# KLASIFIKASI DATA PENDUDUK MISKIN MENGGUNAKAN METODE *NAIVE BAYES* UNTUK MENGETAHUI PENDUDUK MISKIN PADA DESA SOPET JANGKAR SITUBONDO

**SKRIPSI**


Oleh:

**MUHAMMAD YASIN**

2021503117


**PROGRAM STUDI TEKNOLOGI INFORMASI**

**FAKULTAS SAINS DAN TEKNOLOGI UNIVERSITAS IBRAHIMY**

**SITUBONDO**

**2025**

# KLASIFIKASI DATA PENDUDUK MISKIN MENGGUNAKAN METODE *NAIVE BAYES* UNTUK MENGETAHUI PENDUDUK MISKIN PADA DESA SOPET JANGKAR SITUBONDO

**SKRIPSI**

Diajukan untuk Memenuhi Salah Satu Persyaratan dalam Menyelesaikan Program Sarjana (S-1) pada Program Studi Teknologi Informasi Fakultas Sains dan Teknologi Universitas Ibrahimy

Oleh:

**MUHAMMAD YASIN**

2021503117

**PROGRAM STUDI TEKNOLOGI INFORMASI**

**FAKULTAS SAINS DAN TEKNOLOGI UNIVERSITAS IBRAHIMY**

**SITUBONDO**

**2025**

## PERNYATAAN KEASLIAN TULISAN

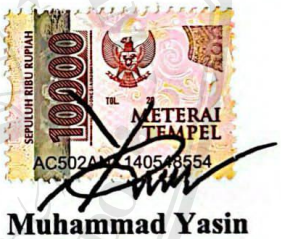
Saya yang bertanda tangan di bawah ini :

| | |
|---|---|
| Nama | : **Muhammad Yasin** |
| NPM/NIM | : 2021503117 |
| Program Studi | : S-1 Teknologi Informasi |
| Fakultas | : Fakultas Sains dan Teknologi |

Menyatakan dengan sebenarnya, bahwa tugas akhir/skripsi ini secara keseluruhan adalah hasil penelitian atau karya saya sendiri, kecuali pada bagian-bagian yang dirujuk sebagai sumber referensi dan disebutkan dalam daftar pustaka. Apabila di kemudian hari terbukti atau dapat dibuktikan bahwa tugas akhir/skripsi ini hasil plagiasi, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Situbondo, 28,01.2026

Saya yang menyatakan,




**Muhammad Yasin**

## PERSETUJUAN PEMBIMBING

Nama : **Muhammad Yasin**

NPM/NIM : 2021503117

Program Studi : **Klasifikasi Data Penduduk Miskin Menggunakan Metode Naive Bayes Untuk Mengetahui Penduduk Miskin Pada Desa Sopet Jangkar Situbondo**

Telah disetujui oleh:

| Pembimbing 1 | Pembimbing 2 |
| --- | --- |
| **Achmad Baijuri, M.Kom.** | **Nur Azise, M.Kom.** |
| NIDN: 715078902 | NIDN: 730108802 |

PENGESAHAN

**SKRIPSI**

**KLASIFIKASI DATA PENDUDUK MISKIN MENGGUNAKAN METODE *NAIVE BAYES* UNTUK MENGETAHUI PENDUDUK MISKIN PADA DESA SOPET JANGKAR SITUBONDO**

**Muhammad Yasin**
2021503117

telah dipertahankan di depan dewan penguji Sidang/Munaqasyah Skripsi pada hari senin , Tanggal 5 Mei 2025 sebagai salah satu syarat memperoleh gelar Sarjana (S.Kom) pada Fakultas Sains dan Teknologi Universitas Ibrahimy.

Tim Penguji,

Ketua Sidang,

**Dr. Ach. Khumaidi, M.P**
NIDN: 0722049001

Sekretaris Sidang,

**Muhammad Khoironi, S. Kom**
NIDN: -

Penguji I,

**Abd. Ghofur, M.Kom.**
NIDN: 0711088303

Penguji II,

**Hermanto, M.Kom.**
NIDN: 708087807

Mengetahui.
Dekan,

**Abd. Ghofur, M.Kom.**
NIDN: 0711088303

## MOTTO

**"Seberapa sulitnya dirimu dan sesakit apapun lukamu, janganlah berputus asa, teruslah kejar mimpimu untuk menjadi yang lebih baik, jadilah lebih kuat dari keadaanmu sekarang, tersenyumlah, berbahagialah, bertahanlah dan jngan pernah jatuh."**

**"ODIK TENNANG"**

## PERSEMBAHAN

1. Terimakasih kepada ayah dan ibu tercinta yang telah memberikan cinta, dan kasih sayang serta dukungan hingga saya sampai di titik ini
2. Terimakasih kepada kakak serta keluarga untuk doa, support dan pengertiannya selama ini
3. Terimakasih kepada kakak kakak tingkat yang sudah berbaik hati meluangkan waktu dan bimbingannya
4. Terimakasih kepada teman teman di Kampus Universitas Ibrahimy Kota yang selalu memberikan support serta motivasi. terkhusus kepada Gus Ahmad Gunawan dan Gus Ma'ruf Ubaidillah meski selalu "absen"
5. Terimakasih kepada semua pembaca skripsi ini

## KATA PENGANTAR

Dengan memanjatkan puji syukur ke hadirat Allah SWT, atas limpahan rahmat dan hidayah-Nya penulis dapat menyajikan Laporan Praktik Kerja Lapangan ini dengan sebaik baiknya, oleh karena itu kami mengucapkan terima kasih kepada :

1. KHR. Ach. Azaim Ibrahimy, S. Sy, M. HI Selaku Pengasuh Pondok Pesantren Salafiyah Syafi'iyah Sukorejo Situbondo.
2. KH. Ach. Fadlail, SH, M.H Selaku Rektor Universitas Ibrahimy.
3. Abd. Ghofur, M. Kom Selaku Dekan Fakultas Sains dan Teknologi Universitas Ibrahimy.
4. Dr. Ach. Khumaidi, M.P Selaku Wakil Dekan 1 Fakultas Sains dan Teknologi Universitas Ibrahimy.
5. Abd. Wafi, M.P Selaku Wakil Dekan 2 Fakultas Sains dan Teknologi Universitas Ibrahimy.
6. Ahmad Lutfi, M. Kom Selaku Wakil Dekan 3 Fakultas Sains dan Teknologi Universitas Ibrahimy.
7. Firman Santoso, M.Kom Selaku Ketua Program Studi Teknologi Informasi
8. Achmad Baijuri, M. Kom dan Nur Aizise, M.Kom. selaku pembimbing I dan II,
9. Seluruh Dosen Fakultas Sains dan Teknologi yang telah memberikan kami ilmu sehingga sampai pada masa akhir ini.

Semoga semua amal baik yang telah diberikan oleh Bapak/Ibu kepada peneliti mendapat balasan yang sebaik mungkin dari Allah SWT, Amin.

Situbondo, 17 Agustus 2025

**Penyusun**

# DAFTAR ISI

## DAFTAR TABEL

## DAFTAR GAMBAR

## DAFTAR RUMUS

# BAB I

# PENDAHULUAN

## 1.1 Latar Belakang

Kemiskinan merupakan masalah yang kompleks di Indonesia dan memiliki banyak sisi, yang telah mendapatkan perhatian sebagai masalah pembangunan. [1]. Pada Maret 2023, terdapat 25,90 juta penduduk miskin di Indonesia, dibandingkan dengan 26,36 juta orang pada Maret 2022, hal ini menunjukkan bahwa penduduk yang hidup dalam kemiskinan masih sangat tinggi dengan nilai proporsi 9,36% [2]. Data penduduk miskin sangat penting untuk menentukan kebijakan dan program bantuan sosial yang tepat sasaran. Namun, proses identifikasi penduduk miskin sering kali menghadapi kendala, yaitu terjadinya kesalahan dalam pencatatan dan kurangnya sistem klasifikasi yang efektif.

Saat ini pendataan penduduk miskin di Desa Sopet, sudah dilakukan secara terkomputerasisasi menggunakan *Microsoft Office Excel*, akan tetapi penghitungan menggunakan *Microsoft Office Excel* masih memiliki benyak kelemahan yaitu pada pengelolaan data yang sangat kompleks dengan ratusan ribu baris, *Excel* dapat berubah menjadi program yang sangat lambat untuk dioperasikan[3]. sehingga dapat menyebabkan perkiraan yang tidak akurat mengenai kondisi masyarakat miskin. Selain itu, tidak adanya pendekatan yang baku dalam pengolahan data sering kali menyebabkan salah sasaran dalam pemberian bantuan. Sehingga diperlukan sebuah metode yang dapat mengklasifikasikan masyarakat miskin dengan lebih tepat.

Berdasarkan permasalahan yang ada maka diusulkan sebuah metode klasifikasi penduduk miskin secara akurat berdasarkan data sosio-ekonomi masyarakat untuk mendukung program bantuan sosial dan pemberdayaan masyarakat. Penggunaan teknologi selain membantu pengolahan data yang berguna untuk perencanaan pembangunan, teknologi juga dapat meningkatkan pengambilan keputusan dan menawarkan sejumlah keuntungan tambahan[4]. yaitu membantu pemerintah desa dalam membuat keputusan yang lebih efektif dan efisien.

Algoritma *Naive Bayes* adalah sebuah metode klasifikasi yang berdasarkan pada teorema Bayes dengan asumsi bahwa semua fitur yang terlibat dalam klasifikasi adalah independen secara statistik. Meskipun asumsi ini jarang benar dalam situasi dunia nyata, metode ini tetap efektif dan efisien dalam banyak kasus, terutama ketika dimensi data cukup tinggi. Metode *Naive Bayes* digunakan dalam berbagai aplikasi, termasuk analisis sentimen, klasifikasi teks, deteksi spam, pengenalan pola, dan banyak lagi. Ini bekerja dengan memanfaatkan probabilitas dan membuat asumsi sederhana tentang ketergantungan antara fitur-fitur dalam data. Keuntungan dari *Naive Bayes* termasuk sederhana dan cepat dalam pelatihan dan pengujian, serta performa yang relatif baik dalam beberapa kasus. Namun, asumsi independensi fitur sering kali tidak realistis dalam situasi dunia nyata, yang dapat menyebabkan kinerja yang lebih rendah jika asumsi ini tidak terpenuhi [5].

Seperti penelitian yang dilakukan oleh Nur Madia, dkk. dengan judul Penentuan Kelayakan Masyarakat Miskin Penerima Bantuan Menggunakan Metode *Naive Bayes* (Studi Kasus: Kabupaten Penajam Paser Utara), menganalisis kelayakan masyarakat miskin. Hasil penelitian ini memperoleh

bahwa metode *Naive Bayes* dapat menghasilkan *accuracy* tertinggi pada skenario ketiga yaitu 60% atau 328 data latih dan 40% atau 218 data uji dengan nilai *accuracy* sebanyak 77,98%[6]. Dalam konteks klasifikasi di Desa Sopet diharapkan dapat membantu pemerintah daerah dalam menentukan kebijakan yang lebih tepat sasaran dengan mengategorikan data penduduk miskin menjadi sangat miskin, miskin, hampir miskin, rentan miskin dan tidak miskin.

Untuk mengetahui klasifikasi penduduk miskin di Desa Sopet secara lebih tepat dan metodis, penelitian ini bertujuan untuk membuat model kategorisasi penduduk miskin dengan menggunakan pendekatan *Naive Bayes*. Metode ini diharapkan dapat membantu pemerintah desa dalam mengklasifikasikan warga sesuai dengan status kesejahteraannya, sehingga kebijakan dan inisiatif bantuan sosial dapat lebih tepat sasaran. Pemerintah Desa diharapkan dapat mengambil manfaat dari penelitian ini untuk meningkatkan efisiensi pengumpulan dan pengelolaan data kemiskinan, serta dapat menurunkan kemungkinan kesalahan dalam menilai tingkat kemiskinan seseorang, dan menawarkan solusi berbasis teknologi untuk memfasilitasi pengambilan keputusan ditingkat desa yang lebih adil dan efektif.

## 1.2 Identifikasi Masalah

Berdasarkan latar belakang yang telah dipaparkan, untuk ditentukan identifikasi masalah yaitu pendataan penduduk miskin pada Desa Sopet sudah dilakukan secara terkomputrisasi menggunakan *Microsoft Office Excel.* Akan tetapi pendataan dan perhitungan menggunakan rumus di *Microsoft Office Excel* masih

memiliki banyak kelemahan pada pengelolaan data, diantaranya data yang membutuhkan banyak baris dan kolom, hal tersebut menyebabkan proses pendataan di *Microsoft Office Excel* membutuhkan proses yang lama.

## 1.3 Rumusan Masalah

Berdasarkan latar belakang dan identifikasi masalah di atas, rumusan masalah dalam penelitian ini yang akan dibahas yaitu "bagaimana membuat program aplikasi untuk mengklasifikasikan data penduduk miskin di Desa Sopet menggunakan *Algoritma Naive Bayes*"

## 1.4 Batasan Masalah

Penelitian ini dibatasi pada masyarakat Desa Sopet, Kecamatan Jangkar, Kabupaten Situbondo, dengan memanfaatkan data yang diperoleh langsung dari pemerintah desa. Fokus kajian hanya pada proses klasifikasi penduduk miskin berdasarkan indikator sosial ekonomi yang tersedia, menggunakan *algoritma Naive Bayes* sebagai metode utama. Analisis variabel dibatasi pada data kependudukan desa, sedangkan faktor lain di luar data tersebut tidak menjadi bagian penelitian. Temuan yang dihasilkan hanya berlaku untuk klasifikasi penduduk miskin di Desa Sopet dan tidak dimaksudkan untuk digeneralisasi pada wilayah lain.

## 1.5 Tujuan Penelitian

Tujuan penelitian ini yaitu membuat program aplikasi untuk mengetahui performa Algoritma *Naive Bayes* dalam mengklasifikasikan data penduduk miskin di Desa Sopet.

## 1.6 Manfaat Penelitian

Manfaat dari penelitian ini antara lain klasifikasi data penduduk miskin menggunakan algoritma *Naive Bayes* memiliki beberapa manfaat:

a. Dengan adanya klasifikasi data penduduk miskin yang lebih akurat, distribusi bantuan sosial dapat dilakukan secara lebih adil dan tepat sasaran, sehingga diharapkan dapat mempercepat peningkatan kesejahteraan masyarakat di Desa Sopet.

b. Penilitian ini dapat menjadi refrensi bagi akademisi dan peneliti lain dalam mengembangkan metode klasifikasi data penduduk miskin menggunakan *Naive Bayes,* Serta memperkaya literatur ilmiah di bidang *data mining* dan analisi sosial ekonomi.

## 1.7 Metode Penelitian

### 1.7.1 Jenis Penelitian

**a. Kuantitatif**

Metodologi penelitian yang digunakan dalam studi ini adalah deskriptif kuantitatif. Tujuan penelitian deskriptif kuantitatif adalah untuk mendeskripsikan, menyelidiki, dan menjelaskan suatu subjek selama proses penyelidikan berlangsung, dengan memanfaatkan data numerik untuk menarik kesimpulan dari kejadian yang dapat diamati.[7].

**b. Eksperimen**

Penelitian eksperimen adalah untuk menunjukkan bagaimana suatu pengobatan mempengaruhi hasil pengobatan tersebut. Menurut Arikunt, ilmuwan menggunakan eksperimen untuk secara sengaja menciptakan

situasi atau peristiwa tertentu, lalu menganalisis hasilnya. Pendekatan untuk menentukan hubungan sebab-akibat antara dua hal adalah penelitian eksperimental, dengan kata lain. Penelitian lapangan adalah metode penelitian yang digunakan. Penelitian yang mengamati fenomena dalam habitat alaminya disebut penelitian lapangan.[8].

### 1.7.2 Metode Pengumpulan Data

Pada metode penelitian ini terdapat beberapa metode pengumpulan data yang digunakan diantaranya :

a. Observasi

Pengamatan adalah salah satu teknik untuk mengumpulkan data atau informasi, yang melibatkan pemantauan aktivitas yang sedang berlangsung. Pengamatan dilakukan untuk memberikan gambaran yang akurat tentang suatu peristiwa atau kejadian guna menjawab pertanyaan penelitian.[9]. sehingga dengan penelitian ini dapat pengetahui data dan kondisi di Balai Desa Sopet.

b. Wawancara

Wawancara adalah salah satu teknik untuk mengumpulkan data penelitian. Secara singkat, wawancara adalah proses atau kesempatan ketika pewawancara berbicara langsung dengan narasumber atau sumber informasi.[10]. Wawancara ini dilakukan pada bagian aparatur balai Desa Sopet.

c. Dokumentasi

Kata “dokumentasi” diambil dari bahasa Inggris. Dokumentasi memiliki dua arti, menurut situs web Kamus *Oxford Learner's Dictionary* yang resmi. Pertama, berkaitan dengan penyediaan dokumen resmi atau informasi yang berguna untuk menjaga catatan. Kedua, mencakup proses mendokumentasikan dan mengklasifikasikan data dalam berbagai format, termasuk teks, gambar, video, dan lainnya. Pencarian sistematis, penggunaan, penyelidikan, pengumpulan, dan penyediaan dokumen untuk mengumpulkan pengetahuan, informasi, dan bukti, serta mendistribusikannya kepada pihak yang berkepentingan, dikenal sebagai dokumentasi.[11]. Dokumen yang diteliti adalah terkait data penduduk yang ada pada Desa Sopet.

d. Studi Literatur

Proses pencarian referensi teoretis yang berkaitan dengan kasus atau topik yang sedang diteliti untuk menganalisisnya dengan cara menelusuri materi yang telah diterbitkan sebelumnya dikenal sebagai studi literatur. Para peneliti dapat memanfaatkan semua data dan konsep yang relevan serta mengembangkan pemahaman yang lebih komprehensif dan mendalam tentang topik yang sedang diteliti dengan melakukan tinjauan literatur.[12].

## 1.8 Sistematika Pembahasan

Sistematika pembahasan dalam karya tulis ini dibagi menjadi lima bab, yang memiliki suatu tujuan tertentu dalam setiap bab:

**BAB I : PENDAHULUAN**

Dalam pendahuluan berisi tentang latar belakang, identifikasi masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, metode penelitian dan sistematika pembahasan.

**BAB II : TINJAUAN PUSTAKA**

Pada bab dua dipaparkan tentang tinjauan pustaka, hal-hal yang terkait serta referensi penunjang yang sesuai dengan judul laporan yang diangkat, juga memuat landasan teori dan perangkat lunak yang akan digunakan.

**BAB III : ANALISIS DAN PERANCANGAN**

Pada bab ini menjelaskan tentang gambaran alur proses dan desain dari objek penelitian yang diteliti,bterdapat beberapa sub bab pada bagian ini yakni identifikasi dan analisis proses bisnis, identifikasi dan analisis kebutuhan, identifikasi dan alternatf solusi, sedangkan desain sistem mebahas tentang perancangan alur sitem.

**BAB IV : HASIL DAN PEMBAHASAN**

Bab ini adalah hasil dari Analisa yang telah diputuskan sebelumnya. Berisi tentang beberapa penjelasan maupun perhitungan dari data yang telah diolah sehingga menghasilkan sebuah prediksi berdasarkan kemungkinan yang telah didapatkan sebelumnya dan juga tingkat akurasi dari prediksi yang telah dihasilkan.

## BAB V : PENUTUP

Kesimpulan yang merangkum temuan penelitian utama, seperti hasil akurasi model, efektivitas teknik yang digunakan, atau pola signifikan yang teridentifikasi dalam data, semua telah disesuaikan untuk memenuhi tujuan studi. Bagian ini juga memberikan rekomendasi pada penelitian selanjutnya, seperti penggunaan algoritma baru, pengintegrasian data yang lebih beragam, pengujian dalam berbagai kondisi, atau penerapan sistem dalam praktik untuk penilaian tambahan.

# BAB II

# TINJAUAN PUSTAKA

## 2.1 Penelitian Terdahulu

### 2.1.1 Klasifikasi Tingkat Rumah Tangga Miskin Saat Pandemi Dengan *Naive Bayes Classifier*

Penelitian dilakukan oleh Harliana dan Fatra Nonggala Putra dari Program Studi Ilmu Komputer, FakultasEksakta, Universitas Nahdlatul Ulama Blitar. Dipublikasi pada November 2021 di Jurnal Sains dan Informatika.

Di Indonesia, persentase keluarga miskin meningkat menjadi 9,78%, sebagian besar disebabkan oleh pandemi Covid-19. Metode tradisional dalam menentukan status kemiskinan berdasarkan sampel rumah tangga dipandang kurang efektif, terutama karena lanskap sosial-ekonomi yang berubah dengan cepat. Oleh karena itu, diperlukan cara yang lebih cepat dan tepat untuk mengkategorikan status kemiskinan masyarakat, terutama dalam situasi darurat seperti pandemi.

Meskipun pemerintah telah membagi bantuan ke dalam tiga kelompok, banyak rumah tangga yang berpindah antar kelompok akibat perubahan kondisi ekonomi, seperti dari tidak miskin menjadi sangat miskin. Distribusi bantuan yang tidak akurat muncul dari metode manual untuk mengklasifikasikan status ekonomi. Memahami situasi sebenarnya dari masyarakat miskin selama epidemi semakin diperumit oleh keterbatasan data dan teknik analisis tradisional.

Untuk mengatasi masalah ini, penelitian ini menggunakan data dari Survei Sosial Ekonomi Nasional 2020 untuk mengkategorikan tingkat kemiskinan rumah tangga dengan menggunakan algoritma *Naïve Bayes Classifier*. Berdasarkan hasil pengujian dengan menggunakan pendekatan validasi silang 10 kali lipat, algoritme

ini dapat memperoleh akurasi hingga 93,21%, presisi 86,3%, dan *recall* 80,11%. Hal ini menunjukkan bahwa *Naïve Bayes Classifier* merupakan cara yang praktis dan efisien untuk menentukan status kemiskinan seseorang secara lebih akurat sehingga dapat membantu proses pengambilan keputusan dalam pengalokasian bantuan sosial[13].

### 2.1.2 Klasifikasi Masyarakat Miskin Menggunakan Metode *Naive Bayes*: Studi Kasus di Desa Jati Mulya Kecamatan Sepatan Timur

Penelitian dilakukan oleh Dede Latipah, Mochamad Arief Mardiansah, Esthi Adityarini dari Ilmu Komputer, Kesehatan dan Sains, Universitas Muhammadiyah Bogor Raya Kabupaten Bogor, Jawa Barat Magister Teknik Informatika, Universitas Pamulang Tangerang Selatan, Banten Manajemen Bisnis Syariah, Institut Daarul Qur'an Jakarta Kota Tangerang, Banten. Dipublikasi pada 26 Juli 2024 di JURNAL MULTINETICS.

Pada penelitian ini secara konkret berangkat dari permasalahan kemiskinan di Indonesia, khususnya di Desa Jati Mulya, Kecamatan Sepatan Timur. Kemiskinan di wilayah ini menyebabkan keterbatasan akses masyarakat terhadap kebutuhan dasar seperti pangan, tempat tinggal, pendidikan, dan kesehatan. Pemerintah telah berupaya mengatasi masalah ini dengan berbagai program bantuan sosial. Namun, distribusi bantuan sering kali tidak merata dan kurang tepat sasaran, sehingga masih banyak masyarakat miskin yang tidak mendapatkan bantuan yang semestinya.

Untuk mengatasi permasalahan tersebut, penelitian ini mengusulkan penerapan metode klasifikasi menggunakan algoritma *Naive Bayes* dengan

bantuan aplikasi *Orange*. Metode ini dipilih karena kemampuannya dalam menangani berbagai variabel dan memberikan hasil klasifikasi yang cukup akurat. Data yang digunakan mencakup atribut seperti umur, pendidikan, pekerjaan, penghasilan, jumlah tanggungan, dan status perkawinan.

Penelitian ini bertujuan untuk mengembangkan sistem klasifikasi masyarakat miskin yang lebih akurat dan dapat digunakan sebagai referensi bagi pengambil kebijakan dalam mendistribusikan bantuan sosial secara lebih tepat sasaran. Dengan tingkat akurasi 96%, sistem ini diharapkan dapat membantu pemerintah desa dalam mengidentifikasi dan memverifikasi masyarakat miskin dengan lebih efektif[14]

### 2.1.3 Penerapan *Naive Bayes* dalam Mengklasifikasikan Masyarakat Miskin di Desa Lepak

Penelitian dilakukan oleh Wiwit Pura Nurmayanti, Dinda Ayu Lara Saky, Muhammad Malthuf, Muhammad Gazali, Ristu Haiban Hirzi, dari Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Hamzanwadi, Selong, Indonesia, Program Studi Tadris IPS Fakultas Tarbiyah dan Keguruan Universitas Islam Negeri Mataram, Mataram, Indonesia Dipublikasi pada 1 Juni 2021 di Geodika: Jurnal Kajian Ilmu dan Pendidikan Geografi.

Pada penelitian ini berfokus pada permasalahan kemiskinan di Desa Lepak, Lombok Timur, Nusa Tenggara Barat. Kemiskinan merupakan masalah utama di Indonesia yang berdampak pada berbagai aspek sosial, ekonomi,

dan politik. Berdasarkan data BPS tahun 2020, NTB termasuk dalam 10 provinsi dengan tingkat kemiskinan tertinggi di Indonesia.

Desa Lepak memiliki angka kemiskinan yang signifikan, namun data mengenai penduduk miskin di desa ini belum dimanfaatkan untuk analisis lebih lanjut, melainkan hanya menjadi arsip pemerintah desa. Oleh karena itu, penelitian ini bertujuan untuk memanfaatkan data tersebut dengan metode klasifikasi menggunakan *Naive Bayes Classifier***,** sebuah teknik dalam data mining, untuk mengidentifikasi masyarakat miskin dan tidak miskin.

Hasil klasifikasi ini, menunjukkan bahwa pendekatan *Naive Bayes* dapat mengkategorikan masyarakat ke dalam kelompok miskin dan tidak miskin dengan persentase akurasi sebesar 96,63%. Berdasarkan hasil pengujian dengan menggunakan confusion matrix, 148 dari 156 prediksi yang dibuat oleh kelompok miskin sesuai dengan kondisi aktual, sedangkan 110 dari 111 prediksi yang dibuat oleh kelompok tidak miskin dikategorikan secara akurat. Selain itu, pendekatan ini diklasifikasikan sebagai sangat baik dan berhasil dalam proses pengkategorian data kemiskinan di Desa Lepak karena memiliki nilai sensitivitas atau *recall* sebesar 94,87% untuk prediksi individu miskin dan spesifisitas sebesar 99,10% untuk prediksi individu tidak miskin[15].

## 2.2 Landasan Teori

### 2.2.1 Klasifikasi

Klasifikasi dapat dipahami sebagai suatu proses pengelompokan objek berdasarkan karakteristik tertentu yang dimilikinya. Pada dasarnya, klasifikasi merupakan proses pembelajaran (training) untuk membentuk suatu fungsi dengan

target tertentu, sehingga mampu memetakan setiap himpunan objek ke dalam salah satu label kelas yang telah ditentukan sebelumnya. Melalui klasifikasi, analisis dapat dilakukan untuk membantu mengidentifikasi label kelas dari sejumlah sampel yang hendak dikategorikan[16].

Sebagai salah satu bentuk dasar dari analisis data, klasifikasi sering pula dipandang sebagai metode yang digunakan untuk menentukan kategori kelas dari data yang tersedia. Metode ini sangat sesuai digunakan untuk mendeskripsikan dataset dengan tipe data biner maupun nominal. Namun demikian, kelemahan teknik ini adalah kurang tepat apabila diterapkan pada data bertipe ordinal, sebab pendekatan yang digunakan hanya bersifat implisit terhadap kategori data tersebut[16].

### 2.2.2 Data Mining

Data mining merupakan suatu proses untuk menggali informasi penting atau menemukan pola tersembunyi dari sekumpulan data berukuran besar. Inti dari penerapan data mining adalah mengidentifikasi pola yang sebelumnya belum pernah terdeteksi, sehingga dapat memberikan wawasan baru yang bermanfaat serta mendukung proses pengambilan keputusan secara lebih akurat[17].

Secara lebih luas, data mining dapat dipahami sebagai tahapan analisis yang berfungsi untuk menemukan hubungan, kecenderungan, maupun kebiasaan tertentu dalam data berskala besar. Proses ini memanfaatkan beragam pendekatan, seperti teknik statistik, metode matematis, kecerdasan buatan, hingga algoritma pembelajaran mesin, untuk menghasilkan informasi maupun pengetahuan yang relevan dari sebuah basis data. Dengan kata lain, data mining merupakan perpaduan

dari berbagai disiplin ilmu, termasuk statistik, sistem basis data, machine learning, pengenalan pola, hingga visualisasi, yang keseluruhannya digunakan untuk mengelola serta mengekstraksi informasi berharga dari database yang kompleks dan berjumlah besar[17].

### 2.2.3 Data

kualitas dan akurasi data memiliki dampak langsung terhadap validitas dan akurasi hasil yang dihasilkan, data menjadi dasar utama dalam penelitian. Oleh karena itu, melakukan pengumpulan dan analisis data yang teliti sangat penting untuk menghasilkan penelitian yang dapat mengembangkan pengetahuan atau memecahkan masalah dalam konteks yang lebih luas. Ketidakadaan data yang solid dapat meningkatkan kemungkinan kesimpulan penelitian menjadi bias, tidak representatif, atau tidak dapat dipercaya, yang pada akhirnya dapat mengancam tujuan studi. Oleh karena itu, melakukan pengumpulan dan analisis data yang teliti sangat penting untuk menghasilkan penelitian yang mengembangkan pengetahuan atau memecahkan masalah.[18].

Tantangan dalam pemilihan dan pengumpulan data yang relevan dengan tujuan penelitian sering kali berkaitan dengan ketersediaan, kualitas, serta kesesuaian data dengan pertanyaan penelitian. Peneliti perlu memastikan bahwa data yang dikumpulkan tepat sasaran, representatif, dan bebas dari bias, yang memerlukan perencanaan matang dan pemahaman mendalam tentang metode pengumpulan data[18].

### 2.2.4 Penduduk

Penduduk adalah individu atau kelompok individu yang tinggal disuatu tempat (desa, negara, pulau, dll.). Dengan kata lain, penduduk adalah individu atau kelompok individu yang menjadikan suatu wilayah teritorial sebagai tempat tinggal mereka. Kewarganegaraan diwajibkan bagi semua penduduk. Segala hal yang berkaitan dengan warga negara dianggap sebagai kewarganegaraan[19].

"Penduduk adalah warga negara Indonesia dan orang asing yang tinggal di Indonesia," menurut Pasal 1 Ayat 2 Undang-Undang Administrasi, yang mendefinisikan penduduk. Pasal 26(2) Undang-Undang Dasar Negara Republik Indonesia Tahun 1945 (UUD 1945) juga mendefinisikan penduduk, dengan menyatakan bahwa "penduduk merujuk pada warga negara Indonesia dan warga negara asing yang tinggal di Indonesia"[19].

Salah satu persyaratan paling penting dan mendasar dalam pembentukan suatu bangsa adalah populasinya; suatu bangsa tidak dapat eksis tanpa populasi tersebut. Baik imigran maupun warga negara Indonesia membentuk populasi tersebut. Dengan kata lain, warga asing yang tinggal di Indonesia dianggap sebagai bagian dari komunitas Indonesia[19].

### 2.2.5 Miskin

Peraturan Daerah Kabupaten Situbondo Nomor 7 Tahun 2022, Bab 1, Ketentuan Umum, Pasal 1, mendefinisikan kemiskinan sebagai ketidakmampuan individu, keluarga, atau masyarakat untuk memenuhi kebutuhan dasar mereka guna mempertahankan dan mengembangkan kehidupan yang bermartabat. Definisi ini didasarkan pada indikator kemiskinan regional yang telah disesuaikan dengan

standar Tim Nasional Percepatan Penanggulangan Kemiskinan (TNP2K) untuk kriteria penilaian kemiskinan.[20].

Seseorang atau keluarga dianggap berada dalam kemiskinan jika pendapatan mereka sangat rendah sehingga tidak dapat memenuhi kebutuhan dasar, termasuk memiliki rumah yang layak, makan tiga kali sehari, akses ke layanan kesehatan yang berkualitas, dan pendidikan yang layak. Sebagai contoh, seorang buruh tani yang berpenghasilan Rp500.000 per bulan dan harus menafkahi istri dan tiga orang anak sering kali kesulitan untuk membayar listrik, membeli makanan, atau menyekolahkan anaknya. Kurangnya lapangan pekerjaan, meningkatnya biaya hidup, dan terbatasnya akses terhadap bantuan sosial dari pemerintah dapat memperparah kemiskinan. Masyarakat miskin tidak memiliki tiga modal utama. Pertama, modal manusia yang mencakup kemampuan, pola makan, dan kesehatan yang diperlukan untuk berkontribusi pada perekonomian. Kedua, modal usaha, yang terdiri dari peralatan dan fasilitas motor elektronik yang digunakan dalam industri, termasuk sektor jasa dan pertanian. Ketiga, infrastruktur mencakup hal-hal seperti air, sanitasi, listrik, dan telekomunikasi [21].

### 2.2.6 *Naive Bayes*

Salah satu teknik data mining yang didasarkan pada penerapan teorema Bayes untuk masalah klasifikasi adalah *Naive Bayes*. Untuk setiap nilai kemunculan karakteristik target pada setiap sampel data, *Naive Bayes* akan menentukan probabilitas posteriornya. Sampel data selanjutnya akan

dikategorikan dengan menggunakan *Naive Bayes* ke dalam kelas dengan nilai probabilitas posterior terbesar.[22]

*Naive Bayes* yang digunakan dalam metode penggalian data untuk kategorisasi. *Naive Bayes*, yang sering dikenal sebagai teorema *Bayes*, adalah klasifikasi yang menggunakan teknik statistik dan probabilitas yang diusulkan oleh Thomas *Bayes*, seorang ilmuwan. Metode ini memperkirakan peluang di masa depan berdasarkan kinerja masa lalu. Nilai probabilitas adalah dasar dari penelitian *Naive Bayes*[22].

Persamaan yang digunakan dalam *Naive Bayes* antara lain :

a. Menghitung probabilitas dari Vj

$$\boldsymbol{P(v_j) = \frac{|doc\,j|}{|training|}} \quad \textbf{(2.1)}$$

P(vj) adalah probabilitas kategori vj (kategori penduduk miskin)

Docj adalah data testing

b. Menghitung probabilitas kata ai

$$\boldsymbol{P(a_i|v_j) = \frac{ni+1}{|n+kosakata|}} \quad \textbf{(2.2)}$$

P($ai/vj$ ) adalah probabilitas setiap atribut ai terhadap kategori vj

ni adalah jumlah kemunculan kata/atribut $ai$ dalam kategori $vj$

n adalah total kemunculan seluruh kata/atribut dalam kategori vj

Kosakata adalah data training

c. Mengghitung probabilitas pada persamaan 2.1

$$P(X|H) = \frac{P(X|H)P(H)}{P(X)} \quad \textbf{(2.3)}$$

Keterangan :

X : Data dengan class yang belum diketahui

H : Hipotesis data X merupakan suatu class spesifik

P(X|H) : Probabilitas hipotesis H berdasarkan kondisi X (*posteriori probability*)

P(H) : Probabilitas hipotesis H (*posteriori probability*)

d. Menghitung nilal maksimum pada nilal persamaan 2.2

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)} \quad (\mathbf{2.4})$$

Keterangan :

P(X) : bernilai konstan untuk semua class

P(C) : merupakan frek relatif sample class C

Dicari P(C|X) bernilai maksimum, sama halnya dengan P(X|C) •P(C) juga bernilai maksimum.

e. Menghitung persamaan dari *Naive Bayes* pada persamaan 2.3

$$P(Ci \mid X) = \frac{P(X|\ CDP(Ci)}{P(X)} \quad (\mathbf{2.5})$$

Keterangan :

X : Kriteria suatu kasus berdasarkan masukan

Ci : Kelas solusi pola ke-i, dimana i adalah jumlah label kelas

P(Ci|X) : Probabilitas label kelas Ci dengan kriteria masukan X

P(X|Ci) : Probabilitas kriteria masukan X dengan label kelas Ci

P(Ci) : Probabilitas label kelas Ci

### 2.2.7 *Confusion Matrix*

Menilai efektivitas proses klasifikasi sangat penting. Kinerja proses klasifikasi menunjukkan seberapa efektif sistem mengklasifikasikan data. Matriks kebingungan adalah alat yang berguna untuk mengevaluasi kinerja proses ini. Matriks kebingungan adalah matriks yang mencakup data klasifikasi aktual dan prediktif dari sistem klasifikasi. Tabel matriks yang menampilkan kinerja model klasifikasi pada kumpulan data uji dengan nilai yang diketahui disebut matriks kebingungan.[23]. Berikut rumus dari *Confusion matrix*.

a. Menghitung *accuracy*, berikut rumus untuk menghitung *accuracy* :

$$\text{Accuracy} = \frac{\sum TP}{\sum Total\ Data} \quad \textbf{(2.6)}$$

b. Mwnghitung *precision*, berikut rumus untuk menghitung *precision* :

$$Precision = \frac{TP}{TPi + \sum FP} \quad \textbf{(2.7)}$$

c. Menghitung *recall*, berikut rumus untuk menghitung *recall* :

$$\text{Recall} = \frac{TP}{TPi + \sum FN} \quad \textbf{(2.8)}$$

d. Nebghitung *F1-Score*, berikut rumus untuk menghitung *F1-Score* :

$$\text{F1} = 2\ \chi \frac{Precision\ x\ Recall}{Precision + Recall} \quad \textbf{(2.9)}$$

### 2.2.8 Python

Python merupakan bahasa pemrograman yang bersifat dinamis dan banyak dimanfaatkan dalam pengembangan aplikasi di berbagai bidang. Karakteristik ini memberikan fleksibilitas bagi programmer untuk menuliskan program dengan

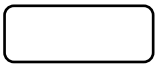
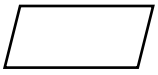
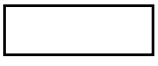
beragam pendekatan. Sebagai contoh, antarmuka grafis dapat dibangun menggunakan paradigma berorientasi objek, sementara proses pengolahan data bisa dilakukan dengan pendekatan fungsional maupun prosedural. Bahasa pemrograman ini juga menyediakan berbagai fitur yang mendukung kebutuhan pengembang perangkat lunak.

Selain dikenal sebagai bahasa yang fleksibel, Python juga memiliki keunggulan lain, yaitu kemampuannya dalam membangun aplikasi dengan Graphic User Interface (GUI) yang lebih menarik dan interaktif[24].

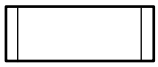
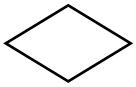
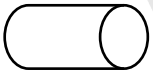
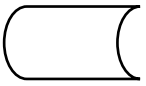
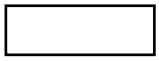
### 2.2.9 Pemodelan

Sistem Aliran, sebagai bagian dari alur kerja keseluruhan sistem, merupakan alat pemodelan yang digunakan. Diagram ini menggambarkan urutan operasi dalam sistem, dan saat merancang aliran, hal ini sering ditentukan oleh fungsi yang mengimplementasikan atau mengawasi subsistem[25]. Simbol *Flowcahrt* ditunjukkan pada tabel 2.1 dibawah ini.

**Tabel 2. 1 Simbol-simbol Bagan Aliran Sistem *(System Flowchart)***

| No | Simbol | Nama | Arti |
|---|---|---|---|
| 1. |  | Terminal | Awal akhir *flowchart* |
| 2. |  | *Input / output* | input data atau output data- data yang di proses atau informasi |
| 3. |  | Proses | Mempresentasikan Operasi |
| 4. |  | Anak Panah | Mempresentasikan alur kerja |

*Tabel 2. 1 Simbol-simbol Bagan Aliran Sistem (System Flowchart) Lanjutan*

| No | Simbol | Nama | Arti |
|---|---|---|---|
| 5. |  | Penghubung | Keluar atau masuk dari bagian lain *flowchart* |
| 6. |  | Sub proses | Rincian operasi berada di tempat lain |
| 7. |  | Keputusan | Kebutuhan dalam proram |
| 8. |  | *Magnetik Disk* | I/O yang menggunakan disk magnetik |
| 9. |  | *Punched Tape* | I/O yang menggunakan pitakes berhubung |
| 10. |  | *Punched card* | I/O yang menggunakan kartu berhubung |
| 11. |  | Magnetik Drum | I/O yang menggunakan drum magnetik |
| 12. |  | *On Line Strorage* | I/O yang menggunakan penyimpanan akan langsung |
| 13. |  | Proses | Mempresentasikan Operasi |

## 2.3 Perangkat Lunak Yang Digunakan

### 2.3.1 *Microsoft Office Excel*

Microsoft Office Excel, yang lebih dikenal dengan sebutan Microsoft Excel, adalah perangkat lunak lembar kerja yang dikembangkan dan dipasarkan

oleh Microsoft Corporation. Beberapa pakar juga memberikan definisi tersendiri mengenai aplikasi ini. Sesuai namanya, Microsoft Excel merupakan bagian dari paket Microsoft Office. Program ini dapat dijalankan pada sistem operasi Windows maupun MacOS. Salah satu keunggulan utama dari Excel adalah kemampuannya dalam melakukan perhitungan serta menyajikan data dalam bentuk grafik. Dengan dukungan strategi pemasaran yang intensif dari Microsoft, tidak mengherankan jika Excel telah menjadi salah satu aplikasi komputer yang sangat populer dan masih digunakan secara luas hingga sekarang[26].

### 2.3.2 *Rapid Miner*

Markus Hofmann dari *Institute of Technology* Blanchardstown dan Rafl Klinkenberg dari rapid-i.com mengembangkan program Rapid Miner, yang memiliki tampilan antarmuka pengguna grafis (GUI) untuk memudahkan orang dalam mengoperasikannya. Rapid Miner dapat digunakan pada sistem operasi apa pun dan merupakan perangkat lunak sumber terbuka yang dikembangkan dengan bahasa pemrograman Java di bawah Lisensi Publik GNU. Karena semua kemampuan yang dibutuhkan sudah tersedia, menggunakan Rapid Miner tidak memerlukan pengetahuan coding khusus. Rapid Miner didesain secara khusus untuk pengguna data mining[27].

### 2.3.3 *Google colab*

Google Colab adalah layanan cloud computing gratis dari Google yang menyediakan lingkungan pengembangan Python yang interaktif dan dapat diakses melalui browser web. Layanan ini memungkinkan pengguna untuk menulis,

menjalankan, dan berbagi kode Python, serta melakukan analisis data dan machine learning dengan mudah tanpa perlu menginstal perangkat lunak di komputer lokal.

Artinya, dengan menggunakan Google Colab kita tidak perlu menginstal sebuah apikasi tambahan ataupun tools seperti yang kita kenal selama ini, seperti IDE. Dengan memanfaatkan google colab kita bisa membuat sebuah program python dengan gratis dan tidak memerlukan komputer dengan spesifikasi tertentu seperti halnya pada IDE lainnya.

Google Colab menyediakan lingkungan pengembangan Python yang lengkap, termasuk dukungan untuk paket-paket populer seperti NumPy, Pandas, TensorFlow, dan Keras. Pengguna dapat mengunggah kode Python mereka ke Google Colab, menjalankannya pada mesin virtual yang disediakan oleh Google, dan menyimpan output dan hasilnya pada Google Drive[28].

# BAB III

# ANALISIS DAN PERANCANGAN

## 3.1 Gambaran Umum Objek Penelitian

Objek penelitian ini berfokus pada orang dengan kondisi sosial ekonomi yang terbatas, yang ditandai dengan rendahnya pendapatan, terbatasnya akses terhadap layanan dasar seperti perumahan yang layak, layanan kesehatan, dan pendidikan, serta ketergantungan yang tinggi terhadap bantuan pemerintah, merupakan kriteria dari data penduduk miskin di Desa Sopet, Kecamatan Jangkar, Kabupaten Situbondo. Dengan tingkat pendidikan yang relatif rendah, penduduk miskin di Desa Sopet ini umumnya bekerja di sektor yang informal seperti buruh tani, nelayan, atau pekerja serabutan. Karena sangat terkait dengan struktur ekonomi lokal, aksesibilitas infrastruktur, dan efektivitas program-program bantuan sosial yang dijalankan oleh pemerintah.

### 3.1.1 Pengumpulan Data

Data yang akan digunakan pada penelitian ini didapatkan melalui wawancara dan observasi terhadap pemerintah Desa Sopet. Berikut merupakan tahapan yang dilakukan untuk mengumpulkan data penduduk miskin di Desa Sopet:

a. Membuat daftar observasi dan protokol wawancara yang sesuai dengan tujuan penelitian.
b. Mengajukan permohonan izin penelitian dan bekerja sama dengan pihak berwenang di Desa Sopet untuk mengatur tahap awal penelitian.

c. Menyiapkan daftar pertanyaan untuk melakukan wawancara kepada kepala desa.
d. Melakukan wawancara kepada masyarakat secara langsung untuk mempelajari lebih lanjut tentang karakteristik penduduk miskin, populasi, dan kondisi desa.
e. Mendatangi kantor desa untuk melakukan wawancara dengan kepala desa sesuai waktu yang sudah ditentukan.
f. Melakukan koordinasi dengan aparatur desa yang telah ditunjuk oleh kepala desa guna mendapatkan data penduduk miskin yang sudah tersedia di desa.
g. Menyortir data menurut kategori tertentu untuk mempermudah proses analisis.

beberapa atribut yang terdapat pada data penduduk di desa Sopet yaitu Nama, Status Hubungan Ruta , Status Perkawinan, Pendidikan Terakhir, Tanggal Lahir, Umur, pekerjaan dan Nilai, data yang telah diperoleh akan dilakukan *Data Preparation* untuk menyeleksi data teks agar lebih terstruktur.

## 3.2 Alur Proses

Dalam sebuah penelitian, Alur proses secara umum sering kali terdiri dari sejumlah langkah yang saling berkaitan, seperti identifikasi proses dan analisis proses. Kemudian membuat rencana penelitian yang mencakup peralatan, strategi pengumpulan data, dan pilihan metode. Dilanjutkan dengan mengumpulkan data di lapangan dengan menggunakan proses yang telah direncanakan sebelumnya, setelah itu data diperiksa untuk membuat kesimpulan berdasarkan hasil. Diakhiri

dengan membuat laporan penelitian yang mencakup temuan, analisis, dan saran yang dapat menjadi dasar sebagai evaluasi pada penelitian yang akan datang.

### 3.2.1 Identifikasi dan Analisis Proses Bisnis

Identifikasi terkait dengan objek penelitian yang sedang dilakukan pada bagian ini yaitu data penduduk miskin di Desa Sopet, Kecamatan Jangkar, Kabupaten Situbondo. pada tahap ini akan dilakukan beberapa proses yang terkait sehingga data siap digunakan.

a. *Data Preparation*
b. *Labeling data*
c. Implementasi metode *Naive bayes*

### 3.2.2 Analisis Proses Bisnis

Pada proses ini dilakukan analisis sebelum proses pembahasan objek dilakukan. Pada umunya proses bisnis berkaitan dengan tahapan pengerjaan terkait obejek penelitian. Dengan analisis ini akan mempermudah penelitian dalam menjabarkan proses yang diperlukan dalam menyelesaikan permasalahan objek penelitian. Berikut tahapan – tahapan analisis proses bisnis antara lain :

a. *Data Preparation*

Data preparation merupakan proses yang kompleks dan melibatkan berbagai langkah, mulai dari menangani data hilang, scaling, hingga augmentasi. Melakukan langkah-langkah ini secara manual bisa memakan waktu dan meningkatkan kemungkinan kesalahan, terutama jika data perlu diolah dalam jumlah besar atau secara berulang. Dengan

pipeline otomatisasi, seluruh proses data preparation dapat dilakukan dengan lebih konsisten, cepat, dan mudah ditelusuri[29].

b. Metode *Naive Bayes*

Berikut merupakan tahapan pengujian algoritma *Naive Bayes Clasifier* di bawah ini:

1. Menghitung probabilitas dari Vj

   Pada data *training* dengan persamaan (2.1) di mana Vj merupakan Kategori data penduduk miskin, yaitu V1 = sangat miskin, V2 = miskin, V3 = hampir miskin V4 = rentan miskin V5 = tidak miskin[30]. Yang di tentukan pada persamaan (2.1)
2. Menghitung probabilitas kata ai Pada kategori Vj dengan persamaan (2.2)
3. Model probabilitas NBC disimpan dan digunakan untuk tahap testing.
4. Menghitung probabilitas tertinggi dari kategori klasifikasi yang diujikan (Vmap) dengan persamaan (2.3)
5. Menghitung nilai maksimum yang diujikan (Vmap) pada persamaan (2.4)
6. Menghitung persamaan dari *Naive Bayes* yang diujikan (Vmap) pada persamaan (2.5)
7. Mencari nilai Vmap paling maksimum dan masukkan data penduduk tersebut pada kategori dengan Vmap maskimum.

c. *Confusion Matrix*

Berikut empat hasil perhitungan menggunakan *Confusion Matrix* untuk pengujian penduduk miskin sebagai berikut :

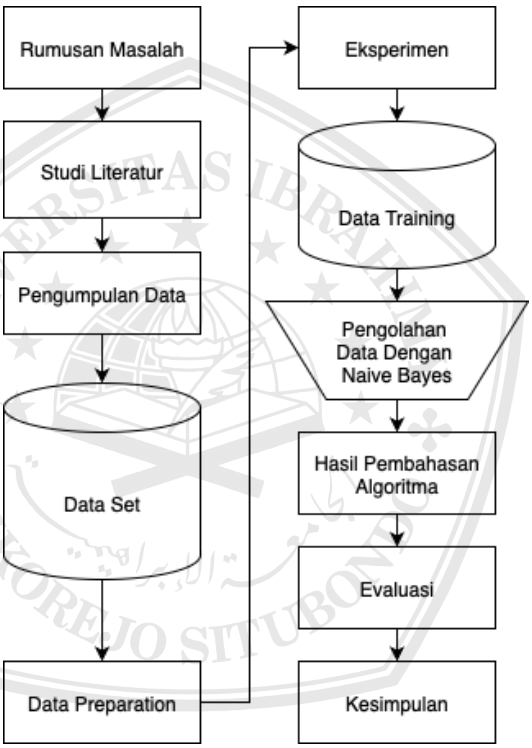
1. Menghitung nilai akurasi, atau proporsi kelas yang diprediksi secara akurat oleh model yang sudah dibuat menggunakan persamaan (2.6).
2. Menentukan nilai presisi, atau tingkat keakuratan antara informasi yang diminta pengguna dan respons sistem dengan persamaan (2.7).
3. Menentukan nilai *Recall*, sering disebut sebagai sensitivitas, yang merupakan proporsi data yang diprediksi oleh program daripada kelas yang sebenarnya dimana kelas 5 (2.8).
4. Menentukan nilai *F1-Score*, salah satu formula penilaian dalam data yang mengintegrasikan akurasi dan *recall*. Rata-rata tertimbang *recall* dan akurasi dibandingkan dalam nilai *F1-Score* (2.9).

d. Alur tahapan klasifikasi

Alur penelitian dengan metode Naïve Bayes diawali dari tahap perumusan masalah untuk menentukan arah serta fokus penelitian. Setelah itu dilakukan studi literatur sebagai dasar teori, kemudian dilanjutkan dengan proses pengumpulan data dari sumber yang sesuai. Data yang terkumpul disusun menjadi sebuah data set dan melalui tahap data preparation agar data bersih serta siap digunakan dalam analisis. Pada tahap eksperimen, data dibagi menjadi data training dan diolah menggunakan algoritma Naïve Bayes untuk proses klasifikasi. Hasil dari pengolahan data ini kemudian dibahas lebih lanjut dalam

analisis algoritma, dilanjutkan dengan evaluasi guna menilai performa model menggunakan ukuran tertentu. Tahap akhir penelitian ditutup dengan penarikan kesimpulan yang merangkum keseluruhan hasil penelitian.

Langkah-langkah yang dilakukan pada Klasifikasi data penduduk miskin ditunjukkan pada gambar 3.1 berikut ini :




**Gambar 3. 1 Tahapan Klasifikasi**

### 3.2.3 Identifikasi dan Analisis Kebutuhan

a. Identifikasi Kebutuhan Fungsional

Kebutuhan fungsional memiliki peran penting dalam sebuah penelitian karena mereka memastikan bahwa proses berjalan dengan lancar.

b. Analisis Kebutuhan Fungsional

Pada tahap ini data akan diolah menggunakan sebuah tahapan data yakni *data preparation*. *Data Preparation* merupakan tahapan awal yang dilakukan untuk memastikan data yang digunakan dalam penelitian berada dalam kondisi yang bersih, konsisten, dan siap untuk dianalisis. Tahap ini mencakup serangkaian proses yang bertujuan meningkatkan kualitas data sehingga hasil analisis dan pemodelan dapat menghasilkan output yang akurat dan dapat diandalkan. Proses data preparation pada penelitian ini meliputi beberapa langkah, antara lain pembersihan data, transformasi data, dan pengkodean data. sehingga data tidak mengalami *error* karena adanya data yang tidak memiliki nilai. Setelah proses *data preparation* selesai dilakukan proses pelabelan dalam penelitian ini ada 5 label yang digunakan yakni label dengan kategori miskin, hampir miskin, rentan miskin, sangat miskin, tidak miskin. Setelah data melewati proses pelabelan data akan dibagi menjadi dua yaitu *data training* dan *data testing*, hal ini dilakukan untuk memudahkan proses implementasi metode algoritma yang akan digunakan yakni algoritma *Naive Bayes*. Pada proses perhitungan menggunakan *algoritma Naive Bayes* akan mencari nilai probabilitas tertinggi untuk mengklasifikasi data uji pada kategori yang paling tepat.

### 3.2.4 Identifikasi dan Analisis Alternatif Solusi

#### a. Identifikasi Alternatif Solusi

Proses ini berfungsi untuk menganalisis alternatif yang digunakan dalam pembuatan program. Terdapat dua alternatif solusi dalam penelitian ini. Tabel 3.1 merupakan tabel identifikasi alternatif solusi.

**Tabel 3. 1 Identifikasi Alternatif Solusi**

| Karkteristik | Alternatif 1 (Python) | Alternatif 2 (RapidMiner) |
|---|---|---|
| Bagian sistem yang terkomputerisasi. | Semua kebutuhan fungsional akan terpenuhi | Semua kebutuhan fungsional akan terpenuhi. |
| Keuntungan, deskripsi ringkas keuntungan yang akan direalisasikan | Pengembangan lebih cepat dan mudah dilakukan bagi pemula | Proses lebih cepat bagi pemula |
| Perangkat lunak aplikasi | perangkat lunak yang akan digunakan adalah *Visual Studio Code* menggunakan bahasa *Python* | Perangkat lunak yang digunakan yang akan digunakan adalah GUI, Java API dan *data mining*. |
| Metode pemrosesan data, pada biasanya kombinasi *online, batch, deferred, remote batch dan reel time* | Editor teks yang dikembangkan oleh Microsoft untuk sistem operasi multiplatform | |
| Alat *Output* | *Monitor* dan *Printer* | *Monitor* dan *Printer* |
| Alat *Input* | *Keyboard* dan *Mouse* | *Keyboard* dan *Mouse* |
| Alat penyimpanan | Data disimpan dalam folder ruang kerja *(.vsWork)* | Data disimpan dalam tabel repositori |

**b. Analisis Kelayakan Alternatif Solusi**

Proses ini berfungsi untuk menganalisis sistem yang akan dibuat berdasarkan dokumen yang didapatkan sebelumnya. Tabel 3.2 menjelaskan identifikasi kelayakan alternatif.

**Tabel 3. 2 Analisis Kelayakan Alternatif Solusi**

| Karkteristik | Bobot | Alternatif 1 (Python) | Alternatif 2 (RapidMiner) |
|---|---|---|---|
| Kelayakan Operasional Fungsional | 40% | Mendukung seluruh kebutuhan fungsional serta pengembangan yang lebih mudah dan bisa dikerjakan oleh banyak orang | Mendukung seluruh kebutuhan fungsional, mudah dilakukan oleh pemula. |
| Skor | | 40 | 30 |
| Kelayakan teknis teknologi | 30% | Teknologi yang digunakan sudah sangat mencukupi dalam keahlian | Teknologi yang digunakan dalam manajemen sudah memadai dan sudah cukup. |
| Skor | | 25 | 30 |
| Kelayakan ekonomis biaya pengembangan | 30% | Pengembangan sistem tidak memerlukan biaya dalam pengoperasiannya | Tidak ada pengembangan dan bersifat komersial |
| Skor | | 30 | 20 |
| Total | 100% | 95 | 80 |

### 3.2.5 Desain Proses

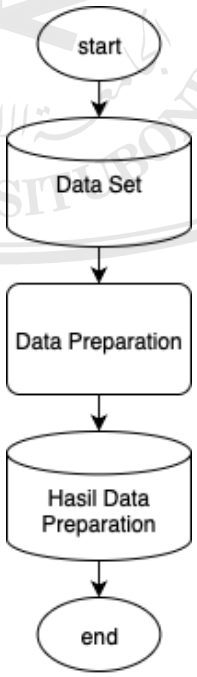
desain proses merupakan tahapan dalam merancang sebuah sistem. Tahapan ini dilakukan agar proses sistem berjalan sesuai dengan alur yang diinginkan.

a. Identifikasi Desain Proses

Pada identifikasi desain proses menjelaskan tentang alur dari analisis yang dilakukan. Bagian ini fokus pada proses yang akan dilakukan yakni proses *input* data, perhitungan dengan metode dan evaluasi hasil dari metode. Berikut adalah desain proses yang akan dilakukan.

1. *Data Preparation*

Pada tahapan ini dilakukan untuk mengolah data menjadi sebuah *dataset*. Data didapatkan dari hasil observasi dan wawancara pada aparatur desa. Data yang diperoleh berupa *file excel* lalu di *input* dan siap dilakukan proses perhitungan dengan algoritma. Pada gambar 3.2 merupakan alur proses *data Preparation*.




**Gambar 3. 2 Data Preparation**

2. *Labeling Data*

Labeling data dilakukan setelah *data preparation.* Labeling data terdiri dari :

a. Sangat miskin

Memiliki ruas tempat tinggal 5 M persesegi ,lantai tanah, dinding bambu yang kualitasnya rendah, tidak memiliki MCK, tidak teralir listrik, sumber air berasal dari sumur yang menggunakan tenaga tangan atau sungai, jenis bahan bakar menggukan kayu, tidak mengkonsumsi daging/ayam dalam seminggu, tidak pernah membeli satu stel pakain baru dalam satu tahun satu kali, makan 2x sehari, tidak sanggup membayar biaya pengobatan di puskemas atau poli klinik, penghasilan di bawah 600.000 perbulan, Pendidikan kepala keluarga tidak sekolah atau tidak tamat SD, tidak memiliki tabungan atau barang yang mudah di jual minimalnya 500.000.

b. Miskin

Memiliki ruas tempat tinggal 8 M persesegi ,lantai tanah, dinding bambu yang kualitasnya rendah, tidak memiliki MCK, tidak teralir listrik, sumber air berasal dari sumur yang menggunakan tenanga tanang, sungai, jenis bahan bakar menggukan kayu, mengkonsumsi daging/ayam Cuma seminggu dalam sehari, membeli satu stel pakain baru dalam

satu tahun satu kali, makan 2x sehari, tidak sanggup membayar biaya pengobatan di puskemas atau poli klinik, penghasilan di bawah 600.000 perbulan, Pendidikan kepala keluarga tidak sekolah atau tidak tamat SD, tidak memiliki tabungan atau barang yang mudah di jual minimalnya 500.000.

c. Hampir miskin

Memiliki ruas tempat tinggal 8 M persesegi ,lantai tehel, dinding kayu yang kualitasnya lumayan, memiliki MCK yang sederhana, teralir listrik, sumber air berasal dari sumur yang menggunakan tenanga mesin/sanyo, jenis bahan bakar menggukan elpiji, mengkonsumsi daging/ayam Cuma seminggu dalam duahari, membeli satu stel pakain baru dalam dua tahun satu kali, makan 3x sehari, Memiliki jaminan kesehatan (BPJS PBI/non-PBI), tetapi kadang menunda pengobatan karena biaya, penghasilan diatas 600.000 perbulan, Pendidikan kepala keluarga tidak sekolah atau tidak tamat SD, memiliki tabungan atau barang yang mudah di jual minimalnya 500.000.

d. Rentan miskin

Memiliki ruas tempat tinggal 15 M persesegi ,lantai kramik, dinding bata/bataku yang kualitasnya lumayan, memiliki MCK, teralir listrik, sumber air berasal dari sumur yang

menggunakan tenanga mesin sanyo, sungai, jenis bahan bakar menggukan kayu, mengkonsumsi daging/ayam dalam seminggu dalam lima hari, membeli satu stel pakain baru dalam satu tahun 10 kali, makan 3x sehari, Memiliki asuransi kesehatan atau sanggup membayar biaya pengobatan di puskemas atau poli klinik, penghasilan 1.500.000 perbulan, Pendidikan kepala keluarga sekolah atau tamat SD, memiliki tabungan atau barang yang mudah di jual minimalnya 1.000.000.

e. Tidak miskin

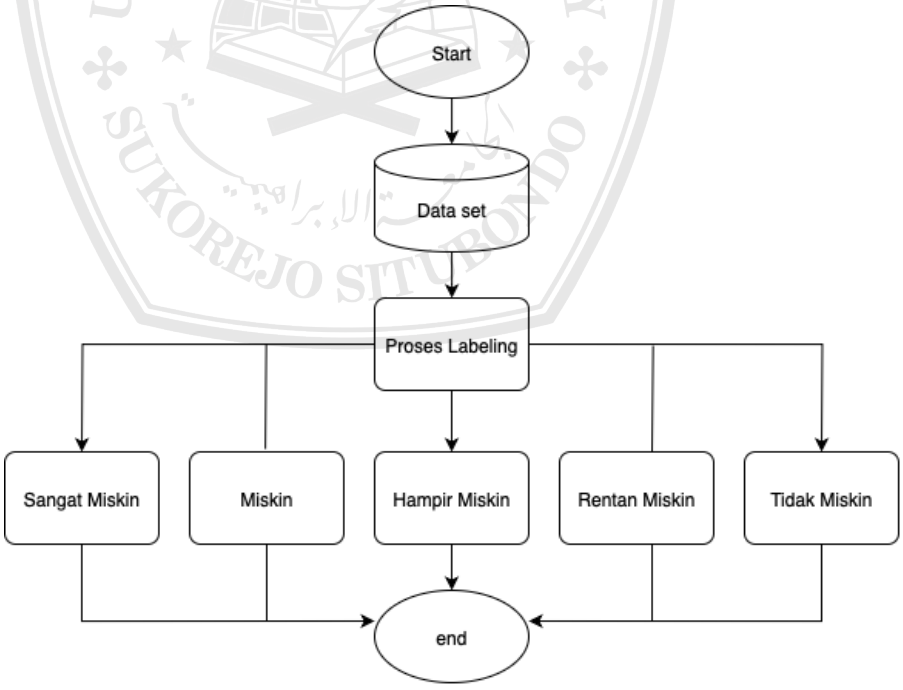
Memiliki ruas tempat tinggal 20 M persesegi ,lantai kramik, dinding batako atau bata yang kualitasnya bagus, memiliki MCK yang bagus, teralir listrik, sumber air berasal PDAM, jenis bahan bakar menggunakan elpiji khusus, mengkonsumsi daging/ayam dalam satu bulan lima hari, membeli satu stel pakain baru dalam satu tahun berkali - kali, makan 3x sehari, Menggunakan asuransi kesehatan dan mampu membayar biaya pengobatan mandiri, penghasilan 2.000.000 perbulan, Pendidikan kepala keluarga sekolah di perguruan tinggi, memiliki tabungan atau barang yang mudah di jual minimalnya diatas 1.500.000.

Tabel rentang pelabelan berikut menunjukkan klasifikasi Data Penduduk Miskin Pada Desa Saopet ke dalam lima kelompok,

yaitu Sangat Miskin, Miskin, Hampir Miskin, Rentan Miskin, dan Tidak Miskin. Pengelompokan ini digunakan sebagai acuan dalam pemberian label pada data, sehingga dapat mempermudah proses analisis serta klasifikasi pada penelitian ini. yang ditunjukkan pada tabel 3.3 dibawah ini, dan Gambar 3.3 merupakan alur proses *labeling data*.

**Tabel 3. 3 3 Range Laeling**

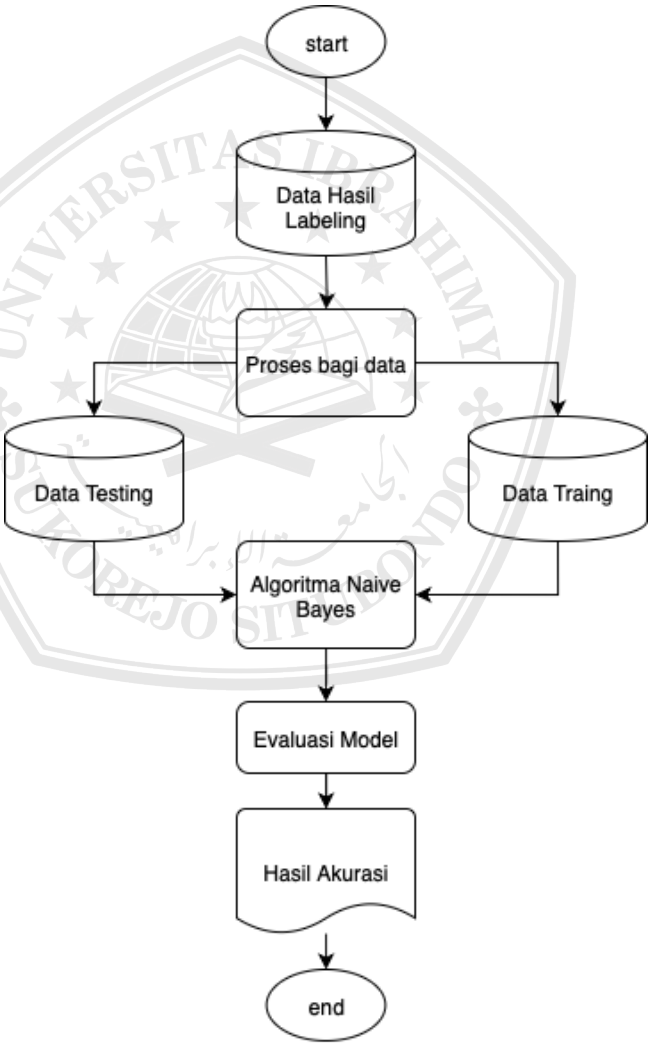
| NO | LABELING | RANGE | |
|---|---|---|---|
| 1 | SANGAT MISKIN | 1000 | 1200 |
| 2 | MISKIN | 1211 | 1410 |
| 3 | HAMPIR MISKIN | 1411 | 1510 |
| 4 | RENTAN MISKIN | 1511 | 1600 |
| 5 | TIDAK MISKIN | 1611 | 1800 |




**Gambar 3. 3 Alur *Labeling***

3. Metode *Naive Bayes*.

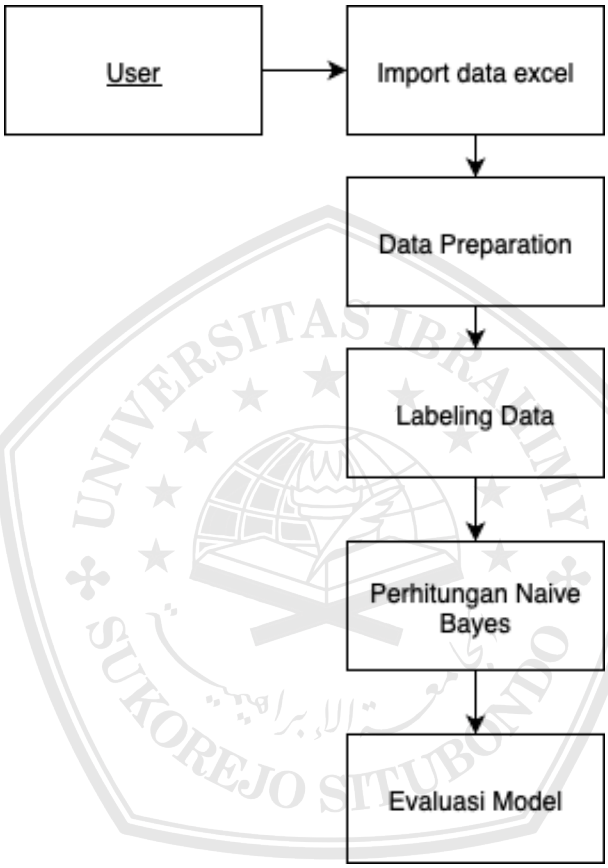
*Menginput* data hasil *labeling*, bagi menjadi *data testing* dan *data training* untuk menghitung akurasi, di *input* data *excel* dari hasil *labeling data*. dilakukan eksperimen melakukan algoritma *Naive Bayes. Output* yang dihasilkan adalah Nilai hasil akurasi. Gambar 3.4 menjelaskan alur metode *Naive Bayes.*




**Gambar 3. 4 Alur Metode Naive Bayes**

b. Arsitektur Sistem

Arsitektur sistem berfungsi sebagai penjelasan dari pola sistem analisis yang akan dilakukan, Gambar 3.5 berikut merupakan arsitektur sistem yang akan dilakukan.




**Gambar 3. 5 Arsitektur Sistem**

# BAB IV

# IMPLEMENTASI SISTEM

## 4.1 Kebutuhan Sistem

Dalam bahasa pemrograman Python sistem dapat dijalankan dengan memperhatikan beberpa perangkat yang bisa menunjang keberhasilan sistem. Berikut adalah perangkat minimal yang bisa digunakan.

a. *Software*

*Software* yang digunakan dalam mendukung sistem adalah sebagai berikut :

a) *Python*

b) *Microsoft excel*

b. *Hardware*

a) *Asus*

b) *Mouse* dan *Keyboard*

c) *Intel Core i3*

d) *SSD* dengan kapasitas 128 GB

e) RAM 4 GB

## *4.2 Data Preparation*

Dalam melakukan *Data Preparation* ada beberapa tahapan yang perlu dilakukan. Adapun tahapan *Data Preparation* adalah sebagai berikut :

a. Mengahapus data duplikat

Tahap awal dalam prosedur penghapusan data adalah menemukan baris duplikat dalam kumpulan data. Fungsi *duplicated()* dari pustaka pandas dalam

bahasa pemrograman *Python* digunakan untuk melakukan operasi ini. Setelah baris duplikat berhasil diidentifikasi, metode *drop_duplicates()* digunakan untuk menghapus data tersebut. Fungsi ini secara *default* akan menghapus semua baris duplikat sambil mempertahankan satu baris data yang dianggap sebagai representasi asli. Berikut ini disajikan potongan program untuk menghapus data duplikat:

Segmen Program menjelaskan tentang segmen program data preparation yang ada pada sistem ini untuk menghapus nama yang berduplikat. Yang ditunjukkan pada segmen 4.1

**Segmen Program 4.1 Menghapus Data Duplikat**

```
from google.colab import drive
drive.mount('/content/drive')
import pandas as pd
file_path ="/content/drive/MyDrive/DATA SOPET SS
MISKIN.xlsx"
df = pd.read_excel(file_path, na_values=['', ' ', 'NA',
'null', 'None', '-', '--'])
```

Berikut ini adalah data sebelum di hapus ada 1008 data yang ditunjukkan pada gambar 4.1 dibawah ini

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1008 entries, 0 to 1007
Data columns (total 11 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   NAMA                  1008 non-null   object
 1   STATUS HUBUNGAN RUTA  1008 non-null   object
 2   STATUS PERKAWINAN     1007 non-null   object
 3   USIA                  1008 non-null   object
 4   JENIS KELAMIN         1007 non-null   object
 5   PENDIDIKAN TERAKHIR   1007 non-null   object
 6   PEKERJAAN             1006 non-null   object
 7   PENGHASILAN/BULAN     1008 non-null   int64
 8   STATUS RUMAH          1008 non-null   object
 9   NILAI                 1003 non-null   object
 10  KATEGORI              1008 non-null   object
dtypes: int64(1), object(10)
memory usage: 86.8+ KB
```

**Gambar 4. 1 Menghapus Data Duplikat**

Berikutnya adalah data yang sudah di hapus duplikat, pada kolom nama data yang sudah di hapus tinggal 1001 karena ada 7 nama yang berduplikat. Yang ditunjukkan pada gambar 4.2 dibawah ini.

```
<class 'pandas.core.frame.DataFrame'>
Index: 1001 entries, 0 to 1007
Data columns (total 11 columns):
 #   Column                Non-Null Count  Dtype 
---  ------                --------------  ----- 
 0   NAMA                  1001 non-null   object
 1   STATUS HUBUNGAN RUTA  1001 non-null   object
 2   STATUS PERKAWINAN     1000 non-null   object
 3   USIA                  1001 non-null   object
 4   JENIS KELAMIN         1000 non-null   object
 5   PENDIDIKAN TERAKHIR   1000 non-null   object
 6   PEKERJAAN             999 non-null    object
 7   PENGHASILAN/BULAN     1001 non-null   int64 
 8   STATUS RUMAH          1001 non-null   object
 9   NILAI                 996 non-null    object
 10  KATEGORI              1001 non-null   object
dtypes: int64(1), object(10)
memory usage: 93.8+ KB
```

**Gambar 4. 2 Hasil Menghapus Data Duplikat**

b. Mengahaspus baris yang kosong

Untuk menjamin integritas dan kualitas data, sangat penting untuk menangani baris dengan nilai kosong selama tahap pra-pemrosesan data. Nilai yang hilang atau baris kosong dapat menyebabkan masalah dalam proses analisis data dan mengancam validitas model yang sedang dikembangkan. Untuk mengatasi masalah ini, menghapus baris yang bermasalah merupakan solusi yang umum digunakan. Metode *dropna()* yang disediakan oleh paket pandas dalam bahasa pemrograman *Python* dapat digunakan untuk melaksanakan prosedur ini. Fungsi ini menghapus baris yang memiliki satu atau lebih nilai kosong, sehingga menghasilkan dataset yang lebih bersih dan siap untuk analisis lebih lanjut. Berikut ini merupakan potongan kode program yang digunakan untuk menghapus baris yang kosong dalam dataset:

Segmen Program menjelaskan tentang segmen program *data preparation* yang ada pada sistem ini untuk Mengahaspus data baris yang kosong. Yang ditunjukkan pada segmen 4.2

**Segmen Program 4.2 Menghapus Baris Yang Kosong**

```
#menghapus baris yang kosong
df = df.dropna(how='any')
df_hilangkan.to_csv('SS PARATION MISKIN .csv',
index=False)
```

Sebelum dilakukan penghapusan baris yang kosong data terdiri dari 1001. Berikut data sebelum di hapus baris yang kosong pada gambar 4.3 dibawah ini.

```
<class 'pandas.core.frame.DataFrame'>
Index: 1001 entries, 0 to 1007
Data columns (total 11 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   NAMA                  1001 non-null   object
 1   STATUS HUBUNGAN RUTA  1001 non-null   object
 2   STATUS PERKAWINAN     1000 non-null   object
 3   USIA                  1001 non-null   object
 4   JENIS KELAMIN         1000 non-null   object
 5   PENDIDIKAN TERAKHIR   1000 non-null   object
 6   PEKERJAAN             999 non-null    object
 7   PENGHASILAN/BULAN     1001 non-null   int64
 8   STATUS RUMAH          1001 non-null   object
 9   NILAI                 996 non-null    object
 10  KATEGORI              1001 non-null   object
dtypes: int64(1), object(10)
memory usage: 93.8+ KB
```

**Gambar 4. 3 Sebelum Menghapus Baris Yang Hilang**

Setelah dilakukan penghapusan baris yang kosong data menjadi 996 baris. Berikut hasil dari penghapusan baris yang kosong pada gambar 4.4.

```
<class 'pandas.core.frame.DataFrame'>
Index: 996 entries, 0 to 1007
Data columns (total 11 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   NAMA                  996 non-null    object
 1   STATUS HUBUNGAN RUTA  996 non-null    object
 2   STATUS PERKAWINAN     996 non-null    object
 3   USIA                  996 non-null    object
 4   JENIS KELAMIN         996 non-null    object
 5   PENDIDIKAN TERAKHIR   996 non-null    object
 6   PEKERJAAN             996 non-null    object
 7   PENGHASILAN/BULAN     996 non-null    int64
 8   STATUS RUMAH          996 non-null    object
 9   NILAI                 996 non-null    object
 10  KATEGORI              996 non-null    object
dtypes: int64(1), object(10)
memory usage: 93.4+ KB
```

**Gambar 4. 4 Hasil Menghapus Baris Yang Hilang**

## 4.3 Labeling Data

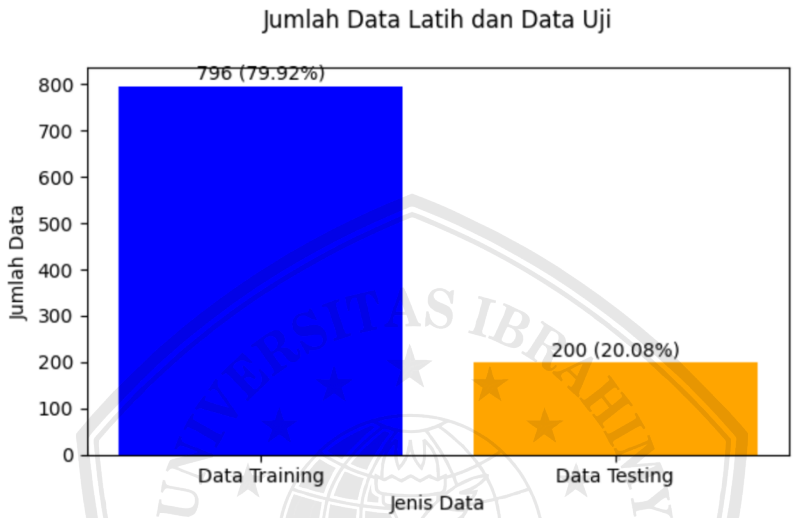
Proses labeling data dilakukan secara manual. Proses ini dilakukan secara teliti dengan mengamati setiap data satu per satu, lalu menetapkan label seperti "Sangat Miskin, Miskin, Hampir Miskin, Rentan Miskin dan Tidak Miskin" sesuai dengan kriteria yang telah ditentukan. Berikut ini data yang sudah di lebeling data manual, berikut tabel 4.1 dibawah ini.

**Tabel 4.1 *Labeling* Data**

| NAMA | STATUS HUBUNGAN RUTA | STATUS PERKAWINAN | USIA | JENIS KELAMIN | PENDIDIKAN TERAKHIR | PEKERJAAN | PENGHASILAN /BULAN | STATUS RUMAH | NILAI | KATEGORI |
|---|---|---|---|---|---|---|---|---|---|---|
| NIWETI | KEPALA RUTA | CERAI HIDUP | 70 Tahun | PEREMPUAN | TIDAK SEDANG SEKOLAH | PETANI/PERKEBUNAN | 300000 | MILIK SENDIRI | 1251,87 | MISKIN |
| MASRIJO | KEPALA RUTA | CERAI MATI | 70 Tahun | LAKI - LAKI | TIDAK SEDANG SEKOLAH | PETANI/PERKEBUNAN | 350000 | MILIK SENDIRI | 1253,67 | MISKIN |
| NURSIA | KEPALA RUTA | CERAI MATI | 75 Tahun | PEREMPUAN | TIDAK SEDANG SEKOLAH | PETANI/PERKEBUNAN | 2500000 | MILIK SENDIRI | 1629,79 | TIDAK MISKIN |
| SUHANI | KEPALA RUTA | CERAI MATI | 45 Tahun | PEREMPUAN | TIDAK SEDANG SEKOLAH | PETANI/PERKEBUNAN | 250000 | NUMPANG | 1099,53 | SANGAT MISKIN |
| NIWATI | KEPALA RUTA | CERAI HIDUP | 54 Tahun | PEREMPUAN | TIDAK SEDANG SEKOLAH | PETANI/PERKEBUNAN | 350000 | MILIK SENDIRI | 1279,99 | MISKIN |
| MAWATI | ANGGOTA RUTA | KAWIN | 41 Tahun | PEREMPUAN | TIDAK SEDANG SEKOLAH | PETANI/PERKEBUNAN | 650000 | MILIK SENDIRI | 1481,21 | HAMPIR MISKIN |
| USNADI | KEPALA RUTA | KAWIN | 35 Tahun | LAKI - LAKI | TIDAK SEDANG SEKOLAH | PETANI/PERKEBUNAN | 1550000 | MILIK SENDIRI | 1591,11 | RENTAN MISKIN |
| MIHATI | KEPALA RUTA | CERAI MATI | 72 Tahun | PEREMPUAN | TIDAK SEDANG SEKOLAH | MENGURUS RUMAH TANGGA | 300000 | MILIK SENDIRI | 1249,58 | MISKIN |

## 4.1 Split Data Training dan Data Testing

Pada proses klasifikasi ini, dataset dibagi menjadi dua data yaitu data training dan data testing. karena proses klasifikasi membutuhkan data training dan testing, Berikut gambar 4.5 dibawah ini.




**Gambar 4. 5 Diagram Batang Split Data.**

Pada proses split data, data Penduduk Miskin dibagi agar model dapat dilatih menggunakan sebagian data, dan diuji performanya. Data Penduduk Miskin yang di gunakan, sebanyak 796 data (79.92%) digunakan sebagai data latih, dan 200 data (20,08%) digunakan sebagai data uji. Yang ditunjukkan pada segmen 4.3

**Segmen Program 4.3 Split Data**

```
1   import matplotlib.pyplot as plt
2   # Contoh nilai (ganti dengan variabel aslimu)
3   train_size = len(x_train)
4   test_size = len(x_test)
5   # Membuat plot
6   plt.figure(figsize=(6, 4))
7   bars = plt.bar(['Data Training', 'Data Testing'],
8   [train_size, test_size], color=['blue', 'orange'])
9   # Menambahkan label untuk setiap bar (jumlah +
10  persentase)
11  for bar in bars:
12      height = bar.get_height()
```

*Segmen Program 4.3 (Lanjutan)*

```
    total = train_size + test_size
    percent = (height / total) * 100
    plt.text(bar.get_x() + bar.get_width() / 2, height +
5, f'{height} ({percent:.2f}%)',
             ha='center', va='bottom')
plt.title('Jumlah Data Latih dan Data Uji\n')
plt.xlabel('Jenis Data')
plt.ylabel('Jumlah Data')
plt.tight_layout()
plt.show()
```

## 4.4 Metode *Naive Bayes*

Pada tahap ini untuk menerjemahkan teks menjadi representasi numerik berdasarkan frekuensi kemunculan kata (token), data pada kolom VALUE kini diproses menggunakan pendekatan *CountVectorizer*. Untuk mengelompokkan data teks sesuai dengan kelasnya sangat miskin, miskin, hampir miskin, rentan miskin, dan tidak miskin setiap kata yang muncul dalam data diubah menjadi fitur dan kemudian digabungkan dengan kolom kategori. Distribusi istilah dalam setiap kategori dapat dianalisis lebih lanjut menggunakan kombinasi ini.

Selanjutnya dilakukan Probabilitas prior, atau persentase data dalam setiap kategori terhadap total data, kemudian ditentukan setelah frekuensi token untuk setiap kategori telah ditentukan (2.1). Algoritma *Naive Bayes* menggunakan probabilitas ini sebagai titik awal untuk memprediksi kategori baru berdasarkan data masukan. Fase ini berfungsi sebagai persyaratan penting dalam proses

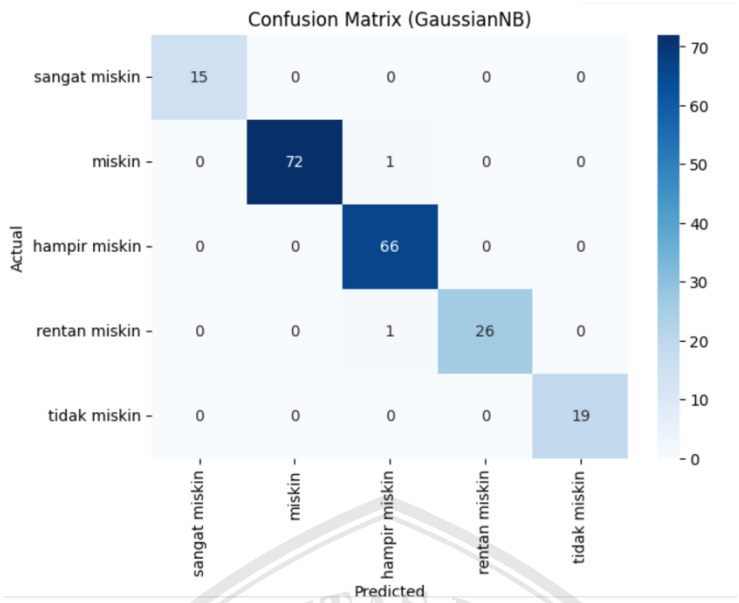
kategorisasi (2.2). berikut hasil aktual dan prediksi pada gaussianNB pada gambar 4.6.

| | PENGHASILAN/BULAN | STATUS RUMAH | NILAI | Actual | Predicted |
|---|---|---|---|---|---|
| 832 | 400000 | MILIK SENDIRI | 1377,56 | miskin | miskin |
| 970 | 400000 | MILIK SENDIRI | 1388,11 | miskin | miskin |
| 96 | 2700000 | MILIK SENDIRI | 1659,69 | tidak miskin | tidak miskin |
| 587 | 750000 | MILIK SENDIRI | 1432,71 | hampir miskin | hampir miskin |
| 450 | 1600000 | MILIK SENDIRI | 1598,31 | rentan miskin | rentan miskin |
| 266 | 400000 | MILIK SENDIRI | 1316,18 | miskin | miskin |
| 290 | 750000 | MILIK SENDIRI | 1471,27 | hampir miskin | hampir miskin |
| 158 | 3000000 | MILIK SENDIRI | 1614,99 | tidak miskin | tidak miskin |
| 668 | 1550000 | MILIK SENDIRI | 1546,6 | rentan miskin | rentan miskin |
| 572 | 750000 | MILIK SENDIRI | 1453,82 | hampir miskin | hampir miskin |

**Gambar 4. 6 Hasil Prediksi Pada Data Klasifikasi**

Selanjutnya adalah mengukur ketepatan klasifikasi pada data training dan data testing dari setiap klasifikasi. Pengukuran ini dilakukan menggunakan confusion matrik berdasarkan hasil dari klasifikasi,

Metode *Gaussian Naive Bayes* pada data dengan lima kelas, *confusion matrik* hasil klasifikasi ditampilkan pada gambar 4.7. Dengan 5, 114, 54, 11, dan 3. titik data yang diidentifikasi dengan benar, masing-masing, temuan ini menunjukkan kinerja model yang kuat, terutama pada kelas 0 dan 1. Namun, masih ada beberapa kesalahan prediksi, termasuk lima titik data kelas 1 yang salah diklasifikasikan sebagai kelas 0 dan beberapa kesalahan kecil di kelas 5, 4, 1 dan 1. Kelas 4 memiliki tiga prediksi yang akurat, sedangkan kelas 2 dan 3 masing-masing memiliki sebelas dan lima prediksi yang benar. Secara keseluruhan, model ini melakukan pekerjaan yang baik dalam mengklasifikasikan data, terutama pada kelas-kelas.




**Gambar 4. 7 Confusion Matrix**

```
Classification Report (GaussianNB):
              precision    recall  f1-score   support

sangat miskin       1.00      1.00      1.00        15
      miskin        1.00      0.99      0.99        73
hampir miskin       0.97      1.00      0.99        66
rentan miskin       1.00      0.96      0.98        27
 tidak miskin       1.00      1.00      1.00        19

     accuracy                           0.99       200
    macro avg       0.99      0.99      0.99       200
 weighted avg       0.99      0.99      0.99       200

========================================
Accuracy (GaussianNB): 0.9900
========================================
```

**Gambar 4. 8 *Classification Results***

Tahap selanjutnya merupakan perhitungan manual untuk menentukan akurasi model klasifikasi sentimen. Jumlah prediksi yang akurat dibandingkan dengan jumlah total data testing untuk menentukan akurasi. Pada hasil *confusion matrix* menunjukkan jumlah prediksi yang benar sangat miskin 5, miskin 114, hampir miskin 54, rentan miskin 11 dan tidak miskin 3 dengan total 187 data yang dikategorikan dengan akurat. Ada 85 data uji secara keseluruhan. Oleh karena itu,

rumus (**2.6**) digunakan untuk menentukan akurasi model, menghasilkan nilai akurasi 0, 9396 atau 93,96%.

$$\text{Akurasi} = \frac{5+114+54+11+3}{5+1+114+5+1+4+54+11+1+3} x100\% =$$

$$\text{Akurasi} = \frac{187}{199} x100\% =$$

$$\text{Akurasi} = \frac{187}{199} x100\% = 0.9396$$

Berikut merupakan perhitungan manual *classification results* seperti *precision, recall,* dan *fl-score* pada tiap kelas :

a. Sangat Miskin

Pada kategori Sangat Miskin, hasil *confusion matrix* memperlihatkan bahwa jumlah data yang berhasil diprediksi dengan benar sebagai Sangat Miskin (*True* Positive/TP) adalah sebanyak 5 data. Namun demikian, terdapat 4 data dari kelas lain yang secara keliru diklasifikasikan sebagai Sangat Miskin (*False Positive*/FP). Selain itu, terdapat pula 1 data Sangat Miskin yang tidak berhasil diidentifikasi dengan tepat oleh model (*False Negative*/FN). Berdasarkan data tersebut, nilai Precision dihitung dengan menggunakan rumus (2.7):

$$\text{Precission} = \frac{5}{5+4} = 0.5556$$

Nilai *precision* sebesar 0,5556 menunjukkan bahwa hanya sekitar 55,56% dari keseluruhan data yang diprediksi sebagai Sangat Miskin benar-benar sesuai dengan kondisi sebenarnya. Dengan kata lain, masih terdapat cukup banyak kesalahan dalam memprediksi data dari kelas lain sebagai Sangat Miskin. Selanjutnya, nilai Recall diperoleh dengan rumus(2.8):

$$\text{Recall} = \frac{5}{5+1} = 0.8333$$

*Recall* sebesar 0,8333 (83,33%) menandakan bahwa sebagian besar data yang sebenarnya termasuk kategori Sangat Miskin berhasil dikenali oleh model, meskipun masih ada sebagian kecil data yang tidak teridentifikasi. Untuk memberikan gambaran kinerja yang lebih seimbang antara precision dan recall, dihitung nilai *F1-Score* dengan rumus rata-rata harmonis (2.9)sebagai berikut:

$$\text{F1} = 2\ \chi\ \frac{0.5556x0.8333}{0.5556+0.8333} = 0.6667$$

Nilai *F1-Score* sebesar 0,6667 mengindikasikan bahwa meskipun model cukup baik dalam mengenali data kategori Sangat Miskin (tinggi *recall*), namun ketepatannya dalam membedakan data ini dari kelas lain masih belum optimal (*precision* rendah). Hal ini berarti model masih sering salah mengklasifikasikan data dari kelas lain ke dalam kategori Sangat Miskin, sehingga perlu adanya peningkatan akurasi pada aspek precision agar hasil klasifikasi lebih seimbang dan reliabel.

b. Miskin

Pada kategori Miskin, hasil pengujian melalui confusion matrix menunjukkan bahwa jumlah data yang berhasil diprediksi dengan benar sebagai Miskin (*True Positive*/TP) adalah sebanyak 114 data. Selain itu, terdapat 1 data dari kelas lain yang secara keliru terklasifikasi sebagai Miskin (*False Positive*/FP). Adapun jumlah data yang seharusnya termasuk ke dalam kategori Miskin namun tidak berhasil dikenali oleh model, atau disebut False Negative (FN), 5+1=6 jadi berjumlah 6 data. Dari hasil tersebut, nilai Precision untuk kategori Miskin dapat dihitung dengan menggunakan formula:

$$\text{Precission} = \frac{114}{114+1} = 0.9913$$

Nilai *precision* yang sangat tinggi, yakni 0,9913 (99,13%), menunjukkan bahwa hampir seluruh data yang diprediksi sebagai Miskin benar-benar sesuai dengan kondisi aslinya. Hal ini mengindikasikan bahwa model memiliki tingkat ketepatan yang sangat baik dalam meminimalkan kesalahan klasifikasi dari kelas lain ke dalam kategori Miskin. Selanjutnya, nilai Recall dihitung melalui rumus:

$$\text{Recall} = \frac{114}{114 + 6} = 0.95$$

Nilai *recall* sebesar 0,95 (95%) menunjukkan bahwa sebagian besar data yang sebenarnya termasuk ke dalam kelas *Miskin* berhasil diidentifikasi dengan benar oleh model. Meskipun demikian, masih ada sejumlah kecil data (6 data) yang tidak berhasil dikenali sebagai Miskin.

Untuk memperoleh ukuran performa yang lebih seimbang antara precision dan recall, dilakukan perhitungan *F1-Score* dengan menggunakan rata-rata harmonis dari kedua nilai tersebut. Perhitungannya sebagai berikut:

$$\text{F1} = 2\ \chi\ \frac{0.9913x0.95}{0.9913+0.95} = 0.9702$$

Nilai *F1-Score* sebesar 0,9702 menegaskan bahwa model memiliki kinerja yang sangat baik dalam mengklasifikasikan data pada kategori Miskin. Tingginya nilai precision dan recall yang saling mendekati menunjukkan bahwa model tidak hanya mampu mengenali sebagian besar data Miskin dengan baik, tetapi juga sangat jarang melakukan kesalahan prediksi terhadap kelas lain. Dengan demikian, dapat disimpulkan bahwa untuk kategori ini, performa model berada pada tingkat yang optimal dan dapat diandalkan.

c. Hampir Miskin

Pada kategori Hampir Miskin, hasil confusion matrix memperlihatkan bahwa jumlah data yang berhasil diprediksi dengan benar sebagai Hampir Miskin (*True Positive*/TP) adalah sebanyak 54 data. Namun, masih terdapat 5 data dari kelas lain yang secara keliru terklasifikasi ke dalam kategori ini (*False Positive*/FP). Selain itu, ada 4 data Hampir Miskin yang gagal dikenali oleh model dan justru terprediksi sebagai kelas lain (*False Negative*/FN). Berdasarkan data tersebut, nilai Precision dapat dihitung sebagai berikut:

$$\text{Precission} = \frac{54}{54 + 5} = 0.9153$$

Hasil ini menunjukkan bahwa sekitar 91,53% dari data yang diprediksi sebagai Hampir Miskin memang benar-benar sesuai dengan label aslinya. Dengan kata lain, model cukup baik dalam menekan kesalahan prediksi dari kelas lain agar tidak salah dikategorikan ke dalam kelas ini. Selanjutnya, nilai *Recall* dihitung dengan formula:

$$\text{Recall} = \frac{54}{54 + 4} = 0.9310$$

Nilai *recall* sebesar 0,9310 (93,10%) mengindikasikan bahwa sebagian besar data yang sebenarnya termasuk dalam kategori Hampir Miskin dapat dikenali oleh model dengan baik, meskipun masih ada 4 data yang luput terdeteksi. Untuk memperoleh ukuran kinerja yang lebih seimbang antara precision dan *recall*, digunakan nilai *F1-Score* yang dihitung sebagai berikut:

$$\text{F1} = 2\,\chi\,\frac{0.9153x0.9310}{0.9153+0.9310} = 0.9231$$

Nilai *F1-Score* sebesar 0,9231 menunjukkan bahwa model memiliki performa yang cukup stabil dan seimbang dalam mengklasifikasikan kategori Hampir Miskin. Tingginya nilai *precision* dan *recall* yang saling mendekati memperlihatkan bahwa model mampu menjaga keseimbangan antara ketepatan dalam prediksi dan kelengkapan dalam mengenali data sebenarnya. Secara keseluruhan, kinerja model pada kategori ini dapat dikatakan baik, meskipun masih terdapat peluang peningkatan agar kesalahan klasifikasi semakin berkurang.

d. Rentan Miskin

Pada kategori Rentan Miskin, hasil evaluasi menggunakan confusion matrix menunjukkan bahwa nilai *True Positive* (TP) yang diperoleh adalah sebesar 11. Artinya, terdapat 11 data yang berhasil diprediksi dengan benar sebagai kelompok rentan miskin. Sementara itu, jumlah *False Positive* (FP) bernilai 0, yang menunjukkan bahwa tidak ada data dari kelas lain yang salah teridentifikasi sebagai rentan miskin. Di sisi lain, terdapat satu data yang masuk dalam *False Negative* (FN = 1), yang berarti ada satu kasus rentan miskin yang tidak berhasil terdeteksi dengan benar oleh model.

$$\text{Precission} = \frac{11}{11 + 0} = 1.0$$

Berdasarkan data tersebut, nilai Precision dihitung dengan rumus TP/(TP+FP). Dengan TP = 11 dan FP = 0, maka diperoleh nilai *Precision* sebesar 11/(11+0) = 1.0. Nilai ini menunjukkan bahwa seluruh prediksi yang diberikan model untuk kelas rentan miskin benar adanya, atau dengan kata lain tidak terdapat kesalahan prediksi positif.

$$\text{Recall} = \frac{11}{11+1} = 0.9167$$

Nilai *Recall* dihitung dengan rumus TP/(TP+FN). Dari perhitungan diperoleh 11/(11+1) = 0,9167. Nilai *recall* ini menandakan bahwa model mampu mengenali sekitar 91,67% dari total data aktual yang termasuk ke dalam kategori rentan miskin, meskipun masih ada sebagian kecil data yang terlewatkan.

$$\text{F1} = 2\,\chi\,\frac{1.0x0.9167}{1.0+0.9167} = 0.9565$$

Untuk memberikan ukuran kinerja yang lebih seimbang antara *Precision* dan *Recall*, dilakukan penghitungan *F1-Score* dengan menggunakan rumus 2 × (*Precision* × *Recall*) / (*Precision* + *Recall*). Dari hasil substitusi nilai Precision = 1.0 dan Recall = 0,9167, diperoleh nilai *F1-Score* sebesar 0,9565. Hal ini mengindikasikan bahwa performa model dalam mengklasifikasikan kategori rentan miskin tergolong sangat baik, karena nilai *F1-Score* mendekati angka sempurna (1.0)

e. Tidak Miskin

Pada kategori Tidak Miskin, hasil pengujian model yang ditampilkan dalam *confusion matrix* menunjukkan bahwa jumlah data yang berhasil diprediksi benar sebagai Tidak Miskin (*True Positive*/TP) adalah sebanyak 3 data. Sementara itu, terdapat kesalahan prediksi di mana data dari kelas lain ikut terklasifikasi sebagai Tidak Miskin, yaitu 1+1=2 jadi data sebanyak 2 (*False Positive*/FP). Adapun kesalahan prediksi berupa data Tidak Miskin yang tidak teridentifikasi dengan benar (*False Negative*/FN) tercatat tidak ada atau bernilai 0.

Berdasarkan hasil tersebut, nilai Precision untuk kategori Tidak Miskin dihitung menggunakan rumus:

$$\text{Precission} = \frac{3}{3+2} = 0.6$$

Artinya, dari seluruh data yang diprediksi sebagai Tidak Miskin, hanya 60% yang benar-benar sesuai dengan label aslinya. Selanjutnya, perhitungan *Recall* diperoleh dengan rumus:

$$\text{Recall} = \frac{3}{3+0} = 1.0$$

Nilai *recall* sebesar 1,0 (100%) menunjukkan bahwa seluruh data yang sebenarnya termasuk dalam kelas Tidak Miskin berhasil dikenali dengan benar oleh model. Dengan kata lain, model tidak melewatkan satu pun data Tidak Miskin. Untuk mendapatkan ukuran kinerja yang lebih seimbang antara precision dan recall, digunakan perhitungan *F1-Score*, yang merupakan rata-rata harmonis dari kedua nilai tersebut. Perhitungan *F1-Score* ditunjukkan sebagai berikut:

$$\text{F1} = 2\ \chi\ \frac{0.6x1.0}{0.6+1.0} = 0.75$$

Nilai *F1-Score* sebesar 0,75 memberikan gambaran bahwa meskipun recall menunjukkan hasil sempurna, nilai *precision* yang relatif rendah membuat akurasi keseluruhan pada kategori ini tidak maksimal. Secara umum, hasil ini mengindikasikan bahwa model lebih baik dalam mengenali semua data Tidak Miskin (tinggi *recall*), tetapi masih kurang tepat dalam menghindari salah klasifikasi dari kelas lain sebagai Tidak Miskin. Yang ditunjukkan pada segmen 4.4

**Segmen Program 4.4 Program *Naive Bayes***

```
1  import pandas as pd
2  import numpy as np
3  from sklearn.model_selection import train_test_split
4  from sklearn.naive_bayes import GaussianNB
```

*Segmen Program 4.4 (Lanjutan)*

```
from sklearn.metrics import confusion_matrix,
classification_report, accuracy_score
from sklearn.feature_extraction.text import
TfidfVectorizer
import seaborn as sns
import matplotlib.pyplot as plt
tfidf = TfidfVectorizer()
# menggabungkan ketiga kolom menjadi satu string per
baris
text_data = df_data[['PENGHASILAN/BULAN', 'STATUS
RUMAH', 'NILAI']].astype(str).agg(' '.join, axis=1)
x = tfidf.fit_transform(text_data).toarray()
y = df_data['KATEGORI']
# Split data
X_train, X_test, y_train, y_test = train_test_split(x,
y, test_size=0.2, random_state=42)
# Initialize and train model GaussianNB
gnb = GaussianNB()
gnb.fit(X_train, y_train)
# Predict with GaussianNB
y_pred_gnb = gnb.predict(X_test)
# Tentukan urutan label secara manual
label_order = ["sangat miskin", "miskin", "hampir
miskin", "rentan miskin", "tidak miskin"]
# Evaluate GaussianNB dengan urutan label tertentu
conf_matrix_gnb = confusion_matrix(y_test, y_pred_gnb,
labels=label_order)
class_report_gnb = classification_report(y_test,
y_pred_gnb, labels=label_order,
target_names=label_order)
accuracy_gnb = accuracy_score(y_test, y_pred_gnb)
print("GaussianNB Results")
print("=====================================")
print("Confusion Matrix (GaussianNB):")
print(conf_matrix_gnb)
print("=====================================")
print("\nClassification Report (GaussianNB):")
print(class_report_gnb)
print("=====================================")
print(f"Accuracy (GaussianNB): {accuracy_gnb:.4f}")
print("=====================================")
# Plot confusion matrix untuk GaussianNB dengan label
yang jelas
plt.figure(figsize=(7, 5))
sns.heatmap(conf_matrix_gnb, annot=True, fmt='d',
cmap='Blues',
            xticklabels=label_order,
yticklabels=label_order)
plt.title('Confusion Matrix (GaussianNB)')
plt.xlabel('Predicted')
plt.ylabel('Actual')
plt.show()
```

*Segmen Program 4.4 (Lanjutan)*

```
57  # Simpan hasil prediksi ke CSV
58  results_gnb = df_data.loc[y_test.index,
59  ['PENGHASILAN/BULAN', 'STATUS RUMAH', 'NILAI']].copy()
60  results_gnb['Actual'] = y_test.values
61  results_gnb['Predicted'] = y_pred_gnb
62  results_gnb.to_csv('Hasil_pred_GaussianNB.csv',
63  encoding='utf8', index=False)
64  results_gnb.head(10)
```

# BAB V

# KESIMPULAN

## 5.1 KESIMPULAN

Berdasarkan hasil pengujian klasifikasi data penduduk miskin pada Desa Sopet menggunakan *Algoritma Naive Bayes* menghasilkan akurasi 93,97%. Dengan nilai *Precision* pada kategori hampir miskin 92%, miskin 99%, rentan miskin 100%, sangat miskin 56%, tidak miskin 60%. Nilai *Recall* pada kategori penduduk miskin hampir miskin 93%, miskin 95%, rentan miskin 92%, sangat miskin 83%, tidak miskin 100%. sedangkan nilai *F1 Score* pada kategori hampir miskin 92%, miskin 97%, rentan miskin 96%, sangat miskin 67%, tidak miskin 75%. Dimana hasil perhitungan tersebut berdasarkan beberapa variabel, yaitu Nama, Status Hubungan Ruta , Status Perkawinan, Pendidikan Terakhir, Tanggal Lahir, Umur, pekerjaan, nilai, dan penghasilan.

Hasil tingkat akurasi, *Precision, Recall, dan F1 Score* menunjukkan kinerja yang sangat baik dari *Algoritma Naïve Bayes Clasifier,* terutama pada kelas miskin dan hampir miskin. Hasil ini menunjukkan bahwa metode *Naive Bayes Clasifier* dapat menjadi alat yang bermanfaat untuk memfasilitasi pengambilan keputusan yang lebih terfokus dalam mengalokasikan bantuan sosial.

## 5.2 SARAN

Untuk memberikan hasil klasifikasi yang lebih mendalam, disarankan agar metode klasifikasi penduduk miskin ini bisa diperluas untuk penelitian selanjutnya dengan memasukkan variabel-variabel penentu penduduk miskin tambahan, seperti partisipasi sosial, kepemilikan aset produktif, dan pertimbangan kesehatan. Dan untuk membandingkan kinerjanya algoritma *Naive Bayes* disarankan juga agar

menggunakan algoritma lain seperti *Random Forest* dan *Support Vector Machine* juga dapat digunakan. Sehingga dengan mengintegrasikan sistem informasi desa untuk menyediakan pengumpulan data otomatis dan pembaruan informasi secara berkelanjutan, Keputusan mengenai bantuan sosial dapat diambil dengan lebih cepat, akurat, dan efisien jika penelitian selanjutnya memanfaatkan data *real-time*.

## Daftar Pustaka


[1] S. Gabriel Christian1, *IMPLEMENTASI ALGORITMA NAÏVE BAYES CLASSIFIER PADA KLASIFIKASI PENDUDUK MISKIN (STUDI KASUS: DESA TEMBUNG)*. 2023. [Online]. Available: https://publisher.unimed.ac.id

[2] E. Purwanti, "Analisis Deskriptif Profil Kemiskinan Indonesia Berdasarkan Data BPS Tahun 2023," *Jurnal Mahasiswa Humanis*, vol. 4, no. 1, pp. 1–10, 2024.

[3] *Mengenal Microsoft Office*. Elex Media Komputindo, 2024.

[4] Muhammad Ridho, "Penerapan Teknologi Informasi untuk Mendorong Kemandirian Desa di Era Digital," *Merkurius : Jurnal Riset Sistem Informasi dan Teknik Informatika*, vol. 2, no. 6, pp. 150–158, Oct. 2024, doi: 10.61132/merkurius.v2i6.450.

[5] F. Marisa, S. Risnanto, R. Hardi, B. Pudjoatmojo, E. T. E. Handayani, and A. L. Maukar, *Algoritma Populer Dalam Intelligent Sistem Beserta Contoh Kasus*. Deepublish, 2023. [Online]. Available: https://books.google.co.id/books?id=G-NMEQAAQBAJ

[6] N. Madia, A. Septiarini, H. Rahmania Hatta, H. Hamdani, and M. Wati, "Penentuan Kelayakan Masyarakat Miskin Penerima Bantuan Menggunakan Metode Naïve Bayes (Studi Kasus: Kabupaten Penajam Paser Utara)," 2023.

[7] Florens Dianni Nurhabiba, "KEMAMPUAN HIGHER ORDER THINKING SKILL (HOTS) DALAM PEMBELAJARAN BERDIFERENSIASI SD 19 PALEMBANG," vol. 09, pp. 492–504, 2023.

[8] M. Farhan Arib, M. Suci Rahayu, R. A. Sidorj, and M. Win Afgani, "Experimental Research Dalam Penelitian Pendidikan," *Journal Of Social Science Research*, vol. 4, pp. 5497–5511, 2024.

[9] M. M. Agung Mahardini, "Analisis Situasi Penggunaan Google Classroom pada Pembelajaran Daring Fisika," *Jurnal Pendidikan Fisika*, vol. 8, no. 2, p. 215, Sep. 2020, doi: 10.24127/jpf.v8i2.3102.

[10] P. A. W. Annisa Rizky Fadilla, "LITERATURE REVIEW ANALISIS DATA KUALITATIF: TAHAP PENGUMPULAN DATA," *MITITA JURNAL PENELITIA*, vol. 1, pp. 34–46, 2023.

[11] H. Hasan Sistem Informasi and S. Tidore Mandiri, "PENGEMBANGAN SISTEM INFORMASI DOKUMENTASI TERPUSAT PADA STMIK TIDORE MANDIRI," 2022. [Online]. Available: http://www.php.net

[12] Mahasiswa PGSD C, *ANEKA INOVASI PEMBELAJARAN DARI STUDI KEPUSTAKAAN*. Uwais Inspirasi Indonesia, 2024.

[13] H. Harliana and F. N. Putra, "Klasifikasi Tingkat Rumah Tangga Miskin Saat Pandemi Dengan Naïve Bayes Classifier," *Jurnal Sains dan Informatika*, vol. 7, no. 2, pp. 165–173, Dec. 2021, doi: 10.34128/jsi.v7i2.339.

[14] J. Mulya Kecamatan Sepatan Timur *et al.*, "Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes : Studi Kasus di Desa

Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes : Studi Kasus di Desa Jati Mulya Kecamatan Sepatan Timur," 2024.
[15] W. P. Nurmayanti, "Penerapan Naive Bayes dalam Mengklasifikasikan Masyarakat Miskin di Desa Lepak," *Geodika: Jurnal Kajian Ilmu dan Pendidikan Geografi*, vol. 5, no. 1, pp. 123–132, Jun. 2021, doi: 10.29408/geodika.v5i1.3430.
[16] V. T. S. W. S. E. P. Rahayu Mayang Sari, *Klasifikasi Forecasting Mengunakan Algoritma Naive Bayes*. Indonesia: Serasi Media Teknologi, 2024.
[17] M. M. M. M. F. R. M. E. K. A. M. T. A. H. F. Y. W. Carudin Carudin, *Buku Ajar Data Mining*. 2024.
[18] M. M. Undari Sulung, "MEMAHAMI SUMBER DATA PENELITIAN : PRIMER, SEKUNDER, DAN TERSIER," *Edu Research*, vol. 5, pp. 110–116, 2024.
[19] Rachmadi Usman, *Hukum pencatatan sipil*. Sinar Grafika, 2019.
[20] "Bupati situbondo Provinsi Jawa Timur perbup no. 7-KEMISKINAN.2022".
[21] Y. Triani, F. Ekonomi, B. Islam, and R. Sumantri, "ANALISIS PENGETASAN KEMISKINAN DI KOTA PALEMBANG Maya Panorama," 2020.
[22] V. T. S. W. S. E. P. Rahayu Mayang Sari, *Klasifikasi Forecasting Mengunakan Algoritma Naive Bayes*. Serasi Media Teknolog, 2024.
[23] F. A. Hawari and I. D. Sholihati, "PERBANDINGAN METODE NAÏVE BAYES DAN K-NEAREST NEIGHBORS DALAM KLASIFIKASI KEPUASAN MAHASISWA TERHADAP LAYANAN WI-FI DI UNIVERSITAS NASIONAL," 2025.
[24] A. L. H. Y. S. S. M. Y. E. B. P. W. S. U. N. T. S. S. S. M. K. E. P. H. Z. A. I. S. K. A. S. R. M. Siti Maesaroh, *Bahasa Pemrograman Python*, vol. 201. Indonesia: Sada Kurnia Pustaka, 2024.
[25] P. Studi, F. Teknologi, and D. Informatika, "Rancang Bangun Aplikasi Pengelolaan Inventaris Barang Pada Dinas Pekerjaan Umum Cipta Karya Dan Tata Ruang Provinsi Jawa Timur Kerja Praktek," 2015.
[26] Seliwati, *Pengenalan Teknologi Komputer Memahami Perkembangan Hardware dan Software pada Komputer*, vol. 76. Indonesia: Indie Press, 2022.
[27] M. F. A. M. P. A. Y. P. P. M. A. Zunan Setiawan, *BUKU AJAR DATA MINING*. PT. Sonpedia Publishing Indonesia, 2023.
[28] T. I. R. S. T. , M. Sc. Grasberg Nahumarury, *PENGANTAR DATA SCIENCE Teori & Aplikatif*. Penerbit Andi, 2025.
[29] M. Eng. , IPM. , A. Eng. , Prof. Ir. P. I. S. M. Sc. , Ph. D. , IPU. · Dr. Ir. Rianto, *Data Preparation untuk Machine Learning & Deep Learning*. Penerbit Andi, 2025.
[30] Suyanto, *Machine learning : tingkat dasar dan lanjut*. INFORMATIKA, 2018.