang dapat dilihat dari penjumlahan garis makanan (GKM) dan kemiskinan non-makanan (GKNM). GKM ini mengarahkan pada kebutuhan yang dikeluarkan dalam perihal makanan dengan takaran 2.100 kkal untuk setiap makanan. Sedangkan untuk GKNM ditentukan berdasarkan kebutuhan antara lain seperti perumahan, sandang, pendidikan, kesehatan dan lain sebagainya [10].

Masalah kemiskinan ini ditimbulkan dari beberapa faktor yang ada seperti besarnya tingkat pengangguran yang dari tahun ke tahun semakin bertambah, dimana lapangan pekerjaan yang sulit untuk didapatkan. Selain itu juga upah pekerja yang tidak selaras dengan kebutuhan hidup dan kualitas hidup masyarakat yang bisa terbilang rendah [11].

2.2 *Naïve Bayes*

Klasifikasi *naïve bayes* adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas dalam sebuah *class*. Berdasarkan teorema Bayes yang terdapat klasifikasi yang sama dengan *decision tree* dan *neural network*. Klasifikasi bayesian terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar. Terdapat rumus dalam perhitungan *naïve bayes*, bentuk umum teorema bayes adalah sebagai berikut [2]

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad \dots (2.1)$$

Keterangan :

X = Data dengan *class* yang belum diketahui

H = Hipotesis dari data X didefinisikan sebagai *class* yang spesifik

$P(H|X)$ = Probabilitas suatu hipotesis H berdasarkan kondisi dari X

$P(H)$ = Probabilitas hipotesis dari H

$P(X|H)$ = Probabilitas X berdasarkan pada kondisi tersebut

$P(X)$ = Probabilitas X [12]

Untuk cara kerja dalam algoritma *naïve bayes* ini akan dijabarkan dengan flowchart di bawah ini :



Gambar 2. 1 Alur Algoritma *Naive Bayes*

1. Langkah awal adalah dengan membaca data training
2. Selanjutnya menghitung jumlah probabilitas pada kelas yang ada

3. Kemudian mendapatkan tabel probabilitas dan menghasilkan sebuah solusi [13]

Adapun kelebihan dan kekurangan dalam algoritma naïve bayes antara lain sebagai berikut :

a. Kelebihan

- Dapat digunakan untuk data kuantitatif maupun kualitatif
- Tidak memerlukan jumlah data yang banyak
- Tidak melakukan data training yang banyak
- Apabila ada nilai yang hilang, maka bisa diabaikan dalam perhitungan.
- Perhitungannya bisa terbilang cepat dan efisien
- Mudah untuk dipahami

b. Kekurangan

- Jika probabilitas kondisionalnya bernilai nol, maka probabilitas prediksi juga akan bernilai nol
- Diasumsikan bahwa masing-masing variabel independent dapat membuat berkurangnya akurasi, karena biasanya ada korelasi antara variabel yang satu dengan variabel yang lain
- Keakuratannya tidak dapat diukur menggunakan satu probabilitas saja. Dibutuhkan bukti-bukti lain untuk dapat membuktikannya.

2.3 Metode Naïve

2.4 Klasifikasi

Klasifikasi merupakan sebuah model yang mendeskripsikan dan membedakan data yang memiliki kelas yang bertujuan untuk dapat memprediksi kelas dari objek yang kelasnya belum diketahui [12].

Sistem klasifikasi dapat berupa jenis, ukuran, warna, abjad judul dan juga abjad pengarang. Tetapi, biasanya perpustakaan menggunakan sistem kategori dalam subjek pustaka [14]. Definisi klasifikasi bisa disimpulkan sebagai pengelompokan dalam

beberapa objek, gagasan, buku atau benda yang dilandasi oleh subjek atau karakteristik yang sama dalam penataannya yang teratur yang bisa diselaraskan dengan subjek yang saling berdekatan posisinya.

2.5 Confusion Matrix

Confusion matrix merupakan sebuah tabel yang isinya terdapat klasifikasi data yang sedang diuji dengan benar dan jumlah data uji yang salah. Contoh dari *confusion matrix* dalam klasifikasi biner dapat dilihat pada tabel dibawah ini :

Tabel 2. 1 *Confusion Matrix*

Tabel 2.1 *Confusion Matrix*

		Kelas Prediksi	
		1	0
Kelas Sebenarnya	1	TP	FN
	0	FP	TN

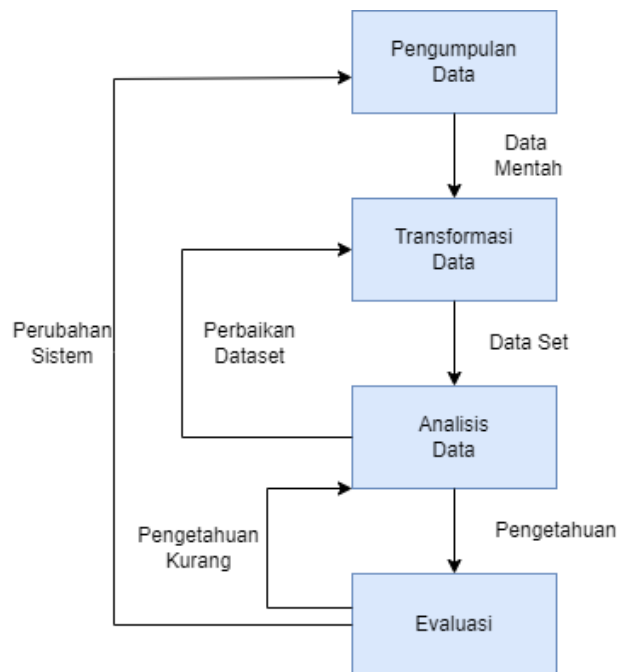
Keterangan :

1. TP (*True Positive*) = Dari jumlah berkas dikategorikan kelas 1 yang benar maka diklasifikasikan sebagai kelas 1
2. TN (*True Negative*) = Dari jumlah berkas dikategorikan kelas 0 yang benar maka diklasifikasikan sebagai kelas 0
3. FP (*False Positive*) = Dari jumlah berkas dikategorikan kelas 0 yang salah maka diklasifikasikan sebagai kelas 1
4. FN (*False Negative*) = Dari jumlah berkas dikategorikan kelas 1 yang salah maka dikalsifikasikan sebagai kelas 0 [15]

2.6 Data Mining

Definisi dari data mining sendiri yaitu pembagian pada model dari penjelasan untuk menyimpan sebuah *database* [16]. Data mining sering disebut sebagai *Knowledge Discovery in Database* (KDD). KDD merupakan proses yang terdiri dari

pembersihan data (*data cleaning*), integrasi data (*data intergration*), pemilihan data (*data selection*), transformasi data, data mining, evaluasi pola, dan penyajian pengetahuan. Proses dari data mining sendiri diawali dengan *preprocessing*, dimana proses tersebut adalah mengumpulkan data hingga menghasilkan data mentah yang dibutuhkan oleh data mining. Selanjutnya adalah transformasi data yang berfungsi untuk mengubah data mentah menjadi sebuah format yang akan diproses oleh data mining. Hasil dari transformasi data ini digunakan sebagai analisa data yang akan menimbulkan sebuah pengetahuan dengan teknik seperti analisis statistik, *machine learning*, dan visualisasi informasi [17].



Gambar 2. 2 Alur Informasi Data Mining

Klasifikasi data mining memiliki kemiripan dalam konsep klasifikasi pada bidang biologi. Klasifikasi memiliki tujuan yaitu melakukan sebuah prosedur pengelompokan. Dilihat dari perbedaannya, di bidang biologi klasifikasi memiliki tujuan untuk mengelompokkan makhluk hidup sedangkan di bidang data mining prosedur klasifikasi memiliki tujuan untuk mengelompokkan data menjadi beberapa kelompok. Dalam prosesnya, pengelompokan ini menggunakan data yang sudah

diketahui kelasnya. Data yang belum memiliki kelas, mampu menentukan kelompoknya melalui perbandingan data yang telah diketahui kelompoknya. Adapun algoritma klasifikasi yang banyak diketahui adalah *decision tree*, *naïve bayes*, *K-Nearest Neighbor* (KNN), jaringan syaraf tiruan, algoritma genetik, *Support Vector Machine* (SVM).

Beberapa manfaat yang didapati dalam penggunaan data mining salah satunya adalah mengenai ancaman keamanan dalam pengintaian penggunaan komputer sehari-hari. Contoh ancaman yang sering dijumpai seperti *malware*, *phising*, *spam*, akses buatan yang ada di dalam komputer dapat diminimalisir menggunakan data mining. Definisi *malware* sendiri yaitu perangkat lunak yang dirancang untuk mengerjakan sebuah aktifitas yang tidak diijinkan dalam komputer, sebagai contoh pencurian data. Dalam penggunaan berbagai metode klasifikasi, pola-pola *malware* dapat diidentifikasi agar penyebarannya dapat diatasi [18].

2.7 Penelitian Terkait

Dalam sebuah penelitian diperlukan adanya penelitian sebelumnya yang relevan dengan penelitian yang akan dilakukan. Dengan tujuan agar penelitian sebelumnya dapat membantu proses selama penelitian dilakukan. Penelitian terkait tentang “Implementasi Data Mining Menggunakan Algoritma C4.5 Untuk Penentuan Penerima Bansos (Studi Kasus : Kantor Kelurahan Martoba)” menghasilkan validasi yang sama dan tingkat akurasi sebesar 100%. Perhitungan yang dilakukan dengan metode tersebut juga mendapatkan hasil nilai gain yaitu 0,8474359 [7].

Penelitian terkait tentang “Implementasi Data Mining Dengan *Naïve Bayes* Dalam Pengklasifikasian Penerima Bantuan PKH di Desa Wae Jare” menghasilkan tingkat akurasi dengan menggunakan *confusion matrix* sebesar 82,14% dimana jumlah data training yang digunakan sebanyak 154 dan testing sebanyak 56 data [19]. Penelitian terkait tentang “Implementasi Dengan *Naïve Bayes* Untuk Pengklasifikasian Dalam Penerima Bantuan di Desa Pelangian” menghasilkan tingkat akurasi sebesar 83,67%, presisi sebesar 78,79% dan recall sebesar 74,29% [4].

Penelitian terkait tentang “Implementasi Data Mining Dengan Metode *Naïve Bayes* Dalam Memprediksi Penerimaan Bantuan Pangan Non Tunai (BPNT) di Desa Wanacala” menghasilkan akurasi sebesar 96% dan *error* sebesar 4%. Pada penelitian ini menggunakan perhitungan manual dan juga dengan bantuan *tools RapidMiner* dengan hasil dugaan yang sama [5]. Penelitian terkait tentang “Sistem Pendukung Keputusan Menggunakan *Naïve Bayes* Dalam Penentuan PKH (Studi Kasus : Kelurahan Karanganyar Gunung Semarang)”. Penelitian ini menggunakan *RapidMiner* dan memperoleh akurasi sebesar 93,33%, nilai presisi sebesar 87,50% dan recall sebesar 100% [6].

Penelitian terkait tentang “Penerapan *Naïve Bayes* Dalam Pengklasifikasian Penerima PKH”. Penelitian ini menggunakan model *confusion matrix* dan menghasilkan tingkat akurasi sebesar 88%, nilai presisi sebesar 91%, recall sebesar 88% dan f1-score sebesar 87%. Dimana nilai yang dihasilkan oleh penelitian ini termasuk kedalam kategori sebagai *Good Classification* [20].

Tabel 2. 2 Penelitian Terkait

Pengarang, Tahun	Data	Metode	Pengujian	Hasil
Winda Lidysari, Heru Staria Tambunan, Hendry Qurniawan, 2022	Kantor Kelurahan Martoba	<i>Naïve Bayes</i>	Algoritma C4.5	Penelitian ini menguji menggunakan Algoritma C4.5 menghasilkan tingkat akurasi sebesar 100%
Apolinaria Ifon Purnama, Abdul Aziz,	Dataset dari Sekretaris Desa	<i>Naïve Bayes</i>	<i>Confusion Matrix</i>	Pada penelitian ini menghasilkan

Anggri Sartika Wiguna, 2020				tingkat akurasi sebesar 82,14%. Data yang digunakan sebanyak 154 data training dan 56 data testing.
Warisa, Nurahman, 2022	Data BLT dari Desa Pelangsian	<i>Naïve Bayes</i>		Pada penelitian ini, tingkat akurasi yang dihasilkan sebesar 83,67%, nilai presisi 78,79%, dan recall sebesar 74,29%
Edy Widodo dan Ahmad Jaelani, 2022	Data warga sasaran BPNT hasil sensus di Desa Wanacala Kecamatan Harjamukti Kota Cirebon	<i>Naïve Bayes</i>		Akurasi yang dihasilkan dari penelitian ini adalah sebesar 96%, kelas yang digunakan sebanyak 2 kelas yaitu

				“layak” dan “tidak layak”
Amin Abdullah Sidiq dan Febrian Wahyu Christanto, 2020	Staf IT Kelurahan Karanganyar Gunung setelah mendapat persetujuan dari Lurah Karanganyar Gunung.	<i>Naïve Bayes</i>	RapidMiner	Penelitian ini menghasilkan tingkat akurasi sebesar 93,33%, nilai presisi sebesar 87,50% dan recall sebesar 100%
Andhini Asri Awaliyah Arifin, Wiwin Handoko dan Zulfan Efendi, 2022	Dinas Sosial Kabupaten Asahan	<i>Naïve Bayes</i>	<i>Confusion Matrix</i>	Menghasilkan tingkat akurasi sebesar 88%, nilai presisi sebesar 91%, recall sebesar 88% dan f1- score sebesar 87%

BAB III

METODE USULAN

Pada tahap metode usulan yang akan dilakukan oleh peneliti, peneliti melakukan beberapa tahapan beserta penjelasan diantaranya adalah sebagai berikut :



Gambar 3. 1 Tahapan Metode Usulan

3.1 Identifikasi Masalah

Identifikasi masalah adalah langkah awal dalam penelitian. Peneliti mencari permasalahan yang bertujuan menemukan ide untuk melakukan sebuah penelitian. Dalam pencarian tersebut, peneliti melakukan observasi mengenai subjek dan objek dari masalah yang terjadi pada tempat tertentu.

3.2 Analisa Sistem

Analisa sistem merupakan pembagian dan penyelidikan dalam sebuah benda untuk mencari komponen yang penting dalam pembangunan sistem informasi [21]. Dalam analisa sistem ini, peneliti dapat mengetahui baik atau tidaknya sistem berjalan menggunakan metode *naïve bayes*. Ini merupakan tahap awal dalam merancang sistem yang penting untuk dilakukan peneliti.

3.2.1 Analisa Kebutuhan

Pada rancangan sistem yang akan dibangun membutuhkan beberapa hal agar kebutuhan sistem yang dibangun dapat berjalan dengan baik. Kebutuhan tersebut antara lain :

a. Kebutuhan Antarmuka

Peneliti merancang dalam implementasi Algoritma *Naïve Bayes* dengan membuat desain yang cukup mudah untuk digunakan. Ada beberapa menu yang akan dibuat dalam sistem ini, yaitu :

- Halaman Awal (Beranda)

- Data
- Perhitungan *Naïve Bayes*
- Hasil klasifikasi *Naïve Bayes*
- Logout (Keluar)

b. Kebutuhan Data

Untuk mengetahui apakah sistem yang dirancang dapat menghasilkan kinerja yang baik, maka dibutuhkan data pendukung dalam menentukan klasifikasi penerima bantuan. Data yang diproses adalah data dengan bentuk excel.

c. Kebutuhan Fungsional

Dalam kebutuhan fungsional, peneliti harus menyediakan beberapa hal, antara lain :

1. Membutuhkan sebuah sistem yang dapat mengklasifikasikan data dengan cepat dan efisien.
2. Membutuhkan sebuah sistem yang dapat menyimpan data.
3. Membutuhkan sebuah sistem yang dapat mengolah data dengan baik.

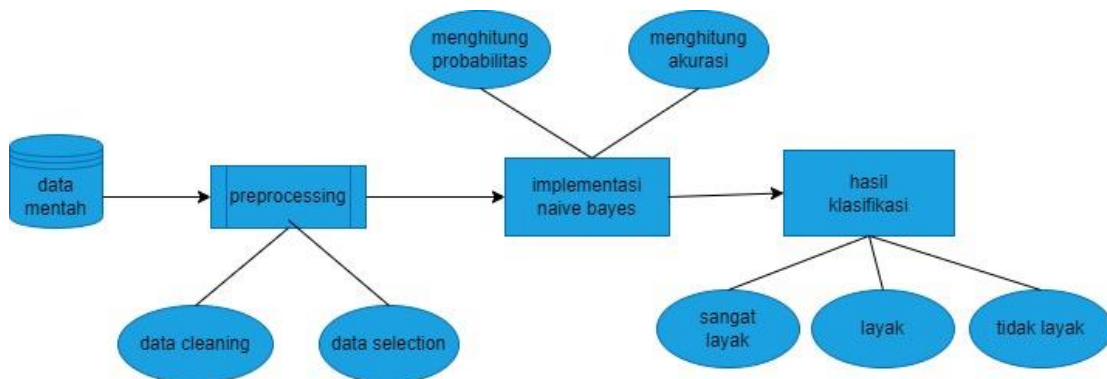
d. Kebutuhan Non Fungsional

Pada kebutuhan non fungsional ini, peneliti menyediakan komponen dalam merancang sebuah sistem, komponen tersebut adalah :

1. Processor Intel Dual Core N3060
2. Memory 2GB
3. Hardisk 500GB
4. Windows 10
5. Google Chrome

3.2.2 Analisa Arsitektur Sistem

Dalam tahap arsitektur sistem ini memiliki tujuan untuk mengidentifikasi sistem yang nantinya akan dibangun. Gambaran arsitektur yang akan dibangun, dapat dilihat sebagai berikut :



Gambar 3. 2 Arsitektur Sistem

Penjelasan mengenai arsitektur sistem adalah sebagai berikut :

Pada proses awal dimana data mentah yang diperoleh akan melakukan *preprocessing*. *Preprocessing* yang dilakukan berupa *data cleaning* dan *data selection*. *Data cleaning* bertujuan untuk memastikan tidak adanya data yang kosong, begitu juga untuk *data selection* yang bertujuan untuk memilih data apa saja yang digunakan dalam penelitian. Selanjutnya adalah mengimplementasi pada metode, dimana metode yang digunakan adalah metode *naïve bayes*. Pada pengimplementasian ini, peneliti menghitung nilai probabilitas dan akurasi yang dihasilkan. Pada akhir nanti, hasil klasifikasinya terdapat 3 kategori yaitu sangat layak, layak dan tidak layak.

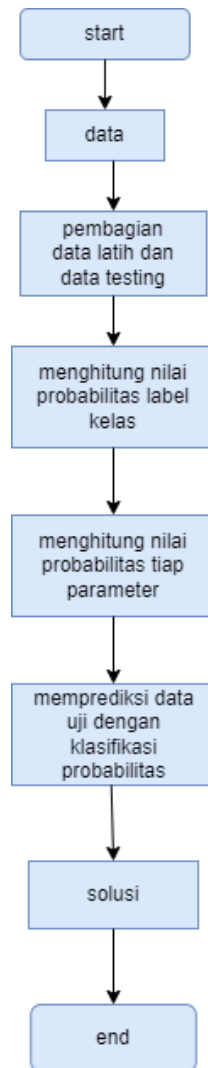
3.3 Desain Rancang Sistem

Pada tahap ini, peneliti merancang desain sesuai dengan kebutuhan pengguna dan juga dapat menentukan kinerja sistem berjalan dengan baik atau tidak. Ada beberapa bahasan yang akan dibahas pada sub bab ini seperti flowchart, desain database, rancangan basis data dan antarmuka sistem.

3.3.1 Flowchart

Flowchart bisa dikatakan sebagai alur proses dalam sebuah sistem. Flowchart menggambarkan secara detail dan dapat menerangkan hubungan yang telah terjadi dengan bentuk bagan.

3.2.1.1 Flowchart *Naïve Bayes*

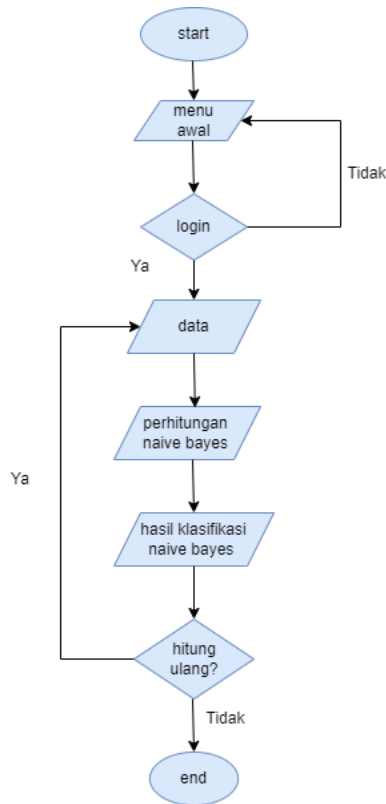


Gambar 3. 3 Flowchart *Naive Bayes*

Penjelasan :

Langkah awal adalah menyiapkan data, dimana data yang sudah ada akan dibagi menjadi 2 bagian yaitu data training dan data testing. Setelah itu, menghitung nilai probabilitas *naïve bayes* dengan menggunakan rumus *naïve bayes*. Kemudian menghitung probabilitas kelas lalu memprediksi data yang diujikan dengan klasifikasi pada probabilitas. Langkah terakhir melakukan analisa terhadap hasil klasifikasi.

3.2.1.2 Flowchart Sistem Algoritma *Naïve Bayes*



Gambar 3. 4 Flowchart Sistem *Naive Bayes*

Penjelasan :

Dapat dilihat pada gambar 3.2 adalah flowchart sistem *naïve bayes* yang akan di rancang. Dimana dibagian awal akan ditampilkan halaman login, pengguna akan diminta untuk memasukkan *username* dan password. Jika berhasil masuk, pengguna akan diperlihatkan dengan beranda.

3.3 Perhitungan Manual

Ada beberapa tahapan yang dilakukan oleh peneliti sebelum memasuki perhitungan manual dengan menggunakan metode *naïve bayes* diantaranya adalah sebagai berikut:

1. Analisis masalah

Tahap ini adalah tahap awal yang dilakukan peneliti untuk memastikan permasalahan yang ada. Setelah memastikan permasalahan tersebut, peneliti menganalisa cara untuk menyelesaikan permasalahan tersebut dan menentukan ruang lingkup masalah yang diteliti. Langkah selanjutnya yaitu dengan mempelajari teori-teori yang berhubungan dengan masalah yang akan diteliti.

2. Mengumpulkan data

Data yang diperoleh ini berupa file *excel* berasal dari Bappeda (Badan Perencanaan Pembangunan Daerah) Kabupaten Sumenep

3. Implementasi dan pengujian

Adapun metode yang digunakan dalam proses data mining, antara lain:

- a. *Data Cleaning*

Pada tahap awal proses data mining ini adalah dengan memastikan bahwa data yang telah diperoleh tidak ada nilai kosong.

- b. *Data Selection*

Dalam data *selection* ini adalah dengan memilih variabel yang akan digunakan dalam penelitian. Data yang akan digunakan adalah pendidikan, kepemilikan rumah, mempunyai simpanan uang/ternak/perhiasan/lainnya, jenis atap, jenis dinding, jenis lantai, sumber penerangan, sumber air minum, bahan bakar memasak, fasilitas BAB, penerima BPUM, BPNT, PKH, BST, sembako dan status. Dalam variabel status, digunakan sebagai label kelas dalam perhitungan manual. Data yang tidak tercantum dalam variabel penelitian yang dipakai akan diabaikan.

- c. Implementasi *Naïve Bayes*

Dalam perhitungan manual ini, peneliti memiliki tujuan untuk lebih mengenal algoritma *naïve bayes* sebelum diimplementasikan ke *coding*. Dalam tahap ini data yang digunakan sebanyak 20 data yang diambil. Berikut adalah data sampel yang akan digunakan untuk perhitungan manual, sebagai berikut :

Nama	Pendidikan	Keperiln Rumah	Mem	Jenis Atap	Jenis Dinding	Jenis Lantai	Sumer Penerangan	Bahan Baka	Sumber Air Minum	Fasilitas BAB	Pener	Pener	Pener	Pener	Pener	Pener	status
ACH. M	Tamat SMP	Milik Sendiri	Tidak	Genteng	Tembok	Keramik/Gr	Listrik Pribadi > 900	Listrik/Gas	Sumur Bor	Ya, dengan Se	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	Layak
ABDUS	Tamat SMA	Milik Sendiri	Tidak	Genteng	Tembok	Semen	Listrik Pribadi > 900	Listrik/Gas	Ledeng/PAM	Ya, dengan Se	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	Sangat Layak
MARSU	Tamat SD/s4	Milik Sendiri	Ya	Genteng	Tembok	Keramik/Gr	Listrik Pribadi s/d 9	Listrik/Gas	Sumur Terlindung	Ya, dengan Se	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	Sangat Layak
DWI F	Tamat Pergu	Milik Sendiri	Tidak	Beton	Tembok	Keramik/Gr	Listrik Pribadi > 900	Listrik/Gas	Sumur Bor	Ya, dengan Se	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	Layak
RUDIY	Tamat Pergu	Milik Sendiri	Tidak	Genteng	Tembok	Keramik/Gr	Listrik Pribadi s/d 9	Listrik/Gas	Ledeng/PAM	Ya, dengan Se	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	Layak
SAHRIN	Tamat SD/s4	Milik Sendiri	Ya	Genteng	Tembok	Semen	Listrik Pribadi s/d 9	Listrik/Gas	Sumur Terlindung	Ya, dengan Se	Tidak	Tidak	Ya	Tidak	Tidak	Tidak	Sangat Layak
M. RAN	Tamat SMA	Milik Sendiri	Ya	Genteng	Tembok	Keramik/Gr	Listrik Pribadi > 900	Listrik/Gas	Sumur Bor	Ya, dengan Se	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	Layak
HAKKO	Siswa SD/se	Milik Sendiri	Tidak	Genteng	Tembok	Keramik/Gr	Listrik Pribadi > 900	Listrik/Gas	Air Kemasan/Isi Ula	Ya, dengan Se	Tidak	Tidak	Ya	Tidak	Tidak	Tidak	Layak
HAIRY	Tamat SD/s4	Bebas Sewa	Ya	Genteng	Tembok	Keramik/Gr	Listrik Pribadi s/d 9	Listrik/Gas	Sumur Terlindung	Ya, dengan Se	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	Sangat Layak
MOHA	Tamat SMA	Milik Sendiri	Tidak	Genteng	Tembok	Keramik/Gr	Listrik Pribadi > 900	Listrik/Gas	Ledeng/PAM	Ya, dengan Se	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	Layak
ALIYAH	Tidak tamat	Milik Sendiri	Tidak	Genteng	Tembok	Semen	Listrik Pribadi s/d 9	Listrik/Gas	Sumur Bor	Ya, dengan Se	Ya	Tidak	Ya	Tidak	Ya	Tidak	Sangat Layak
ABDUR	Tamat SMA	Milik Sendiri	Ya	Genteng	Tembok	Keramik/Gr	Listrik Pribadi s/d 9	Listrik/Gas	Sumur Terlindung	Ya, dengan Se	Tidak	Ya	Tidak	Tidak	Tidak	Tidak	Sangat Layak
HOIFU	Tamat SMA	Menumpang	Ya	Genteng	Tembok	Keramik/Gr	Listrik Pribadi s/d 9	Listrik/Gas	Ledeng/PAM	Ya, dengan Se	Tidak	Ya	Tidak	Tidak	Tidak	Tidak	Layak
PATRA	Tidak tamat	Kontrak/Sewa	Ya	Genteng	Bambu	Semen	Listrik Pribadi s/d 9	Arang/Kayu	Sumur Terlindung	Tidak, Jamba	Ya	Tidak	Ya	Ya	Ya	Ya	Sangat Layak
RB. AB	Tamat SMA	Bebas Sewa	Ya	Genteng	Tembok	Keramik/Gr	Listrik Pribadi s/d 9	Listrik/Gas	Ledeng/PAM	Ya, dengan Se	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	Sangat Layak
TOHI S	Tamat SMP	Milik Sendiri	Ya	Genteng	Tembok	Keramik/Gr	Listrik Pribadi > 900	Listrik/Gas	Sumur Bor	Ya, dengan Se	Ya	Tidak	Ya	Ya	Ya	Ya	Layak
MOHA	Tamat SMA	Milik Sendiri	Tidak	Genteng	Tembok	Keramik/Gr	Listrik Pribadi s/d 9	Listrik/Gas	Sumur Bor	Ya, dengan Se	Ya	Tidak	Ya	Tidak	Ya	Tidak	Layak
WARN	Tamat SD/s4	Milik Sendiri	Ya	Genteng	Tembok	Keramik/Gr	Listrik Pribadi s/d 9	Listrik/Gas	Sumur Terlindung	Ya, dengan Se	Ya	Tidak	Ya	Tidak	Ya	Tidak	Sangat Layak
DARM	Tamat SMA	Milik Sendiri	Tidak	Genteng	Tembok	Keramik/Gr	Listrik Pribadi > 900	Listrik/Gas	Ledeng/PAM	Ya, dengan Se	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	Layak
ASIYA	Tamat SD/s4	Milik Sendiri	Tidak	Genteng	Tembok	Keramik/Gr	Listrik Pribadi > 900	Listrik/Gas	Ledeng/PAM	Ya, dengan Se	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	Sangat Layak

Gambar 3. 5 Data Sampel Perhitungan Manual

3.3.1 Perhitungan manual menggunakan Algoritma *Naïve Bayes*

1. Menyiapkan data sampel dan membagi data menjadi data training dan data testing (uji)

Data sampel yang digunakan adalah data pada dataset yang ke 794-813, data tersebut dibagi menjadi dua, yaitu data latih dan data uji. Untuk data latih yang akan digunakan adalah sebagai berikut :

DATA LATIH																
Nama	Pendidikan	Keperiln Rumah	Mem	Jenis Atap	Jenis Dinding	Jenis Lantai	Sumer Penerangan	Bahan Baka	Sumber Air Minum	Fasilitas BAB	Pener	Pener	Pener	Pener	Pener	status
ACH. M	Tamat SMP	Milik Sendiri	Tidak	Genteng	Tembok	Keramik/Gr	Listrik Pribadi > 900	Listrik/Gas	Sumur Bor	Ya, dengan Se	Tidak	Tidak	Tidak	Tidak	Tidak	Layak
ABDUS	Tamat SMA	Milik Sendiri	Tidak	Genteng	Tembok	Semen	Listrik Pribadi > 900	Listrik/Gas	Ledeng/PAM	Ya, dengan Se	Tidak	Tidak	Tidak	Tidak	Tidak	Sangat Layak
MARSU	Tamat SD/s4	Milik Sendiri	Ya	Genteng	Tembok	Keramik/Gr	Listrik Pribadi s/d 9	Listrik/Gas	Sumur Terlindung	Ya, dengan Se	Tidak	Tidak	Tidak	Tidak	Tidak	Sangat Layak
DWI F	Tamat Pergu	Milik Sendiri	Tidak	Beton	Tembok	Keramik/Gr	Listrik Pribadi > 900	Listrik/Gas	Sumur Bor	Ya, dengan Se	Tidak	Tidak	Tidak	Tidak	Tidak	Layak
RUDIY	Tamat Pergu	Milik Sendiri	Tidak	Genteng	Tembok	Keramik/Gr	Listrik Pribadi s/d 9	Listrik/Gas	Ledeng/PAM	Ya, dengan Se	Tidak	Tidak	Tidak	Tidak	Tidak	Layak
SAHRIN	Tamat SD/s4	Milik Sendiri	Ya	Genteng	Tembok	Semen	Listrik Pribadi s/d 9	Listrik/Gas	Sumur Terlindung	Ya, dengan Se	Tidak	Tidak	Ya	Tidak	Tidak	Sangat Layak
M. RAN	Tamat SMA	Milik Sendiri	Ya	Genteng	Tembok	Keramik/Gr	Listrik Pribadi > 900	Listrik/Gas	Sumur Bor	Ya, dengan Se	Tidak	Tidak	Tidak	Tidak	Tidak	Layak
HAKKO	Siswa SD/se	Milik Sendiri	Tidak	Genteng	Tembok	Keramik/Gr	Listrik Pribadi > 900	Listrik/Gas	Air Kemasan/Isi Ula	Ya, dengan Se	Tidak	Tidak	Ya	Tidak	Tidak	Layak
HAIRY	Tamat SD/s4	Bebas Sewa	Ya	Genteng	Tembok	Keramik/Gr	Listrik Pribadi s/d 9	Listrik/Gas	Sumur Terlindung	Ya, dengan Se	Tidak	Tidak	Tidak	Tidak	Tidak	Sangat Layak
MOHA	Tamat SMA	Milik Sendiri	Tidak	Genteng	Tembok	Keramik/Gr	Listrik Pribadi > 900	Listrik/Gas	Ledeng/PAM	Ya, dengan Se	Tidak	Tidak	Tidak	Tidak	Tidak	Layak
ALIYAH	Tidak tamat	Milik Sendiri	Tidak	Genteng	Tembok	Semen	Listrik Pribadi s/d 9	Listrik/Gas	Sumur Bor	Ya, dengan Se	Ya	Tidak	Ya	Tidak	Ya	Sangat Layak
ABDUR	Tamat SMA	Milik Sendiri	Ya	Genteng	Tembok	Keramik/Gr	Listrik Pribadi s/d 9	Listrik/Gas	Sumur Terlindung	Ya, dengan Se	Tidak	Ya	Tidak	Tidak	Tidak	Sangat Layak
HOIFU	Tamat SMA	Menumpang	Ya	Genteng	Tembok	Keramik/Gr	Listrik Pribadi s/d 9	Listrik/Gas	Ledeng/PAM	Ya, dengan Se	Tidak	Ya	Tidak	Tidak	Tidak	Layak
PATRA	Tidak tamat	Kontrak/Sewa	Ya	Genteng	Bambu	Semen	Listrik Pribadi s/d 9	Arang/Kayu	Sumur Terlindung	Tidak, Jamba	Ya	Tidak	Ya	Ya	Ya	Sangat Layak
RB. AB	Tamat SMA	Bebas Sewa	Ya	Genteng	Tembok	Keramik/Gr	Listrik Pribadi s/d 9	Listrik/Gas	Ledeng/PAM	Ya, dengan Se	Tidak	Tidak	Tidak	Tidak	Tidak	Sangat Layak
TOHI S	Tamat SMP	Milik Sendiri	Ya	Genteng	Tembok	Keramik/Gr	Listrik Pribadi > 900	Listrik/Gas	Sumur Bor	Ya, dengan Se	Ya	Tidak	Ya	Ya	Ya	Layak
MOHA	Tamat SMA	Milik Sendiri	Tidak	Genteng	Tembok	Keramik/Gr	Listrik Pribadi s/d 9	Listrik/Gas	Sumur Bor	Ya, dengan Se	Ya	Tidak	Ya	Tidak	Ya	Layak
WARN	Tamat SD/s4	Milik Sendiri	Ya	Genteng	Tembok	Keramik/Gr	Listrik Pribadi s/d 9	Listrik/Gas	Sumur Terlindung	Ya, dengan Se	Ya	Tidak	Ya	Tidak	Ya	Sangat Layak
DARM	Tamat SMA	Milik Sendiri	Tidak	Genteng	Tembok	Keramik/Gr	Listrik Pribadi > 900	Listrik/Gas	Ledeng/PAM	Ya, dengan Se	Tidak	Tidak	Tidak	Tidak	Tidak	Layak
ASIYA	Tamat SD/s4	Milik Sendiri	Tidak	Genteng	Tembok	Keramik/Gr	Listrik Pribadi > 900	Listrik/Gas	Ledeng/PAM	Ya, dengan Se	Tidak	Tidak	Tidak	Tidak	Tidak	Sangat Layak

Gambar 3. 6 Data Latih Perhitungan Manual

Data yang digunakan sebagai data uji adalah sebagai berikut :

DATA UJI														
Nama	Pendidikan	Kepemilikan Rumah	Memor	Jenis Atap	Jenis Dinding	Jenis Lantai	Sumber Penerangan	Bahan Bak	Sumber Air Minum	Fasilitas BAB	Pener	Pener	Pener	status
ALIYAH	Tidak tamat	Milik Sendiri	Tidak	Genteng	Tembok	Semen	Listrik Pribadi s/d 9	Listrik/Gas	Sumur Bor	Ya, dengan Se	Ya	Tidak	Ya	Sangat Layak
ABDUR	Tamat SMA	Milik Sendiri	Ya	Genteng	Tembok	Keramik/Gra	Listrik Pribadi s/d 9	Listrik/Gas	Sumur Terlindung	Ya, dengan Se	Tidak	Ya	Tidak	Sangat Layak
MOHA	Tamat SMA	Milik Sendiri	Tidak	Genteng	Tembok	Keramik/Gra	Listrik Pribadi > 90Q	Listrik/Gas	Ledeng/PAM	Ya, dengan Se	Tidak	Tidak	Tidak	Layak
HAKKI	Siswa SD/24	Milik Sendiri	Tidak	Genteng	Tembok	Keramik/Gra	Listrik Pribadi > 90Q	Listrik/Gas	Air Kemasan/Isi Ula	Ya, dengan Se	Tidak	Tidak	Ya	Tidak Layak
ASIYA	Tamat SD/24	Milik Sendiri	Tidak	Genteng	Tembok	Keramik/Gra	Listrik Pribadi > 90Q	Listrik/Gas	Ledeng/PAM	Ya, dengan Se	Tidak	Tidak	Tidak	Sangat Layak
RUDIY	Tamat Pergu	Milik Sendiri	Tidak	Genteng	Tembok	Keramik/Gra	Listrik Pribadi s/d 9	Listrik/Gas	Ledeng/PAM	Ya, dengan Se	Tidak	Tidak	Tidak	Tidak Layak
DARM	Tamat SMA	Milik Sendiri	Tidak	Genteng	Tembok	Keramik/Gra	Listrik Pribadi > 90Q	Listrik/Gas	Ledeng/PAM	Ya, dengan Se	Tidak	Tidak	Tidak	Layak
MOHA	Tamat SMA	Milik Sendiri	Tidak	Genteng	Tembok	Keramik/Gra	Listrik Pribadi s/d 9	Listrik/Gas	Sumur Bor	Ya, dengan Se	Ya	Tidak	Ya	Tidak Layak
ACH. N	Tamat SMP	Milik Sendiri	Tidak	Genteng	Tembok	Keramik/Gra	Listrik Pribadi > 90Q	Listrik/Gas	Sumur Bor	Ya, dengan Se	Tidak	Tidak	Tidak	Tidak Layak

Gambar 3. 7 Data Uji Perhitungan Manual

2. Menghitung nilai probabilitas pada label kelas

Untuk menghitung probabilitas menggunakan rumus sebagai berikut :

$$P(A|B) = \frac{P(B|A).P(A)}{P(B)} \quad \dots (3.1)$$

Peneliti melakukan perhitungan manual untuk mencari nilai probabilitas ini menggunakan excel. Jika sudah menemukan hasil dari setiap atribut yang ada, kemudian hasil tersebut dijumlahkan dan mendapatkan hasil total sebagai berikut :

Probabilitas Label Kelas	
Atribut	Nilai Probabilitas
Sangat Layak	0.50
Layak	0.20
Tidak Layak	0.30
Total	1.00

3. Menghitung nilai setiap parameter

Untuk menghitung nilai setiap parameter yang ada, peneliti menggunakan rumus sebagai berikut :

$$\sigma = \sqrt{\sum_x^n (xi - \mu)} \quad \dots (3.2)$$

Perhitungan pada parameter pendidikan dengan 6 kriteria yaitu tamat perguruan tinggi, tamat SMA, tamat SMP, tamat SD, siswa SD dan tidak tamat SD.

Pendidikan Label ?	Sangat Layak	Layak	Tidak Layak
Tamat SMP/ sederajat	0.00	0.25	0.17
Tamat SMA/ sederajat	0.30	0.50	0.50
Tamat SD/ sederajat	0.50	0.00	0.00
Tamat Perguruan Tinggi	0.00	0.00	0.33
Siswa SD/ sederajat	0.00	0.25	0.00
Tidak tamat SD/ sederajat	0.20	0.00	0.00
Total	1.00	1.00	1.00

K. Rumah Label ?	Sangat Layak	Layak	Tidak Layak
Milik Sendiri	0.70	1.00	0.83
Bebas Sewa	0.20	0.00	0.00
Menumpang	0.00	0.00	0.17
Kontrak/ Sewa	0.10	0.00	0.00
Total	1.00	1.00	1.00

M. Simpanan Label ?	Sangat Layak	Layak	Tidak Layak
Ya	0.70	0.25	0.33
Tidak	0.30	0.75	0.67
Total	1.00	1.00	1.00

S. Penerangan Label ?	Sangat Layak	Layak	Tidak Layak
Listrik Pribadi > 900 Watt	0.20	1.00	0.50
Listrik Pribadi s/d 900 Watt	0.80	0.00	0.50
Total	1.00	1.00	1.00

Bahan Bakar Memasak Label ?	Sangat Layak	Layak	Tidak Layak
Listrik/ Gas	0.90	1.00	1.00
Arang/ Kayu	0.10	0.00	0.00
Total	1.00	1.00	1.00

S. Air Minum Label ?	Sangat Layak	Layak	Tidak Layak
Sumur Bor	0.10	0.25	0.67
Ledeng/PAM	0.30	0.50	0.33
Sumur Terlindung	0.60	0.00	0.00
Air Kemasan/Isi Ulang	0.00	0.25	0.00
Total	1.00	1.00	1.00

Jenis Lantai Label ?	Sangat Layak	Layak	Tidak Layak
Keramik/Granit/Marmel	0.60	1.00	1.00
Semen	0.40	0.00	0.00
Total	1.00	1.00	1.00

Fasilitas BAB Label ?	Sangat Layak	Layak	Tidak Layak
Ya, dengan Septic Tank	0.90	1.00	1.00
Tidak, Jamban Umum/Bersam	0.10	0.00	0.00
Total	1.00	1.00	1.00

Penerima BPUM Label ?	Sangat Layak	Layak	Tidak Layak
Ya	0.10	0.00	0.17
Tidak	0.90	1.00	0.83
Total	1.00	1.00	1.00

Penerima BPNT Label ?	Sangat Layak	Layak	Tidak Layak
Ya	0.30	0.25	0.17
Tidak	0.70	0.75	0.83
Total	1.00	1.00	1.00

Penerima PKH Label ?	Sangat Layak	Layak	Tidak Layak
Ya	0.10	0.25	0.00
Tidak	0.90	0.75	1.00
Total	1.00	1.00	1.00

Penerima Sembako Label ?	Sangat Layak	Layak	Tidak Layak
Ya	0.30	0.25	0.17
Tidak	0.70	0.75	0.83
Total	1.00	1.00	1.00

4. Mencari Prediksi Data Uji Dengan Klasifikasi Probabilitas Data Diatas

Data Uji ke -n	Kelas Prediksi	Sangat Layak	Layak	Tidak Layak
1	Sangat Layak	1.42851E-05	0	0
2	Sangat Layak	0.000408317	0	0
3	Layak	0.000196867	0.01977539	0.007442177
4	Layak	0	0.00494385	0
5	Sangat Layak	0.000328112	0	0
6	Tidak Layak	0	0	0.004961452
7	Layak	0.000196867	0.01977539	0.007442177
8	Tidak Layak	3.21416E-05	0	0.000297687
9	Tidak Layak	0	0.00494385	0.004961452

Peneliti melakukan perhitungan dengan data uji yang ada menggunakan rumus sebagai berikut :

$$P(X) = \sum_{n=1}^n p(x|h_j)p(h) \quad \dots (3.3)$$

Dilihat pada gambar diatas, pada data uji pertama di prediksi sebagai “sangat layak” jika dihitung, hasil yang ada hanya pada kriteria “sangat layak” karena pada kriteria tersebut nilai yang dihasilkan lebih besar daripada nilai pada kriteria yang lain sedangkan di kriteria “layak” dan “tidak layak” hasilnya menjadi nol artinya ada probabilitas atribut yang bernilai 0 dan apabila dikalikan maka hasilnya menjadi nol. Begitupun seterusnya.

3.4 Pengujian Sistem

Pengujian sistem dilakukan untuk mengetahui sistem yang dirancang telah sesuai dengan yang diharapkan atau tidak. Pengujian sistem pada penelitian ini berisi mengenai pengujian akurasi dan skenario pengujian.

3.4.1 Pengujian Akurasi

Pengujian akurasi dilakukan untuk mengetahui nilai akurasi pada sistem klasifikasi penerima bantuan yang dilakukan secara manual dengan menggunakan metode *confusion matrix*. Pengujian akurasi akan membandingkan hasil perhitungan dari sistem yang menggunakan metode *Naïve Bayes* dengan hasil data yang dilakukan oleh pihak Bappeda Sumenep untuk mengetahui seberapa akurat sistem yang dibangun. Peneliti melakukan pengujian akurasi dengan rumus sebagai berikut :

$$\frac{TP+TN}{Total} \quad \dots(3.4)$$

Keterangan :

TP (*True Positive*) = Jumlah data aktual yang memiliki nilai kelas positif dan pada model diprediksi sebagai positif

TN (*True Negative*) = Jumlah data aktual yang memiliki nilai kelas negatif dan pada model diprediksi sebagai negatif

Tabel 3. 1 *Confusion Matrix* Kelas Prediksi

Data Aktual	Kelas Prediksi		
	Sangat Layak	Layak	Tidak Layak
Sangat Layak	3	2	0
Layak	2	3	2
Tidak Layak	1	1	3

REFERENSI

- [1] Prof. Dr. B. Suyanto, *Kemiskinan, Konflik, dan Ekses Pembangunan*. Yogyakarta: Suluh Media, 2018.
- [2] H. Annur, “KLASIFIKASI MASYARAKAT MISKIN MENGGUNAKAN METODE,” vol. 10, pp. 160–165, 2018.
- [3] A. Lasarudin, H. Gani, and M. Tomayahu, “Perbandingan Metode Naïve Bayes dan C4 . 5 Klasifikasi Status Gizi Bayi Balita,” vol. 6, no. 3, pp. 273–283, 2022.
- [4] Warisa and Nurahman, “Penerapan Metode Naive Bayes Dalam Mengklasifikasi Penerima BLT Pada Desa Pelangsian,” *Jurnal JUPITER*, vol. 14, no. 2, pp. 64–70, 2022.
- [5] E. Widodo and A. Jaelani, “Penerapan Data Mining untuk Prediksi Penerimaan Bantuan Pangan Non Tunai (BPNT) di Desa Wanacala Menggunakan Metode Naive Bayes,” *Jurnal Teknologi Pelita Bangsa*, vol. 13, no. 3, pp. 153–158, 2022.
- [6] Amin Abdullah Sidiq and Febrian Wahyu Christanto, “Algoritma Naive Bayes Untuk Penentuan Pkh (Program Keluarga Harapan) Berbasis Sistem Pendukung Kepu-Tusan (Studi Kasus: Kelurahan Karanganyar Gunung Se-Marang),” *Jurnal Riptek*, vol. 14, no. 1, pp. 65–71, 2020.
- [7] W. Lidysari, H. S. Tambunan, and H. Qurniawan, “Penerapan Data Mining Dalam Menentukan Kelayakan Penerima Bantuan Sosial Pemko Dengan Algoritma C4.5 (Kasus Kantor Kelurahan Martoba),” *Kesatria : Jurnal Penerapan Sistem Informasi (Komputer dan Manajemen)*, vol. 3, no. 1, pp. 53–61, 2022, doi: 10.30645/kesatria.v3i1.97.
- [8] C. Journal, I. W. Saputro, B. W. Sari, P. Studi, I. Komputer, and J. Informatika, “Uji Performa Algoritma Naïve Bayes untuk Prediksi Masa Studi Mahasiswa,” vol. 6, no. 1, pp. 1–11, 2019.
- [9] F. K. Pratama, D. W. Widodo, and N. Shofia, “Implementasi Metode Naïve Bayes dalam Mengklasifikasi Penerima Program Keluarga Harapan (PKH) Desa Minggiran Kediri,” *Seminar Nasional Inovasi Teknologi UN PGRI Kediri*, pp. 23–28, 2021, [Online]. Available: <https://proceeding.unpkediri.ac.id/index.php/inotek/article/view/1072%0Ahttps://proceeding.unpkediri.ac.id/index.php/inotek/article/download/1072/685>
- [10] S. Q. dan R. H. Taslima, *Tinjauan Kritis Ketahanan Sosial Masyarakat Miskin Perkotaan dan Perdesaan*. Jakarta: LIPI Press, 2019.

- [11] M. L. Prayoga, M. Muchtolifah, and S. Sishadiyati, "Faktor Kemiskinan Di Kabupaten Sidoarjo," *Jambura Economic Education Journal*, vol. 3, no. 2, pp. 135–142, 2021, doi: 10.37479/jeej.v3i2.11058.
- [12] H. Annur, "KLASIFIKASI MASYARAKAT MISKIN MENGGUNAKAN METODE NAÏVE BAYES," 2018.
- [13] K. Y. Raharja, H. Oktavianto, and R. Umilasari, "Perbandingan Kinerja Algoritma Gaussian Naive Bayes Dan K-Nearest Neighbor (Knn) Untuk Mengklasifikasi Penyakit Hepatitis C Virus (Hcv)," pp. 1–12, 2021.
- [14] B. I. Saputro, "Penerapan Sistem Klasifikasi Perpustakaan Arkeologi di Perpustakaan Balai Arkeologi Daerah Istimewa Yogyakarta," *Berkala Ilmu Perpustakaan dan Informasi*, vol. 13, no. 2, p. 107, 2017, doi: 10.22146/bip.23453.
- [15] D. Normawati and S. A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," *Jurnal Sains Komputer & Informatika (J-SAKTI)*, vol. 5, no. 2, pp. 697–711, 2021, [Online]. Available: <http://ejurnal.tunasbangsa.ac.id/index.php/jsakti/article/view/369>
- [16] I. Zulfa, R. Rayuwati, and K. Koko, "Implementasi data mining untuk menentukan strategi penjualan buku bekas dengan pola pembelian konsumen menggunakan metode apriori," *Teknika: Jurnal Sains dan Teknologi*, vol. 16, no. 1, p. 69, 2020, doi: 10.36055/tjst.v16i1.7601.
- [17] D. Firdaus, "Penggunaan Data Mining dalam Kegiatan Sistem Pembelajaran Berbantuan Komputer," 2017.
- [18] A. S. Y. Adinugroho Sigit, *Implementasi Data Mining Menggunakan Weka*. Malang: UB Press, 2018.
- [19] A. I. Purnama, A. Aziz, and A. S. Wiguna, "Penerapan Data Mining Untuk Mengklasifikasi Penerima Bantuan Pkh Desa Wae Jare Menggunakan Metode Naïve Bayes," *Kurawal - Jurnal Teknologi, Informasi dan Industri*, vol. 3, no. 2, pp. 173–180, 2020, doi: 10.33479/kurawal.v3i2.348.
- [20] A. A. A. Arifin, W. Handoko, and Z. Efendi, "Implementasi Metode Naive Bayes Untuk Klasifikasi Penerima Program Keluarga Harapan," *J-Com (Journal of Computer)*, vol. 2, no. 1, pp. 21–26, 2022, doi: 10.33330/j-com.v2i1.1577.
- [21] M. K. Nur Azis, S.Kom., *Analisis Perancangan Sistem Informasi*. Bandung: Widina Bhakti Persada Bandung, 2022.

