Ibrahimy LIBRARY

IMPLEMENTASI ALGORITMA NAÏVE BAYES KLASIFIKASI PENERIMA BANTUAN BLT DI DESA PESANGGRAHAN

SKRIPSI

Oleh:

AHMAD GUNAWAN

2021503102

PROGRAM STUDI TEKNOLOGI INFORMASI

FAKULTAS SAINS DAN TEKNOLOGI UNIVERSITAS IBRAHIMY

SITUBONDO

2025

Ibrahimy Repository

Universitas Ibrahimy

repository.ibrahimy.ac.id

IMPLEMENTASI ALGORITMA NAÏVE BAYES KLASIFIKASI PENERIMA BANTUAN BLT DI DESA PESANGGRAHAN

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:

AHMAD GUNAWAN

2021503102

PROGRAM STUDI TEKNOLOGI INFORMASI

FAKULTAS SAINS DAN TEKNOLOGI UNIVERSITAS IBRAHIMY

SITUBONDO

2025

repository.ibrahimy.ac.id

Ibrahimy LIBRARY

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini :

Nama : **Ahmad Gunawan**

NPM/NIRM : 2021503102

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, ...

Saya yang menyatakan,

Ahmad Gunawan

Ibrahimy Repository
Universitas Ibrahimy
repository.ibrahimy.ac.id

Ibrahimy LIBRARY
Ibrahimy Repository
Universitas Ibrahimy

PERSETUJUAN PEMBIMBING

Nama : **Ahmad Gunawan**

NPM/NIRM : 2021503102

Program Studi : **IMPLEMENTASI ALGORITMA NAÏVE BAYES KLASIFIKASI PENERIMA BANTUAN BLT DI DESA PESANGGRAHAN**

Telah disetujui oleh:

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

repository.ibrahimy.ac.id

PENGESAHAN

SKRIPSI

IMPLEMENTASI ALGORITMA NAÏVE BAYES KLASIFIKASI PENERIMA BANTUAN BLT DI DESA PESANGGRAHAN

Ahmad Gunawan
2021503102

telah dipertahankan di depan dewan penguji Sidang/Munaqasyah Skripsi pada hari Sabtu, Tanggal 23 agustus 2025 sebagai salah satu syarat memperoleh gelar Sarjana (S.Kom) pada Fakultas Sains dan Teknologi Universitas Ibrahimy.

Tim Penguji,

Ketua Sidang,	Sekretaris Sidang,
Dr. Ach. Khumaidi, M.P.	**Muhammad Khoirroni, S.Kom**
NIDN: 0722049001	NIDN: -
Penguji I,	Penguji II,
Abd. Ghofur, M.Kom.	**Hermanto, M.Kom.**
NIDN: 0711088303	NIDN: 708087807

Mengetahui.

Dekan,

Abd. Ghofur, M.Kom.
NIDN: 0711088303

repository.ibrahimy.ac.id

Ibrahimy LIBRARY
Ibrahimy Repository
Universitas Ibrahimy

MOTTO

“ Jek abes lakona kawan abes pangaterrona Dinah la tekkaah lakona nyaman tape tadek terronah tak kera olle apa jek dhina tekkak kalakoanna tak nyaman tekkak bejerenna kenik tape raje pangaterronah bennyak ollena kawan area se ekocak aghi mangkat sossah mole senneng kawan kan baik mara ria”

repository.ibrahimy.ac.id

Ibrahimy LIBRARY
Ibrahimy Repository
Universitas Ibrahimy

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 Kakek dan nenek yang selalu support dan pengertiannya selama ini
3. Terimakasih kepada dosen pembimbing yang sudah berbaik hati meluangkan waktu dan bimbingannya
4. Terimakasih kepada sahabatku yasin dan makruf dan teman teman di Kampus Universitas Ibrahimy yang selalu memberikan support serta motivasi
5. Terimakasih kepada semua pembaca skripsi ini

repository.ibrahimy.ac.id

Ibrahimy LIBRARY

KATA PENGANTAR

Segala puji syukur peneliti sampaikan kepada Allah SWT, karena atas Rahmat dan Hidayah-Nya, perencanaan, pelaksanaan dan penyelesaian tugas akhir/skripsi dengan judul "Implementasi Algoritma Naïve Bayes Klasifikasi Penerima Bantuan BLT Di Desa Pesanggrahan" sebagai salah satu syarat penyelesaian program diploma/sarjana dapat terselesaikan dengan baik dan lancar. Kesuksesan ini dapat peneliti peroleh karena dukungan beberapa pihak. Peneliti menyampaikan 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. Sekaligus Dosen Pembimbing PKL, yang telah memberikan arahan dan bimbingan
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 Azise M.Kom selaku pembimbing I dan II,
9. Seluruh Dosen Fakultas Sains dan Teknologi yang telah memberikan kami ilmu sehingga sampai pada masa Skripsiini.

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

Universitas Ibrahimy
Ibrahimy Repository
repository.ibrahimy.ac.id

Ibrahimy LIBRARY
Ibrahimy Repository
Universitas Ibrahimy

ABSTRAK

Proses penentuan penerima Bantuan Langsung Tunai (BLT) di Desa Pesanggrahan masih dilakukan secara manual sehingga sering terjadi ketidaktepatan sasaran. Kondisi ini menimbulkan potensi ketidakadilan dan kecemburuan sosial di masyarakat. Oleh karena itu, diperlukan pendekatan berbasis teknologi yang dapat meningkatkan objektivitas dan akurasi dalam seleksi penerima bantuan sesuai kriteria yang telah ditetapkan, seperti keluarga miskin, tidak menerima bantuan lain, rentan sakit, dan lansia.

Penelitian ini mengimplementasikan algoritma klasifikasi *Naïve Bayes Gaussian* dengan tahapan data preparation, labeling, pembagian data latih dan uji menggunakan rasio 80:20, serta pengolahan atribut dengan TF-IDF *Vectorizer*. Evaluasi performa dilakukan melalui *confusion matrix*, *classification report*, serta perhitungan akurasi, *precision*, *recall*, dan *f1-score*.

Hasil penelitian menunjukkan bahwa model mampu mengklasifikasikan seluruh data uji tanpa kesalahan dengan nilai akurasi, *precision*, *recall*, dan *f1-score* sebesar 100%. Hal ini membuktikan bahwa algoritma *Naïve Bayes* efektif dalam menentukan kelayakan penerima BLT secara tepat sasaran. Penelitian berikutnya disarankan untuk membandingkan hasil dengan algoritma lain seperti *Support Vector Machine* (SVM), *Random Forest*, *Particle Swarm Optimization* (PSO), atau *Genetic Algorithm* (GA) guna memperoleh gambaran performa yang lebih luas.

Kata Kunci : Data *Naïve Bayes*, Data Mining, Bantuan Langsung Tunai, Klasifikasi, Pesanggrahan

repository.ibrahimy.ac.id

Ibrahimy LIBRARY
Ibrahimy Repository
Universitas Ibrahimy

ABSTRACT

The process of determining beneficiaries of the Direct Cash Assistance (BLT) program in Pesanggrahan Village is still carried out manually, which often leads to inaccurate targeting. This situation creates potential inequality and social jealousy among the community. Therefore, a technology-based approach is needed to improve the objectivity and accuracy of the selection process based on predefined criteria, including low-income households, families not receiving other aid, the elderly, and those with chronic illnesses.

This study applies the Gaussian Naïve Bayes classification algorithm through several stages: data preparation, manual labeling, data splitting with an 80:20 ratio for training and testing, and feature transformation using TF-IDF Vectorizer. The model's performance was evaluated using a confusion matrix, classification report, as well as accuracy, precision, recall, and f1-score metrics.

The results show that the model successfully classified all testing data without error, achieving 100% accuracy, precision, recall, and f1-score. These findings indicate that the Naïve Bayes algorithm is highly effective in classifying eligible BLT recipients with high reliability. Future research is recommended to compare this method with other algorithms such as Support Vector Machine (SVM), Random Forest, Particle Swarm Optimization (PSO), or Genetic Algorithm (GA) to provide broader insights into classification performance

***Keywords**: Naïve Bayes, Data Mining, Direct Cash Assistance, Classification, Pesanggrahan*

repository.ibrahimy.ac.id

DAFTAR ISI

HALAMAN JUDUL ……..……i
PERNYATAAN KEASLIAN TULISAN iii
PERSETUJUAN PEMBIMBING..... iv
PENGESAHAN v
MOTTO vi
PERSEMBAHAN vii
KATA PENGANTAR viii
ABSTRAK ix
ABSTRACT x
DAFTAR ISI xi
DAFTAR TABEL xiv
DAFTAR GAMBAR xv
DAFTAR RUMUS xvi
BAB I 1
PENDAHULUAN 1
1.1 Latar Belakang 1
1.2 Identifikasi Masalah 3
1.3 Rumusan Masalah 4
1.4 Batasan Masalah 4
1.5 Tujuan Penelitian 5
1.6 Manfaat Penelitian 5
1.7 Metode Penelitian 5
1.7.1 Jenis Penelitian 5
1.7.2 Metode Pengumpulan Data 6
1.8 Sistematika Pembahasan 8
BAB II 10
TINJAUAN PUSTAKA 10
2.1 Penelitian Terdahulu 10

repository.ibrahimy.ac.id

2.1.1 Implementasi Algoritma *Naive Bayes* Untuk Prediksi Calon Penerima BLT (Studi Kasus Desa Tridayasakti, Kabupaten Bekasi) 10
2.1.2 IMPLEMENTASI *NAIVE BAYES* UNTUK KLASIFIKASI KELAYAKAN PENERIMA BANTUAN SOSIAL 11
2.1.3 Impelementasi Algoritma *Navie Bayes* Untuk Prediksi Penerima Bantuan Sosial Di Desa Cigayam 12
2.2 Landasan Teori 14
2.2.1 Implementasi 14
2.2.2 Algoritma 15
2.2.3 Naive Bayes 15
2.2.4 Confusion Matrix 18
2.2.5 Klasifikasi 20
2.2.6 Bantuan Langsung Tunai (BLT) 21
2.2.7 Data Mining 22
2.3 Pemodelan 23
2.2.8 Python 26
2.4 Perangkat Lunak yang Digunakan 27
2.2.9 Microsoft Excel 27
2.2.10 RapidMiner 28
2.2.11 Gogle Colab 28
BAB III 30
ANALISIS DAN PERANCANGAN 30
3.1 Gambaran Umum Objek Penelitian 30
3.2 Pengumpulan Data 30
3.3 Alur Proses 31
3.2.1 Identifikasi dan Analisis proses Bisnis 31
3.2.2 Analisis Proses Bisnis 31
3.2.3 Identifikasi dan Analisis Kebutuhan 35
3.2.4 Identifikasi dan Analisis Alternatif Solusi 36
3.2.5 Desain Proses 37
BAB IV 42

repository.ibrahimy.ac.id

HASIL DAN PEMBAHASAN 42
4.1 Kebutuhan system 42
4.2 Data Preparation 42
4.3 Labeling Data 48
4.4 Split Data Training dan Data Testing 48
4.5 Metode Naïve Bayes 51
BAB V 59
PENUTUP 59
5.1 Kesimpulan 59
5.2 Saran 60
DAFTAR PUSTAKA 61

repository.ibrahimy.ac.id

Ibrahimy LIBRARY

DAFTAR TABEL

Tabel 2. 1 Contoh C*onfusion Matrix* 18
Tabel 3. 1 *Confusion Matrix* Penerima Bantuan BLT 33
Tabel 3. 2 Identifikasi Alternatif Solusi 36
Tabel 3. 3 Analisis Kelayakan Alternatif Solusi 37
Tabel 4. 1 Labeling Data 48

Ibrahimy Repository
Universitas Ibrahimy

repository.ibrahimy.ac.id

DAFTAR GAMBAR

Gambar 3. 1 Implementasi Algoritma Naïve Bayes 35
Gambar 3. 2 *Data preparation* 38
Gambar 3. 3 Alur Labeling 39
Gambar 3. 4 Alur Metode *Naive Bayes* 40
Gambar 3. 5 Arsitektur Sistem 41
Gambar 4. 1 Menghapus Data Duplikat 44
Gambar 4. 2 Menghapus Data Duplikat 45
Gambar 4. 3 Menghapus Baris Kosong 47
Gambar 4. 4 Menghapus Baris Kosong 47
Gambar 4. 5 Split Data Training dan Data Testing 49
Gambar 4. 6 Prediksi Naïve bayes GaussianNB 52
Gambar 4. 7 confusion matrix 53
Gambar 4. 8 Classification Report Dan Accuracy 53

repository.ibrahimy.ac.id

Ibrahimy LIBRARY
Ibrahimy Repository
Universitas Ibrahimy

DAFTAR RUMUS

Rumus 2.1 Naive Bayes 16
Rumus 2.2 Akurasi 19
Rumus 2.3 Precission 19
Rumus 2.4 Recall 19
Rumus 2.5 F1-Score 19
Rumus 3.1 Probabilitas Kategori 32
Rumus 3.2 Probabilitas Atribut 32
Rumus 3.3 Mencari nilai Vmap 33

UNIVERSITAS IBRAHIMY SUKOREJO SITUBONDO

repository.ibrahimy.ac.id

Ibrahimy LIBRARY

BAB I

PENDAHULUAN

1.1 Latar Belakang

Dalam era digital saat ini, jumlah data yang dihasilkan oleh berbagai sistem informasi terus meningkat secara eksponensial. Data yang berasal dari transaksi bisnis, media sosial, perangkat IoT, hingga sistem pemerintahan memiliki potensi besar untuk dianalisis guna mendukung pengambilan keputusan yang lebih akurat dan efisien. Namun, tanpa metode yang tepat, data tersebut hanya akan menjadi sekumpulan informasi yang tidak terstruktur dan sulit dimanfaatkan secara optimal[1].

Data mining merupakan salah satu teknik yang dapat digunakan dalam sistem informasi untuk mengekstraksi pola dan pengetahuan yang tersembunyi dari kumpulan data besar[2]. Pengelolaan data menjadi aspek penting dalam berbagai sektor, termasuk dalam penyaluran bantuan sosial seperti Bantuan Langsung Tunai (BLT). BLT merupakan program pemerintah yang bertujuan untuk membantu masyarakat kurang mampu dalam memenuhi kebutuhan dasar mereka dan meringankan beban hidup masyarakat miskin[3].

Desa Pesanggrahan adalah sebuah desa yang terletak di ujung barat Kecamatan Jangkar, Kabupaten Situbondo, Jawa Timur. Mayoritas pekerjaan masyarakat Pesanggrahan petani dan buruh tani. Dalam upaya menjaga kesejahteraan masyarakat Program Bantuan Langsung Tunai (BLT) dengan menggunakan dana desa, pemerintah berupaya memperkuat jaring pengaman sosial (JPS) [4]. Upaya penyaluran BLT ini bertujuan untuk mengatasi berbagai

Universitas Ibrahimy
Ibrahimy Repository
repository.ibrahimy.ac.id

masalah yang dihadapi oleh semua penduduk, terutama yang berkaitan dengan ekonomi. Dalam rangka mempercepat penanggulangan kemiskinan yang parah, Instruksi Presiden (Inpres) Nomor 4 Tahun 2022 menggariskan kebijakan pemberian bantuan keuangan melalui BLT[5].

Proses pemilihan penerima BLT di Desa Pesanggrahan, Kecamatan Jangkar, masih dilakukan secara manual dan tidak sepenuhnya bergantung pada data dan sistem yang terorganisir. Prosedur ini sering kali bergantung pada pendapat atau saran dari warga tertentu, yang dapat mengakibatkan teorganisir kesalahan dalam penyaluran BLT. Proses pendataan penerima bantuan BLT dianggap tidak merata, yang menyebabkan kecemburuan sosial dikalangan masyarakat. diperlukan kriteria yang tepat untuk mencegah kesalahan dalam pendataan penerima bantuan BLT[6]. Program BLT harus dilaksanakan dengan benar, transparan dan teratur agar BLT yang diberikan dapat diterima oleh masyarakat miskin yang berhak membutuhkan[7].

Algoritma *Naive Nayes* merupakan Algoritma klasifikasi dengan metode probabilitas, statistik dan memprediksi peluang dimasa depan berdasarkan pengalaman dimasa sebelumnya, sehingga dikenal dengan *Teorema Bayes*[8]. *Navie Bayes* adalah salah satu algoritma pembelajaran induktif terbaik dan paling efisien untuk *machine learning* dan *data mining*[9]. Dalam penelitian yang dilakukan oleh Wirasa dan Nurahman. Dengan judul Penerapan Metode *Naive Bayes* Dalam Mengklasifikasi Penerima BLT Pada Desa Pelangsian menunjukkan tingkat akurasi BLT di desa pelangsian menggunakan menggunakan Algoritma *Neive Bayes* Berdasarkan analisis yang sudah dilaksanakan nilai yang didapatkan

repository.ibrahimy.ac.id

pada accuracy 83.67%, precision 78.79% dan recall 74.29[10]. Pada penelitian Implementasian Algoritma *Navie Bayes* dalam klasifikasi BLT menunjukan akurasi yang baik. dengan adanya algoritma *Navie Bayes* dapat membantu aparat desa Pesanggrahan dalam klasifikasi penerima bantuan BLT. Implementasi Algoritma *Naïve Bayes*, dapat membantu proses sleksi penerima BLT berdasarkan kreteria yang sudah ditentukan sehingga menghasilkan suatu pengambilan keputusan terhadap calon penerima bantuan berdasarkan kreterian yang sudah ditetapkan kepala desa Nomor 05 Tahun 2023 tentang penetapan keluarga sasaran penerima manfaat Bantuan Langsung Tunai Desa (BLT Desa) yaitu, kelurga miskin, kelurga miskin tidak terkait bantuan lain, rentan sakit dan lansia[11]. Dengan adanya kreteria tersebut calon penerima bantuan BLT tepat sasaran berdasarkan kreteria.

Berdasarkan uraian di atas maka Algoritma klasifikasi penerima BLT menjadi solusi untuk menyelasaikan permasalahan yang ada[12]. Berdasarkan hasil tersebut Implementasi Algoritma *Naïve Bayes*, dapat membatu proses seleksi penerima BLT berdasarkan kreteria yang sudah ditentukan yaitu kelurga miskin, kelurga miskin tidak terkait bantuan lain, rentan sakit dan lansia.

1.2 Identifikasi Masalah

Berdasarkan latar belakang yang telah dipaparkan, ditemukan beberapa identifikasi masalah yaitu:

a. Proses seleksi penerima bantuan BLT di Desa Pesanggrahan masih manual, yaitu berdasarkan pendapat atau saran warga tertentu

repository.ibrahimy.ac.id

sehingga terjadi ketidak tepat sasaran dalam penyaluran bantuan penerima BLT.

b. Dibutuhkan sebuah algoritma klasifikasi dalam menetukan penerima bantuan BLT di Desa Pesanggrahan.

1.3 Rumusan Masalah

Berdasarkan latar belakang dan indentifikasi masalah di atas, rumusan masalah dalam penelitian ini yang akan dibahas yaitu "Pembuatan Aplikasi dengan Implementasi Algoritma *Naive Bayes* dalam klasifikasi dan menentukan calon penerima BLT ".

1.4 Batasan Masalah

Batasan masalah dalam penelitian ini dikarenakan cakupan teori Data Mining cukup luas untuk di bahas, sehingga memuat beberapa batasan sebagai berikut:

a. Pada penelitian menggunakan Algoritma *Navie Bayes* untuk pengklasifikasi penerima BLT di Desa Pesanggrahan.

b. Data pada penelitian didapatkan data penerima bantuan BLT di Desa Pesanggrahan, dan berdasarkan kreteria yang sudah di tentukan yaitu, kelurga miskin, kelurga miskin tidak terkait bantuan lain, rentan sakit dan lansia.

Ibrahimy LIBRARY
Ibrahimy Repository
Universitas Ibrahimy
repository.ibrahimy.ac.id

1.5 Tujuan Penelitian

Tujuan dari penelitian ini adalah Pembuatan Aplikasi dengan mengimpelementasikan Algoritma *Naive Bayes* untuk klasifikasi penerima bantuan BLT di Desa Pesanggrahan, dan membantu proses seleksi penerima bantuan BLT berdasarkan kreteria yang sudah ditentukan.

1.6 Manfaat Penelitian

Klasifikasi penerima bantuan BLT menggunakan algoritma *Naive Bayes* memilik beberapa manfaat:

a. Dapat memberikan solusi berbasis teknologi Data Mining untuk meningkatkan akurasi prediksi dalam proses seleksi penerima bantuan BLT tepat sasaran.
b. Dengan adanya Implementasi Algoritma *Navie Bayes* dapat membantu kepala desa dan aparat desa dalam Proses seleksi penerima bantuan BLT Berdasarkan kreteria yang di tentukan.

1.7 Metode Penelitian

1.7.1 Jenis Penelitian

a. Kuantitatif

Jenis penelitian yang digunakan adalah deskriptif kuantitatif. Penelitian deskriptif kuantitatif adalah teknik yang menggambarkan, menjelaskan, atau meringkas berbagai pengaturan, situasi, kejadian, atau variabel penelitian berdasarkan kejadian aktual yang dapat diamati, difoto, diwawancarai, dan didokumentasikan[13].

Ibrahimy LIBRARY
Ibrahimy Repository
Universitas Ibrahimy
repository.ibrahimy.ac.id

b. Eksperimental

Metode eksperimental merupakan cara untuk memberikan perlakuan tertentu pada unit eksperimen, di mana penataan unit eksperimen tersebut dikenal sebagai proses perancangan atau desain eksperimen. Tujuan dari prosedur ini adalah untuk menganalisis dampak perubahan suatu variabel terhadap variabel lainnya.[14]. Penelitian ini menggunakana bahasa pemrograman *Python* Dan *RapidMiner* untuk mengelola dan menganalisis data penerima bantuan BLT di Desa Pesanggrahan.

1.7.2 Metode Pengumpulan Data

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

a. Observasi

Observasi adalah upaya untuk menghimpun data yang dilakukan dengan cara mengamati dan mencatat gejala (fenomena) secara sistematis terhadap suatu subjek objek pengamatan[15].melakukan observasi langsung ke kantor desa pesanggrahan.

b. .Wawancara

Wawancara adalah metode komunikasi di mana pertanyaan dan tanggapan digunakan untuk mendapatkan informasi, pendapat, atau cerita dari seseorang[16]. Dalam penelitian ini melakukan wawancara langsung ke kepala desa.

repository.ibrahimy.ac.id

c. Studi Literatur

Studi literatur adalah mencari referensi teoretis yang berkaitan dengan kasus atau masalah yang sedang diselidiki untuk menyelesaikannya dengan melacak sumber-sumber literatur yang telah diterbitkan sebelumnya. Peneliti dapat menggunakan semua data dan ide yang relevan, serta pemahaman yang lebih komprehensif dan mendalam tentang masalah yang sedang diselidiki, dengan melakukan studi literatur[17]. Dengan menggunakan metode literatur ini, para peneliti menggunakan data sekunder dari buku-buku, jurnal penelitian terdahulu, dan sejumlah *e-journal*, seperti *Google Schoolar*, untuk mendukung dan membantu pembuatan laporan.

d. Dokumentasi

Metode dokumentasi merupakan teknik penelitian yang memanfaatkan berbagai sumber tertulis, seperti buku, majalah, dokumen, peraturan, notulen rapat, catatan harian, dan media sejenis. Dokumentasi juga mencakup segala bentuk bahan tertulis maupun film yang tidak dibuat khusus atas permintaan peneliti. Dalam pengumpulan data, metode ini digunakan untuk memperoleh informasi sebanyak mungkin dari berbagai media cetak yang dapat dijadikan acuan atau pedoman penelitian. Tujuannya adalah melengkapi data yang diperoleh melalui metode observasi dan wawancara.[18].Dalam dokumentasi penelitian ini mendapatkan data penerima bantuan BLT di kantor desa pesanggrahan.

Ibrahimy LIBRARY
Ibrahimy Repository
Universitas Ibrahimy
repository.ibrahimy.ac.id

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 penelitian yang diangkat, juga memuat landasan teori dan perangkat lunak yang akan digunakan.

BAB III : METODELOGI PENELITIAN

Bab ini membahas tentang bagaimana peneliti mendapatkan data yang digunakan dalam penelitian ini dan bagaimana cara mendapatkannya berikut memprosesnya sebelum melakukan analisa lebih lanjut menggunakan algoritma *Naïve Bayes*.

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 menghadilkan sebuah prediksi berdasarkan kemungkinan yang telah didapatkan sebelumnya dan juga tingkat akurasi dari prediksi yang telah dihasilkan.

Ibrahimy LIBRARY
Ibrahimy Repository
Universitas Ibrahimy
repository.ibrahimy.ac.id

BAB V : PENUTUP

Kesimpulan dari penelitian yang dibuat serta saran atau rekomendasi bagi penulis tentang sistem yang dibuat untuk pengembangan aplikasi, lembaga maupun instansi. Semua proses dalam penyusunan penelitian ini disajikan secara ringkas berdasarkan analisis dan data yang telah di uraikan pada bab sebelumnya.

Ibrahimy LIBRARY
Ibrahimy Repository
Universitas Ibrahimy
repository.ibrahimy.ac.id

Ibrahimy LIBRARY

BAB II

TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

2.1.1 Implementasi Algoritma *Naive Bayes* Untuk Prediksi Calon Penerima BLT (Studi Kasus Desa Tridayasakti, Kabupaten Bekasi)

Penelitian dilakukan oleh Muhamad Adhi Mukti, Wiyanto Wiyanto, Ismasari Nawangsih dari Universitas Pelita Bangsa. Dipublikasikan pada tahun 2023 di JINTEKS (Jurnal Informatika Teknologi dan Sains).

Penelitian ini dilakukan di Desa Tridayasakti, Kabupaten Bekasi, untuk mengatasi permasalahan dalam menentukan penerima Bantuan Langsung Tunai (BLT), di mana masyarakat merasa bahwa proses seleksi tidak objektif dan kurang transparan. Oleh karena itu, penelitian ini menerapkan Algoritma *Naive Bayes* dalam teknik Data Mining untuk membantu pengambilan keputusan secara lebih akurat mengenai kelayakan penerima BLT. Metode yang digunakan dalam penelitian ini adalah CRISP-DM, yang terdiri dari enam tahapan*: Business Understanding*, Data *Understanding*, Data *Preparation*, *Modeling*, *Evaluatio*n, dan *Deployment*. Pengujian dilakukan dengan dua cara, yaitu secara manual menggunakan 282 data training dan 2 data testing, serta dengan RapidMiner menggunakan 197 data *training* dan 85 data t*esting*.

Penelitian ini memiliki beberapa manfaat utama, di antaranya meningkatkan transparansi dalam proses seleksi penerima BLT, mempermudah pihak desa dalam melakukan seleksi berdasarkan data yang lebih akurat, mengurangi kesalahan subjektif dalam pemilihan penerima bantuan,

repository.ibrahimy.ac.id

mengimplementasikan sistem berbasis web menggunakan bahasa pemrograman PHP yang dapat diakses oleh masyarakat, serta meningkatkan efisiensi waktu dalam pengolahan data penerima BLT dibandingkan metode manual.

Tujuan utama penelitian ini adalah mengembangkan sistem berbasis *Naive Bayes* yang dapat membantu dalam klasifikasi penerima BLT secara lebih objektif dan akurat. Hasil penelitian menunjukkan bahwa pengujian manual menggunakan *Confusion Matrix* menghasilkan nilai *Accuracy*, *Precision*, dan *Recall* masing-masing sebesar 100%. Sementara itu, pengujian menggunakan *RapidMiner* menghasilkan nilai *Accuracy* sebesar 98,82%, *Precision* sebesar 100%, dan *Recall* sebesar 96,97%. Dengan hasil tersebut, sistem berbasis *Naive Bayes* terbukti sangat akurat dalam mengklasifikasikan calon penerima BLT[19].

2.1.2 IMPLEMENTASI *NAIVE BAYES* UNTUK KLASIFIKASI KELAYAKAN PENERIMA BANTUAN SOSIAL

Penelitian dilakukan oleh Aisyah Fatmawati DKK.dari Universitas k STMIK IKMI CIREBON Publikasi Tahun 2024 di Jurnal JATI (Jurnal Mahasiswa Teknik Informatika).

Penelitian ini dilakukan untuk meningkatkan efisiensi dan keadilan dalam pendistribusian bantuan sosial, terutama Bantuan Pangan Non Tunai (BPNT) dan Program Keluarga Harapan (PKH), dengan menerapkan Algoritma Naïve Bayes. Penelitian ini berfokus pada Desa Jungjang, Kecamatan Aljawinangun, yang memiliki populasi penduduk terbesar dibandingkan desa lainnya. Salah satu permasalahan utama dalam distribusi bantuan sosial adalah masih banyak warga kurang mampu yang tidak mendapatkan bantuan karena

repository.ibrahimy.ac.id

pendataan yang belum optimal dan penetapan status keluarga miskin yang kurang akurat. Oleh karena itu, metode *Naïve Bayes* diterapkan untuk mengklasifikasikan penerima bantuan sosial secara lebih objektif.

Manfaat dari penelitian ini antara lain mempermudah lembaga penyedia bantuan sosial dalam menentukan penerima yang tepat sasaran, meningkatkan akurasi dalam klasifikasi kelayakan penerima bantuan, serta mengoptimalkan penggunaan sumber daya pemerintah dalam pendistribusian bantuan sosial. Dengan penggunaan *RapidMiner*, analisis dan klasifikasi data dapat dilakukan secara lebih cepat dan akurat dibandingkan dengan metode manual yang memakan waktu lebih lama.

Tujuan utama dari penelitian ini adalah untuk mengembangkan model klasifikasi yang mampu memprediksi kelayakan penerima bantuan sosial secara lebih akurat dan efisien, sehingga dapat membantu pemerintah dalam pengambilan keputusan yang lebih tepat. Hasil penelitian menunjukkan bahwa nilai akurasi klasifikasi menggunakan algoritma *Naïve Bayes* mencapai 86,36%, dengan *precision* sebesar 98,41% dan *recall* sebesar 87,32%. Dengan hasil tersebut, sistem berbasis *Naïve Bayes* terbukti cukup efektif dalam mengklasifikasikan penerima bantuan sosial sesuai kriteria yang telah ditentukan.

2.1.3 Impelementasi Algoritma *Navie Bayes* Untuk Prediksi Penerima Bantuan Sosial Di Desa Cigayam

Penelitian dilakukan oleh Dede Hoeriah, Bani Nurhakim, Sandy Eka Permana, Willy Prihartono, dan Gifthera Dwilestari. Jurnal ini dipublikasikan pada tahun 2024 di TAMIKA,

repository.ibrahimy.ac.id

Penelitian ini dilakukan di Desa Cigayam untuk mengatasi permasalahan terkait ketidakakuratan dalam penyaluran Bantuan Sosial (Bansos). Banyak masyarakat yang kurang mampu tidak mendapatkan bantuan karena kesalahan dalam sistem pendataan. Untuk menyelesaikan masalah ini, penelitian ini menggunakan Algoritma *Naïve Bayes* dengan metode Knowledge Discovery Database (KDD) guna memprediksi penerima bantuan sosial secara lebih akurat. Data yang digunakan berasal dari data kependudukan Desa Cigayam serta data penerima bansos yang telah ada sebelumnya.

Penelitian ini memiliki beberapa manfaat utama, di antaranya meningkatkan akurasi dalam klasifikasi penerima bantuan sosial, mempermudah pemerintah desa dalam menentukan penerima bantuan yang lebih tepat sasaran, mengurangi kesalahan dalam pendataan bantuan sosial, serta memanfaatkan teknologi Data Mining untuk efisiensi analisis data penerima bantuan.

Tujuan utama dari penelitian ini adalah untuk mengembangkan sistem klasifikasi penerima bantuan sosial yang lebih akurat dan efisien. Hasil penelitian menunjukkan bahwa pengujian menggunakan 1047 data dengan 53 data tidak sesuai menghasilkan akurasi tinggi dengan nilai *precision* 95,18%, *recall* 81,17%, dan tingkat kesalahan sebesar 28,38%. Dengan menggunakan kurva ROC, penelitian ini menunjukkan nilai AUC sebesar 0,868, yang dikategorikan dalam tingkat keakuratan "baik". Dengan hasil ini, metode *Naïve Bayes* terbukti efektif dalam meningkatkan ketepatan klasifikasi penerima bantuan sosial di Desa Cigayam[20].

repository.ibrahimy.ac.id

2.2 Landasan Teori

2.2.1 Implementasi

Implementasi merupakan proses pelaksanaan atau penerapan dari suatu rencana yang telah disusun secara rinci dan matang. Tahap ini biasanya dilakukan setelah perencanaan dianggap final, dengan tujuan memastikan kebijakan dapat dijalankan sesuai ketentuan serta mencapai sasaran yang telah ditetapkan. Dengan demikian, implementasi dapat dipandang sebagai langkah nyata untuk mewujudkan dan mengawal keberhasilan suatu kebijakan[21].

Implementasi dapat dipahami sebagai suatu bentuk tindakan nyata dari rencana yang telah disusun secara detail untuk mencapai tujuan tertentu. Proses implementasi baru dapat dilakukan apabila tahap perencanaan dianggap matang. Menurut teori Jones, implementasi merupakan "*those activities directed toward putting a program into effect*" atau kegiatan yang bertujuan merealisasikan suatu program hingga menghasilkan output yang diharapkan. Dengan demikian, implementasi adalah langkah konkret yang dilakukan setelah kebijakan ditetapkan agar tujuan kebijakan tersebut dapat tercapai.

Sementara itu, Nurdin Usman dalam bukunya Konteks Implementasi Berbasis Kurikulum menyatakan bahwa implementasi berkaitan dengan aktivitas, tindakan, atau mekanisme dalam suatu sistem. Ia menekankan bahwa implementasi bukan sekadar kegiatan biasa, melainkan aktivitas yang direncanakan secara sistematis guna mencapai tujuan tertentu[22].

repository.ibrahimy.ac.id

2.2.2 Algoritma

Algoritma adalah Langkah-langkah terperinci dan terorganisir yang digunakan untuk menyelesaikan masalah atau mencapai tujuan. Dalam bidang komputer, algoritma adalah serangkaian instruksi yang jelas dan terdefinisi yang dapat dieksekusi oleh komputer untuk menyelesaikan tugas atau memecahkan masalah. Algoritma dapat digunakan untuk memecahkan berbagai masalah, baik yang sederhana maupun kompleks. Algoritma digunakan dalam berbagai bidang, seperti ilmu komputer, matematika, ilmu data, kecerdasan buatan, dan banyak lagi.[23].

Berikut merupakan beberapa definisi algoritma dari beberapa pakar dunia diantaranya:

a. Abu Ja'far Muhammad Ibnu Musa Al-khawarizmi merupakan seorang ahli dalam bidang matematika, astronomi, astrologi dan geografi yang berasal dari Irak. Menurutnya algoritma adalah suatu metode khusus untuk menyelesaikan sebuah masalah.
b. Donald E. Knuth merupakan seorang ilmuwan komputer terkenal dan saat ini menjadi seorang profesor di Stanford University. Menurutnya algoritma adalah sekumpulan aturan-aturan berhingga yang memberikan deretan proses-proses untuk menyelesaikan suatu masalah yang khusus.

2.2.3 Naive Bayes

Naive Bayes adalah Klasifikasi yang didasarkan pada metode probabilistik dan statistik yang diusulkan oleh ilmuwan Inggris Thomas Bayes, yang memperkirakan probabilitas masa depan berdasarkan pengalaman yang telah

repository.ibrahimy.ac.id

berlalu, disebut teori Bayes. Teori ini terkait dengan Naive Bayes, yang menyatakan bahwa variabel-variabel tidak bergantung pada variabel lainnya. Klasifikasi dengan Naive Bayes didasarkan pada hipotesis bahwa ada atau tidak adanya ciri-ciri khusus dalam suatu kelas tidak bergantung pada ciri-ciri kelas lainnya.[24]. Persamaan dari teorema Bayes adalah sebagai berikut:

P(HX)=(P(X|H).P(H))/(P(X)) (2.1)

1 dari 1 X

Keterangan:

X: Data dengan kelas yang belum diketahui.

H: Hipotesis data X adalah suatu kelas spesifik.

P(HX) Probabilitas hipotesis H sesuai kondisi X (posteriori probability).

P(H): Probabilitas hipotesis H (prior probability).

P(X H): Probabilitas X sesuai kondisi terhadap hipotesis H. P(X):

Probabilitas X

Teori *Naive Bayes* dipandang sebagai salah satu pendekatan statistik yang mendasar dalam bidang pengenalan pola (*pattern recognition*). Pendekatan ini berlandaskan pada kuantifikasi *trade-off* antara berbagai keputusan klasifikasi dengan memanfaatkan probabilitas serta biaya yang mungkin muncul dari suatu keputusan. Dari dasar pemikiran tersebut, lahirlah algoritma *Naive Bayes*, yaitu metode klasifikasi berbasis probabilitas dan statistik. Algoritma ini diperkenalkan oleh ilmuwan asal Inggris, *Thomas Bayes*, yang secara umum teorinya menekankan pada prediksi peluang di masa mendatang dengan mempertimbangkan pengalaman atau data yang terjadi sebelumnya[25].

repository.ibrahimy.ac.id

Terdapat beberapa jenis algoritma yang termasuk dalam keluarga *Naïve Bayes Classifier*, di antaranya *Gaussian Naïve Bayes***,** *Multinomial Naïve Bayes*, serta *Bernoulli Naïve Bayes*. Pada penelitian ini, metode yang dipilih adalah *Gaussian Naïve Bayes (GNB*) merupakan salah satu varian dari algoritma *Naïve Bayes Classifier* yang banyak digunakan dalam permasalahan klasifikasi, khususnya ketika fitur data yang dianalisis berbentuk kontinu atau numerik. Berbeda dengan Bernoulli dan *Multinomial Naïve Bayes* yang lebih cocok untuk data biner maupun data frekuensi kata, *Gaussian Naïve Bayes* menggunakan asumsi bahwa distribusi data mengikuti pola distribusi normal (*Gaussian distribution*).

Dalam penerapannya, setiap atribut atau variabel input diasumsikan memiliki kontribusi yang independen terhadap kelas target. GNB kemudian menghitung probabilitas suatu data masuk ke dalam kelas tertentu dengan menggunakan mean (rata-rata) dan variance (ragam) dari distribusi normal. Rumus distribusi Gaussian digunakan untuk memperkirakan peluang kemunculan sebuah nilai atribut terhadap setiap kelas.Kelebihan algoritma ini adalah kemampuannya menangani data dengan skala kontinu, proses komputasi yang sederhana, serta kecepatan dalam melakukan klasifikasi meskipun dataset yang digunakan cukup besar. Selain itu, *Gaussian Naïve Bayes* sering dipilih karena cukup robust meskipun terdapat asumsi independensi antar fitur yang tidak sepenuhnya terpenuhi di dunia nyata. Dengan demikian, algoritma ini menjadi pilihan yang tepat untuk penelitian yang melibatkan data numerik dan

Ibrahimy LIBRARY
Ibrahimy Repository
Universitas Ibrahimy
repository.ibrahimy.ac.id

membutuhkan metode klasifikasi yang efisien serta mudah diimplementasikan[26].

2.2.4 Confusion Matrix

Metode yang membandingkan nilai kelas asli dengan perkiraan untuk menilai seberapa baik pendekatan klasifikasi memprediksi kelas data. Tabel di bawah ini adalah contohnya. Perhitungan berikut ini dilakukan dalam C*onfusion Matrix*[25]:

Tabel 2. 1 Contoh C*onfusion Matrix*

Skor Perkiraan		Skor Perkiraan	
		(Negative)	(positive)
Skor Perkiraan	**(Positive)**	TP	FP
	(Negative)	FN	TN

Pada table 2.1 *confusion matrix* terdapat empat nilai yang dikeluarkan antara lain:

a. Nilai saat ini yang positif dan diantisipasi secara akurat dikenal sebagai True Positive (TP).
b. Nilai saat ini yang negatif dan diantisipasi secara akurat dikenal sebagai True Negative (TN).
c. False Positive (FP) adalah nilai saat ini yang diproyeksikan positif tetapi sebenarnya negatif.
d. False Negative (FN) adalah hasil yang saat ini positif tetapi diperkirakan negatif.

repository.ibrahimy.ac.id

Empat hasil perhitungan menggunakan *Confusion Matrix* untuk pengujian meliputi :

a. Menghitung akurasi, berikut rumus untuk menghitung akurasi:

$$Akurasi = \frac{TP+TN}{(TP+TN+FP+FN)} X100\% \quad (2.2)$$

b. Menghitung presisi, berikut rumus untuk menghitung presisi:

$$Precission = \frac{TP}{TP+FP} \quad (2.3)$$

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

$$Recall = \frac{TP}{TP+FN} \quad (2.4)$$

d. Menghitung *F1-score*, berikut rumus untuk menghitung F1-*score*:

$$F1 = 2x \frac{precission\ x\ recall}{precission+\ recall} \quad (2.5)$$

Salah satu keunggulan utama dari *Confusion Matrix* adalah kemampuannya memberikan gambaran yang lebih rinci mengenai performa model, khususnya dalam mengidentifikasi jenis kesalahan yang terjadi. Tidak seperti metrik akurasi yang hanya menyoroti seberapa banyak prediksi yang benar secara keseluruhan, *Confusion Matrix* mampu menunjukkan detail kesalahan dalam bentuk *False Positive (FP)* maupun *False Negative (FN)*. Hal ini sangat penting dalam bidang yang sensitif, misalnya pada sistem pendukung keputusan medis. Sebagai contoh, pada kasus diagnosis kanker, akurasi yang tinggi belum tentu menunjukkan kinerja model yang baik. Jika ternyata model menghasilkan tingkat *False Negative* yang tinggi—yakni kasus kanker yang justru diprediksi tidak kanker—maka konsekuensinya dapat berakibat fatal bagi pasien. Oleh sebab itu, metrik lain seperti *recall* yang lebih menekankan pada keberhasilan

Ibrahimy LIBRARY
Ibrahimy Repository
Universitas Ibrahimy
repository.ibrahimy.ac.id

mendeteksi *True Positive* menjadi jauh lebih relevan daripada sekadar bergantung pada nilai akurasi.

Menurut pendapat Tom Mitchell, salah satu pakar dalam bidang *machine learning*, Confusion Matrix merupakan instrumen yang sangat mendasar dalam proses evaluasi model klasifikasi, karena mampu membantu peneliti atau pengembang mengetahui secara jelas pada bagian mana model melakukan kesalahan. Sementara itu, Andrew Ng, seorang tokoh terkemuka dalam pengembangan kecerdasan buatan, menegaskan bahwa pemahaman yang baik mengenai Confusion Matrix sangat krusial terutama saat bekerja dengan data yang tidak seimbang (*imbalanced dataset*). Dalam kondisi seperti ini, akurasi saja bisa menyesatkan, sebab model mungkin cenderung hanya memprediksi kelas mayoritas tanpa benar-benar mempelajari pola dari keseluruhan data. Oleh karena itu, *Confusion Matrix* memberikan kontribusi yang signifikan dalam pemilihan metrik evaluasi yang sesuai dengan kebutuhan serta konteks aplikasi[27].

2.2.5 Klasifikasi

Klasifikasi adalah proses untuk mengelompokkan suatu objek ke dalam kelas, kategori, atau kelompok tertentu berdasarkan prosedur, karakteristik, serta definisi yang telah ditentukan sebelumnya. Dalam bidang data mining, klasifikasi menjadi salah satu komponen penting karena memungkinkan data diolah sehingga memiliki nilai signifikan dan dapat dimanfaatkan lebih lanjut. Proses ini termasuk salah satu tugas utama dalam penambangan data, di mana sebuah *classifier* dibentuk menggunakan kumpulan data pelatihan yang telah memiliki kategori atau label yang telah ditetapkan[8].

repository.ibrahimy.ac.id

Sistem klasifikasi umum merupakan sistem yang hanya menggunakan kelas utama beserta divisi dan sub-divisi dasar. Sebaliknya, sistem klasifikasi khusus mencakup seluruh aspek sub-divisi hingga ke topik atau subjek yang paling spesifik. Baik sistem klasifikasi umum maupun khusus, tujuan utamanya tetap sama, yaitu mengklasifikasikan bahan pustaka atau koleksi agar dapat disusun serta ditempatkan di rak secara teratur dan sistematis[28].

2.2.6 Bantuan Langsung Tunai (BLT)

Program Bantuan Langsung Tunai (BLT) merupakan bentuk kompensasi jangka pendek yang dirancang untuk menjaga agar tingkat konsumsi Rumah Tangga Sasaran (RTS)—yang termasuk kategori sangat miskin, miskin, maupun hampir miskin (near poor)—tidak mengalami penurunan ketika terjadi kenaikan harga BBM di dalam negeri. Meskipun BLT bukan satu-satunya upaya dalam penanggulangan kemiskinan, program ini diharapkan mampu berkontribusi dalam mengurangi angka kemiskinan pada masa penyesuaian harga kebutuhan pokok hingga tercapai keseimbangan baru (Departemen Sosial RI, 2008: 5). BLT sendiri adalah sejumlah dana tunai yang diberikan pemerintah kepada rumah tangga yang membutuhkan, dengan tujuan menjaga kestabilan kesejahteraan mereka saat harga BBM meningkat. RTS didefinisikan sebagai rumah tangga yang berada pada kategori sangat miskin, miskin, atau mendekati miskin (Departemen Sosial RI, 2008: 6).[29].

BLT Desa diberikan kepada keluarga miskin sebagai sasaran penerima manfaat yaitu[11]:

repository.ibrahimy.ac.id

a. Keluarga miskin yang ditetapkan berdasarkan Data Terpadu Kesejahteraan Sosial (DTKS) dan/atau Data Terpadu Daerah Analisis Kependudukan Partisipatif (DTD-AKP).
b. Keluarga miskin yang belum pernah menerima ataupun terdaftar sebagai penerima manfaat program Jaring Pengaman Sosial (JPS), seperti Program Keluarga Harapan (PKH), Bantuan Pangan Non Tunai (BPNT), Kartu Prakerja (KP), maupun bentuk bantuan JPS lainnya.
c. Keluarga miskin ekstrem Antara lain:
    1. Rumah tangga miskin atau tidak mampu yang menetap di desa terkait, dengan prioritas diberikan kepada keluarga miskin yang masuk kategori kemiskinan ekstrem.
    2. Rumah tangga yang memiliki anggota keluarga dengan kondisi rentan terhadap penyakit kronis atau menahun.
    3. Rumah tangga dengan anggota yang merupakan lansia dan tinggal sendiri.
    4. Rumah tangga yang memiliki anggota keluarga penyandang disabilitas.

2.2.7 Data Mining

Data *Mining* merupakan proses untuk menemukan pola atau informasi yang bernilai dari kumpulan data tertentu dengan memanfaatkan teknik atau metode khusus. Beragam teknik, metode, maupun algoritma dapat digunakan dalam proses ini. Pemilihan metode atau algoritma yang sesuai sangat ditentukan

repository.ibrahimy.ac.id

oleh tujuan serta tahapan *Knowledge Discovery in Database* (KDD) secara keseluruhan[30].

Data mining merupakan proses untuk menggali informasi serta menemukan pola-pola berharga dari kumpulan data berukuran besar. Tahapan dalam data mining meliputi pengumpulan, ekstraksi, analisis, hingga pengolahan statistik terhadap data. Konsep ini juga sering disebut dengan istilah *knowledge discovery, knowledge extraction, pattern analysis,* atau information *harvesting* (Arhami & Nasir). Keempat tahapan tersebut pada akhirnya menghasilkan suatu model atau pengetahuan yang bermanfaat.

Menurut Muflikhah , data mining dapat dipahami sebagai upaya untuk menguraikan kompleksitas data sehingga menghasilkan informasi yang sebelumnya tersembunyi atau tidak terlihat secara langsung. Dengan memanfaatkan perangkat otomatis maupun semi-otomatis, data mining berfungsi untuk mengekstraksi serta menganalisis data dalam jumlah besar dengan tujuan menemukan pola yang bermakna. Oleh karena itu, data mining menjadi bagian integral dari konsep *Knowledge Discovery in Databases* (KDD)[31].

2.3 Pemodelan

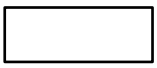
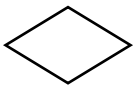
Pemodelan penelitian ini menggunakan *flowchart*, yaitu representasi grafis yang menggambarkan langkah-langkah serta proses suatu program. *Flowchart* membantu analis dan programmer memecah permasalahan menjadi bagian-bagian yang lebih kecil serta mempertimbangkan berbagai alternatif solusi dalam pengoperasian. Umumnya, *flowchart* memudahkan penyelesaian masalah, terutama yang membutuhkan kajian dan evaluasi lebih lanjut. Bentuk *flowchart*

repository.ibrahimy.ac.id

berupa diagram dengan aliran satu atau dua arah secara berurutan. Diagram ini digunakan untuk merepresentasikan maupun merancang program, sehingga harus mampu menggambarkan komponen-komponen yang terdapat dalam bahasa pemrograman.[32]

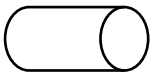
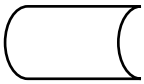
Simbol-simbol yang digunakan dalam *Flowchart* ditunjukkan pada table 2.1.

Tabel 2.2 Simbol *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
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

repository.ibrahimy.ac.id

Tabel 2.2 Simbol Flowchart Lanjutan

No	Simbol	Nama	Arti
12		Magnetik Drum	I/O yang menggunakan drum magnetik
13		*On Line Strorage*	I/O yang menggunakan penyimpanan akan langsung
14		Dokumen	I/O dalam format yang di cetak

Flowchart merupakan representasi visual yang digunakan untuk menggambarkan urutan langkah atau tahapan dalam menyelesaikan suatu masalah maupun menjalankan suatu proses tertentu. Diagram ini banyak dimanfaatkan dalam pemrograman, analisis sistem, hingga perancangan proses bisnis karena mampu menyederhanakan prosedur yang kompleks menjadi lebih mudah dipahami melalui simbol-simbol grafis. *Flowchart* tersusun dari bentuk-bentuk geometri seperti persegi panjang yang melambangkan aktivitas atau instruksi, lingkaran atau oval yang menandakan titik awal dan akhir, belah ketupat yang mewakili keputusan atau kondisi, serta panah yang menunjukkan arah aliran proses. Dengan memadukan simbol-simbol tersebut, alur logika suatu sistem dapat divisualisasikan secara runtut dan sistematis. Kelebihan utama penggunaan *flowchart* adalah kemampuannya memberikan gambaran menyeluruh tentang suatu proses sehingga memudahkan dalam mengidentifikasi potensi kesalahan, hambatan, maupun bagian yang perlu perbaikan. Selain itu, *flowchart* juga menjadi sarana komunikasi yang efektif antara pihak teknis maupun non-teknis karena diagram yang dihasilkan jelas dan mudah dipahami. Oleh sebab itu,

Ibrahimy LIBRARY
Ibrahimy Repository
Universitas Ibrahimy
repository.ibrahimy.ac.id

flowchart sering dipakai baik dalam tahap perencanaan maupun evaluasi sistem. Dalam pengembangan perangkat lunak misalnya, *flowchart* membantu programmer memahami alur algoritma sebelum dituangkan dalam kode, sedangkan dalam dunia bisnis flowchart berfungsi untuk menggambarkan prosedur operasional standar agar koordinasi antarbagian lebih terarah. Dengan demikian, *flowchart* tidak hanya berfungsi sebagai dokumentasi, tetapi juga sebagai alat analisis, perancangan, sekaligus evaluasi proses agar lebih efisien dan terstruktur[33].

2.2.8 Python

Python merupakan bahasa pemrograman bersifat dinamis yang banyak dimanfaatkan dalam pengembangan aplikasi di berbagai bidang. Bahasa ini memungkinkan penulisan program dengan beragam pendekatan secara bersamaan. Contohnya, antarmuka grafis dapat dibuat menggunakan paradigma berorientasi objek, sedangkan proses pengolahan data bisa menggunakan pendekatan fungsional atau prosedural. *Python* juga menawarkan berbagai fitur yang dapat dimanfaatkan oleh pengembang perangkat lunak[34].

Pemrograman *Python* dapat dimanfaatkan untuk membuat visualisasi grafik dari fungsi dengan dua variabel menggunakan pustaka bawaan, seperti *Matplotlib* dan *NumPy*. Pustaka *Matplotlib* berperan dalam menghasilkan berbagai jenis visualisasi data, misalnya *scatter plot*, *line chart*, *bar chart*, maupun *pie chart*, serta memungkinkan pengguna menambahkan elemen pendukung seperti judul atau keterangan lainnya. Sementara itu, *NumPy* sangat bermanfaat pada proses analisis data maupun pemodelan matematis, terutama

repository.ibrahimy.ac.id

karena kemampuannya dalam mengolah data numerik berbentuk *array*. *NumPy* mendukung beragam operasi, mulai dari perhitungan aritmetika, trigonometri, fungsi matematika hingga analisis statistik. Melalui penggunaan *NumPy*, pengolahan data numerik dapat berlangsung lebih cepat dan efisien dalam ekosistem *Python*[35].

2.4 Perangkat Lunak yang Digunakan

2.2.9 Microsoft Excel

Microsoft Excel adalah Sebuah aplikasi atau program yang merupakan bagian dari paket instalasi *Microsoft Office*, digunakan untuk mengolah angka menggunakan lembar kerja yang terdiri dari baris dan kolom untuk menjalankan perintah. *Microsoft Excel* telah menjadi perangkat lunak pengolahan data/angka terbaik di dunia, dan juga tersedia di berbagai *platform*. *Microsoft Excel* tidak hanya dapat diakses di sistem operasi *Windows*, tetapi juga di *MacOS*, *Android*, dan perangkat *Apple*.

Microsoft Excel menggunakan sistem *workbook* yang berisi beberapa *worksheet* atau lembar kerja. Lembar kerja ini terdiri dari baris dan kolom yang membentuk sel-sel. Di dalam sel, pengguna dapat menuliskan data, memasukkan rumus, atau menggunakan fungsi/formula untuk mengolah data administrasi perusahaan secara otomatis, baik untuk menghasilkan laporan berbentuk tabel maupun grafik.[36].

repository.ibrahimy.ac.id

2.2.10 RapidMiner

RapidMiner adalah perangkat lunak yang bersifat *open source* yang digunakan sebagai solusi untuk melakukan analisis data *mining*, *text mining*, serta analisis prediktif. Aplikasi ini memanfaatkan berbagai metode deskriptif dan prediktif untuk memberikan wawasan kepada pengguna sehingga mereka dapat mengambil keputusan yang optimal. *RapidMiner* menyediakan sekitar 500 operator data mining, mencakup fungsi untuk *input*, *output*, pemrosesan data, hingga visualisasi. Perangkat lunak ini dapat digunakan secara mandiri untuk analisis data maupun berfungsi sebagai proses data mining yang bisa diintegrasikan ke dalam produk lain. *RapidMiner* dikembangkan menggunakan bahasa pemrograman *Java*, sehingga kompatibel di berbagai sistem operasi.

RapidMiner menyediakan GUI (*Graphical User Interface*) yang dapat memudahkan pengguna dalam merancang pipeline analisis. Antarmuka ini menghasilkan file XML (*Extensible Markup Language*) yang berisi definisi proses analisis sesuai kebutuhan pengguna untuk diterapkan pada data. File tersebut kemudian dibaca oleh *RapidMiner* sehingga analisis dapat dijalankan secara otomatis.[37].

2.2.11 Gogle Colab

Gogle Colab adalah pengembangan terintegrasi untuk pemrograman dalam *Python*, proses yang di lakukan oleh server *Google* yang dilengkapi dengan perangkat keras berkinerja tinggi. Dari segi perangkat lunak, *Google Colab* telah menyediakan sebagian besar *library* yang dibutuhkan[38].

repository.ibrahimy.ac.id

Secara umum, *Google Colab* memanfaatkan bahasa pemrograman *Python* dalam penerapannya. Proses penyelesaian metode bagi dua dituliskan dalam bentuk skrip, mencakup pendefinisian fungsi, pelaksanaan iterasi, hingga penentuan batas toleransi kesalahan. Setiap hasil iterasi dapat ditampilkan secara otomatis dalam format tabel maupun grafik dengan bantuan pustaka seperti *Matplotlib,* sehingga memudahkan pengguna dalam memvisualisasikan proses konvergensi. Selain itu, *Google Colab* menawarkan fleksibilitas tinggi dalam pengolahan data numerik serta mendukung kecepatan eksekusi yang optimal. Berikut disajikan tahapan penyelesaian persamaan nonlinier dengan metode bagi dua menggunakan *Google Colab*[38].

repository.ibrahimy.ac.id

BAB III

ANALISIS DAN PERANCANGAN

3.1 Gambaran Umum Objek Penelitian

Desa Pesanggrahan yang terletak di Kecamatan Jangkar, Kabupaten Situbondo, merupakan wilayah pedesaan dengan mayoritas penduduk bekerja di sektor pertanian, peternakan, dan buruh harian lepas. Kondisi sosial ekonomi masyarakatnya cukup beragam, dengan sebagian besar berada pada kategori ekonomi menengah ke bawah. Untuk membantu masyarakat kurang mampu, pemerintah menyalurkan Bantuan Langsung Tunai (BLT) yang bersumber dari Dana Desa sebagai bentuk dukungan ekonomi. Namun, dalam pelaksanaannya, masih sering ditemukan kendala dalam hal pendataan dan penentuan penerima bantuan yang tepat sasaran.

3.2 Pengumpulan Data

Data yang digunakan ini dalam penelitian ini data Penerima Bantuan Langsung Tunai (BLT) di desa Pesanggrahan. Untuk mendapatkan data dan informasi yang tepat mengenai penerima BLT di Desa Pesanggrahan, Pengumpulan data pada peneletian ini dilakukan melalui wawancara dan observasi di kantor desa pesanggrahan. Wawancara dilakukan dengan aparat desa guna mengumpulkan data terkait penerima bantuan. Serta melakukan Observasi langsung di kantor desa Pesanggrahan.

repository.ibrahimy.ac.id

3.3 Alur Proses

Alur proses yang merupakan gambaran prosedur yang akan digunakan untuk memecahkan masalah, secara rinci proses ini akan menjelaskan mengenai tahapan sebuah penelitian. Prosedur ini akan memudahkan penelitian untuk berjalan sesuai dengan yang diinginkan.

3.2.1 Identifikasi dan Analisis proses Bisnis

Pada bagian ini dilakukan identifikasi terkait dengan penelitian yang sedang di lakukan yaitu Implementasi Algoritma *Naive Bayes* klasifikasi Penerima Bantuan BLT Di Desa Pesanggrahan. Pada tahap ini menjelasakan alur proses data penerima bantuan BLT.

a. *Data preparation*
b. *Labeling data*
c. Metode *Naive bayes*

3.2.2 Analisis Proses Bisnis

Analisis Proses Bisnis ini terdiri dari beberapa penjelasan tentang proses bisnis yaitu *Data* Preparationdan metode Naive Bayes

a. *Data Preparation*

Data *Preparation* adalah proses multi-langkah yang mencakup penanganan data yang hilang, penskalaan, dan penambahan. Menyelesaikan langkah-langkah ini secara manual dapat memakan waktu dan meningkatkan kemungkinan kesalahan, terutama jika data harus diproses secara berulang-ulang atau dalam jumlah besar.

repository.ibrahimy.ac.id

Seluruh proses persiapan data dapat diselesaikan dengan lebih cepat, konsisten, dan dapat dilacak dengan pipeline otomatisasi[39].

b. Metode Naive Bayes

Berikut merupakan tahapan pengujian algoritma *Naive Bayes Clasifier* di bawah ini[40]:

1. Menghitung probabilitas dari Vj

$$P(\boldsymbol{vj}) = \frac{\text{doc j}}{\text{training}} \quad (3.1)$$

P(vj) adalah probabilitas kategori vj (kategori penerima BLT)

Docj adalah data testing

Pada data training dengan persamaan (3.1) di mana Vj merupakan Kategori Penerima BLT, yaitu V1 = layak, V2 = Tidak layak.

2. Menghitung probabilitas kata ai

$$P(\boldsymbol{ai}|\boldsymbol{vj}) = \frac{ni+1}{n+kosakata} \quad (3.2)$$

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

Pada *kategori* Vj dengan persamaan (3.2)

3. Model probabilitas NBC disimpan dan digunakan untuk tahap testing
4. Menghitung probabilitas tertinggi dari kategori penerima BLT yang diujikan (Vmap) dengan persamaan (2.1)

repository.ibrahimy.ac.id

5. Mencari nilai Vmap paling maksimum dan masukkan penerima BLT tersebut pada kategori dengan Vmap maskimum (3.3)

$$V_{map} = ar\ g\ max\ P(v_j) \prod i\ P\ (a_i|v_j\ ) \qquad (3.3)$$

V_{map} adalah kategori penerima BLT

arg max adalah probabilitas tertinggi dari semua kategori yang diujikan

c. *Confusion Matrix*

Confusion Matrix pada proses Implementasi Algoritma *Naïve Bayes* Klasifikasi Penerima Bantuan BLT. Di tunjukan pada tabel dibawah ini:

Tabel 3. 1 *Confusion Matrix* Penerima Bantuan BLT

		Skor Prediksi	
		Tidak Layak	**Layak**
Skor Prediksi	**Layak**	**TL**	**FL**
	Tidak Layak	**FT**	**TT**

Keterangan:

1. Nilai saat ini yang Layak dan diantisipasi secara akurat dikenal sebagai *True* Layak (TL).
2. Nilai saat ini yang Tidak Layak dan diantisipasi secara akurat dikenal sebagai True Tidak Layak (TT).
3. *False* Layak (FL) adalah nilai saat ini yang diproyeksikan Layak tetapi sebenarnyaTidak Layak.

repository.ibrahimy.ac.id

4. *False* Tidak Layak (FT) adalah hasil yang saat ini Layak tetapi diperkirakan Tidak Layak.

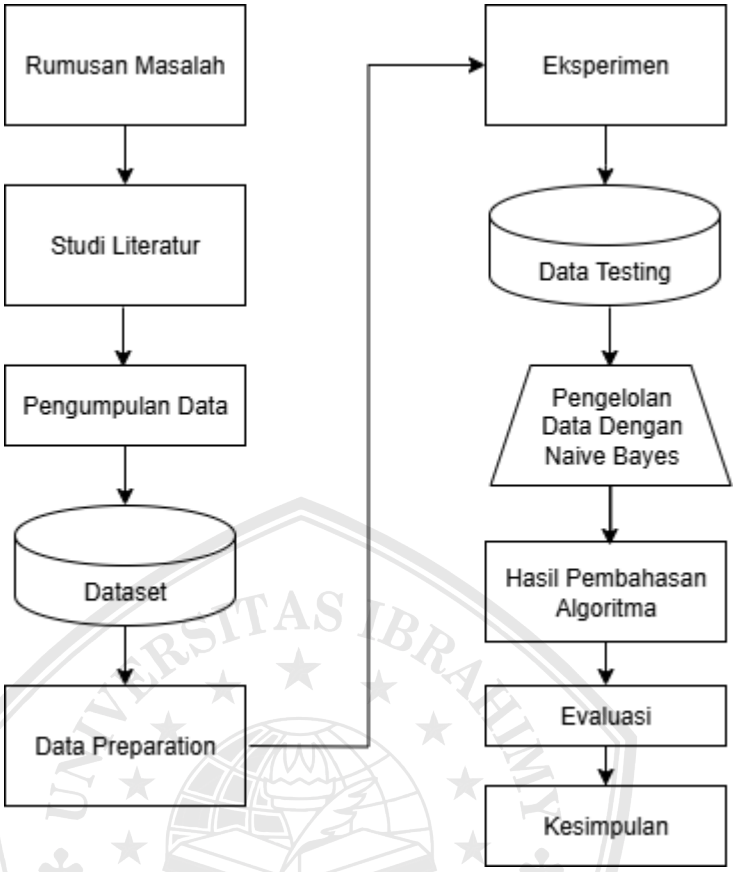
Berikut empat hasil perhitungan menggunakan *Confusion Matrix* untuk pengujian Sebagai berikut :

1. Menghitung nilai akurasi, atau proporsi kelas yang diprediksi secara akurat oleh model yang sudah di buat (2.2).
2. Menentukan nilai presisi, atau tingkat keakuratan antara informasi yang diminta pengguna dan respons sistem (2.3).
3. Menentukan nilai *Recall*, sering disebut sebagai sensitivitas, yang merupakan proporsi data yang diprediksi oleh program daripada kelas yang sebenarnya (2.4).
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* (2.5).

d. Alur Implementasi *Naïve Bayes*

Langkah-langkah yang dilakukan dalam klasifikasi penerima bantuan BLT bisa dilihat pada gambar 3.1 di bawah ini sebagai berikut:

repository.ibrahimy.ac.id



Gambar 3. 1 Implementasi Algoritma Naïve Bayes

3.2.3 Identifikasi dan Analisis Kebutuhan

a. Identifikasi Kebutuhan Fungsional

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

b. Analisis Kebutuhan Fungsional

Pada tahap ini data akan diolah menggunakan sebuah tahapan data yaitu tahap *Data Preparation*.

repository.ibrahimy.ac.id

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 3.2 merupakan tabel identifikasi alternatif solusi.

Tabel 3. 2 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 *ree*l 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

repository.ibrahimy.ac.id

b. Analisis Kelayakan Alternatif Solusi

Proses ini berfungsi untuk menganalisis sistem yang akan dibuat berdasarkan dokumen yang didapatkan sebelumnya. Tabel 3.3 menjelaskan identifikasi kelayakan alternatif.

Tabel 3. 3 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, menurut Kamus APICS, adalah serangkaian tindakan atau tahapan terencana yang digunakan untuk memproses bahan atau prosedur dari satu tahap proses ke tahap berikutnya[41]

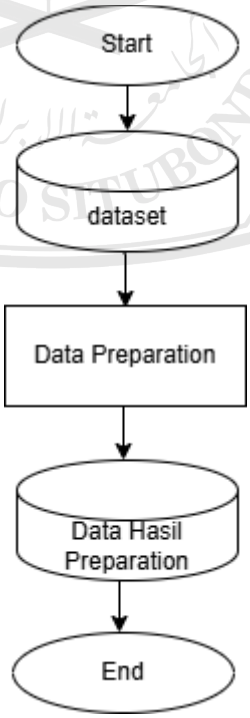
repository.ibrahimy.ac.id

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. Tahapan Data *Preparation*

Proses ini dilakukan untuk mengolah data menjadi sebuah dataset. Data diperoleh dari hasil Wawancara kepaada kepala desa menghasilkan *output* dataset berupa *file excel* lalu di *input*-kan dan siap dilakukan proses perhitungan dengan algoritma. Pada Gambar 3.2 menjelaskan tentang alur proses *Data Prepraration*.

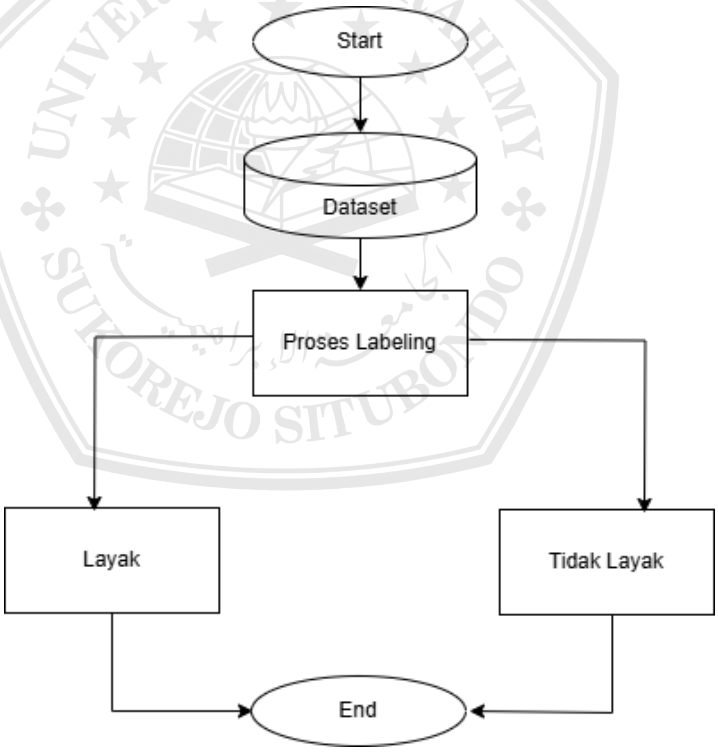


Gambar 3. 2 *Data preparation*

repository.ibrahimy.ac.id

2. *Labeling Data*

Labeling merupakan pemberian label pada data yang telah dikumpulkan[42]. Label pada penelitian ini berperan penting dalam mengkategorikan data penerima bantuan BLT dengan manual berikut kategori, kelurga miskin, kelurga miskin tidak terkait bantuan lain, rentan sakit dan lansia. label yang digunakan adalah label layak atau tidak layak untuk menerima bantun BLT berdasarkan kategori dan selanjutnya dilakukan proses *labeling* Pada Gambar 3.3 merupakan alur proses *labeling data*.

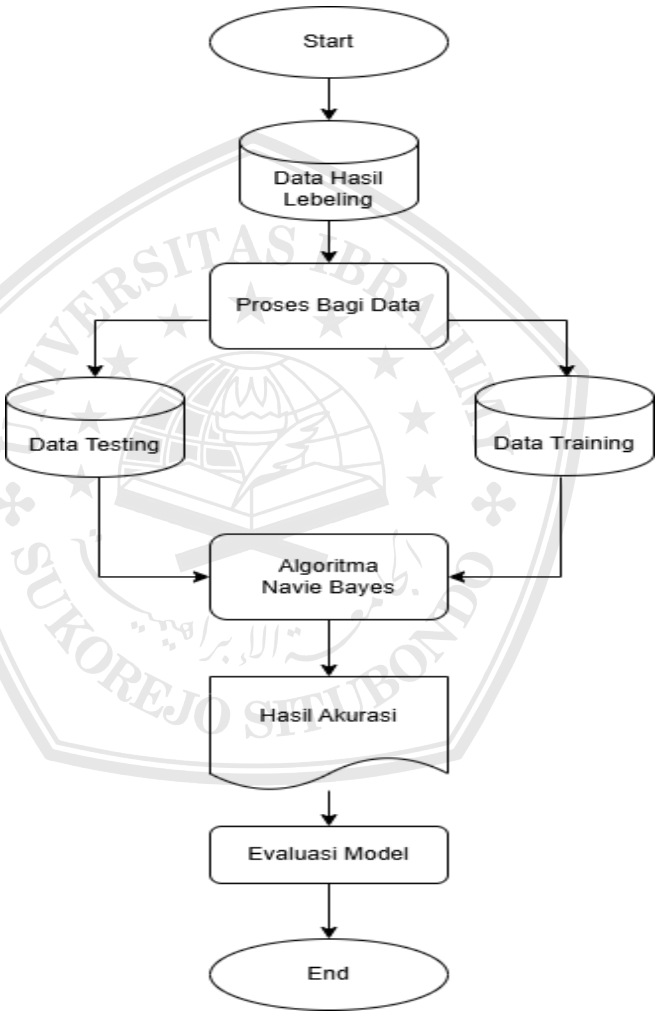


Gambar 3. 3 Alur Labeling

repository.ibrahimy.ac.id

3. Metode *Naive Bayes*.

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

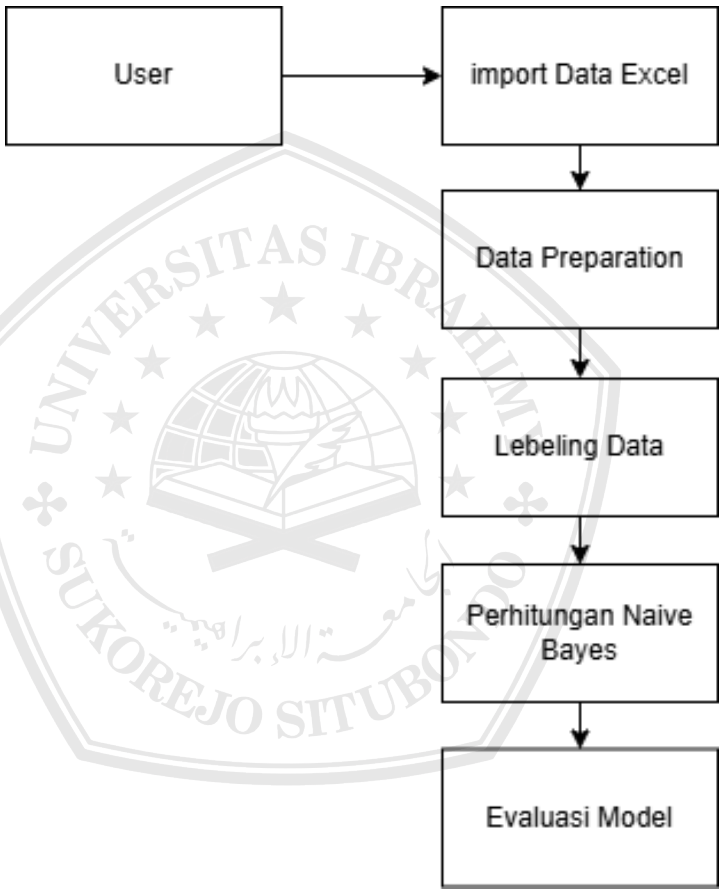


Gambar 3. 4 Alur Metode *Naive Bayes*

repository.ibrahimy.ac.id

b. Arsitektur Sistem

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



Gambar 3. 5 Arsitektur Sistem

repository.ibrahimy.ac.id

BAB IV

HASIL DAN PEMBAHASAN

4.1 Kebutuhan system

Sistem ini dapat dioperasikan menggunakan bahasa pemrograman Python dengan mengawasi sejumlah perangkat yang dapat berkontribusi terhadap keberhasilannya. Berikut adalah perankat yang digunakan:

1. *Software*

Software yang digunaka dalam mendukung sistem adalah sebagai berikut:

a. *Google Collab*

b. *Microsoft excel*

2. Hardware

a. Leptop *acer*

b. *Mouse* dan *Keyboard*

c. *Processor AMD* A4

d. *RAM 4GB*

4.2 Data Preparation

Dalam tahap *data preparation*, dilakukan pembersihan data untuk menjamin mutu serta konsistensi dataset yang akan dipakai pada proses analisis. Salah satu langkah yang ditempuh adalah menghapus data duplikat, yakni baris dengan nama yang tercatat lebih dari satu kali. Data ganda berpotensi menimbulkan bias, sehingga harus dihilangkan agar hasil analisis menjadi lebih valid. Selain itu, baris kosong juga dihapus karena ketidaklengkapan data dapat menghambat pengolahan maupun perhitungan algoritma. Setelah kedua prosedur

repository.ibrahimy.ac.id

ini dilakukan, dataset menjadi lebih rapi, terstruktur, dan siap digunakan pada tahap analisis selanjutnya.. Adapun tahapan *preparation* adalah sebagai berikut :

1. Menghapus data duplikat

Menemukan baris duplikat dalam kumpulan data adalah tahap awal. Hal ini dilakukan dengan menggunakan fungsi duplicated() dari pustaka pandas di Python. Setelah diidentifikasi, data duplikat dapat dihapus dengan fungsi drop_duplicates(). Secara default, fungsi ini akan menghapus semua baris duplikat dan hanya menyisakan satu baris data asli. Berikut ini adalah segmen program menghapus data duplikat sebagai berikut:

Segmen program ini digunakan untuk membaca data Excel dari penyimpanan Google Drive ke dalam environment Google Colab. Pada saat proses pembacaan, nilai yang dianggap kosong seperti 'NA', 'null', 'None', '-', dan '--' secara otomatis dikonversi menjadi NaN agar lebih mudah ditangani. Program kemudian menampilkan informasi jumlah baris dan kolom pada dataset untuk memastikan struktur data yang diperoleh sudah sesuai. Tahap berikutnya adalah menghapus data duplikat berdasarkan atribut NAMA dengan hanya mempertahankan kemunculan pertama. Langkah ini penting dilakukan untuk menjaga konsistensi data dan mencegah bias dalam analisis, karena entri ganda dapat memengaruhi hasil perhitungan maupun klasifikasi.pada Segmen program 4.1.

repository.ibrahimy.ac.id

Segmen Program 4. 1 Menghapus Data Duplikat

```
from google.colab import drive
drive.mount('/content/drive')
Mounted at /content/drive
import pandas as pd
file_path ="/content/drive/MyDrive/datapenduduk.xlsx"
df = pd.read_excel(file_path, na_values=['', ' ', 'NA',
'null', 'None', '-', '--'])
 df
num_rows, num_colums=df.shape
num_rows
#menghapus data yang duplikat
df = df.drop_duplicates(subset='NAMA', keep='first')
df
```

Berikut ini adalah data yang sebelum di hapus duplikat. Berikut ini data sebelum di hapus duplikat ada 153 data pada gambar 4.1:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 153 entries, 0 to 152
Data columns (total 7 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   NO                      153 non-null    int64
 1   NAMA                    153 non-null    object
 2   JENIS KELAMIN           153 non-null    object
 3   USIA                    153 non-null    object
 4   PEKERJAAN               150 non-null    object
 5   TIDAK TERKAIT BANTUAN LAIN  153 non-null    object
 6   DUSUN                   153 non-null    object
dtypes: int64(1), object(6)
memory usage: 8.5+ KB
```

Gambar 4. 1 Menghapus Data Duplikat

Berikut ini adalah data yang sudah di hapus duplikat pada kolom nama yang berduplikat.Berikut ini data sudah di hapus duplikat pada kolom nama data yang sudah di hapus tinggal 150 karna ada 3 nama yang berduplikat. Berikut ini gambar hasil dari menghapus data yang berduplikat:

repository.ibrahimy.ac.id

```
<class 'pandas.core.frame.DataFrame'>
Index: 150 entries, 0 to 151
Data columns (total 7 columns):
 #   Column                      Non-Null Count  Dtype
---  ------                      --------------  -----
 0   NO                          150 non-null    int64
 1   NAMA                        150 non-null    object
 2   JENIS KELAMIN               150 non-null    object
 3   USIA                        150 non-null    object
 4   PEKERJAAN                   147 non-null    object
 5   TIDAK TERKAIT BANTUAN LAIN  150 non-null    object
 6   DUSUN                       150 non-null    object
dtypes: int64(1), object(6)
memory usage: 9.4+ KB
```

Gambar 4. 2 Menghapus Data Duplikat

2. Menghapus baris kosong

Penghapusan baris yang berisi nilai null sepenuhnya merupakan salah satu prosedur yang dilakukan. Baris yang tidak memiliki data lengkap berpotensi menyebabkan kesalahan dalam proses analisis maupun pelatihan model. Oleh karena itu, semua baris yang mengandung nilai kosong pada salah satu kolom diidentifikasi dan dihapus dari dataset Berikut ini adalah segmen program untuk menghapus baris data yang kosong atau bersifat NaN:

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

repository.ibrahimy.ac.id

Segmen Program 4. 2 Menghapus Baris Yosong

```
#cek data nan
df = df.apply(lambda x: x.str.strip() if x.dtype ==
"object" else x)
df.replace(['', ' ', '-', '--'], pd.NA, inplace=True)
df[df.isnull().any(axis=1)]

#menghapus baris yang kosong
df = df.dropna(how='any')

#menyimpan data hasil data preparation
df_hilangkan.to_excel('datapenduduk.xlsx',index=False)
```

Segmen program ini digunakan untuk melakukan pembersihan data dengan cara mengidentifikasi dan menghapus nilai kosong pada dataset. Langkah pertama dilakukan dengan menghapus spasi pada setiap entri bertipe string agar tidak terbaca sebagai nilai kosong. Selanjutnya, simbol tertentu diganti dengan nilai NaN sehingga lebih mudah dikenali sebagai data hilang. Setelah data kosong terdeteksi, baris yang mengandung nilai NaN dihapus menggunakan fungsi dropna.

Sebelum dilakukan penghapusan baris kosong data terdiri dari 150 data.berikut ini data sebelum dihapus baris kolom kosong pada gambar 4.3:

repository.ibrahimy.ac.id

```
<class 'pandas.core.frame.DataFrame'>
Index: 150 entries, 0 to 151
Data columns (total 7 columns):
 #   Column                      Non-Null Count  Dtype
---  ------                      --------------  -----
 0   NO                          150 non-null    int64
 1   NAMA                        150 non-null    object
 2   JENIS KELAMIN               150 non-null    object
 3   USIA                        150 non-null    object
 4   PEKERJAAN                   147 non-null    object
 5   TIDAK TERKAIT BANTUAN LAIN  150 non-null    object
 6   DUSUN                       150 non-null    object
dtypes: int64(1), object(6)
memory usage: 9.4+ KB
```

Gambar 4. 3 Menghapus Baris Kosong

Setelah dilakukan penghapusan hapus baris yang kosong data menjadi 147 data karna ada 3 baris kosong . Berikut hasil dari penghapusan baris yang kosong pada gambar 4.4:

```
<class 'pandas.core.frame.DataFrame'>
Index: 147 entries, 0 to 151
Data columns (total 7 columns):
 #   Column                      Non-Null Count  Dtype
---  ------                      --------------  -----
 0   NO                          147 non-null    int64
 1   NAMA                        147 non-null    object
 2   JENIS KELAMIN               147 non-null    object
 3   USIA                        147 non-null    object
 4   PEKERJAAN                   147 non-null    object
 5   TIDAK TERKAIT BANTUAN LAIN  147 non-null    object
 6   DUSUN                       147 non-null    object
dtypes: int64(1), object(6)
memory usage: 9.2+ KB
```

Gambar 4. 4 Menghapus Baris Kosong

repository.ibrahimy.ac.id

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 Layak atau Tidak Layak sesuai dengan kriteria yang telah ditentukan. Hasil lebeling data manual ditentukan pada Tabel 4.1.

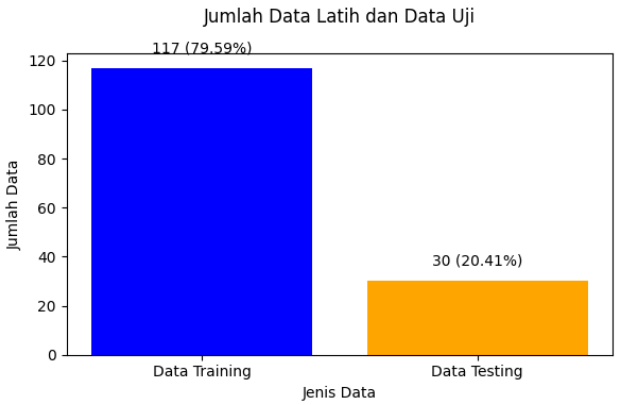
Tabel 4. 1 Labeling Data

NAMA	JENIS KELAMIN	USIA	PEKERJAAN	TIDAK TERKAIT BANTUAN LAIN	DUSUN	STATUS
NING SRI WULANDARI	P	LANSIA	BURUH TANI	TIDAK	GEDANG	LAYAK
YAYUK ERNAWATI	P	LANSIA	BURUH TANI	TIDAK	GEDANG	LAYAK
LASTRI	P	LANSIA	BURUH TANI	TIDAK	GEDANG	LAYAK
SIU	P	LANSIA	BURUH TANI	TIDAK	GEDANG	LAYAK
DIYATI	P	DEWASA	PETANI	TIDAK	GEDANG	TIDAK LAYAK
PANDI	L	DEWASA	PETANI	TIDAK	ACEM	TIDAK LAYAK
SURIJA	P	LANSIA	PETANI	TIDAK	GEDANG	TIDAK LAYAK

4.4 Split Data Training dan Data Testing

Pada tahap *Split Data*, dataset dibagi menjadi dua bagian, yaitu *data training* dan *data testing*. Data training digunakan untuk melatih algoritma agar dapat mengenali pola dari sekumpulan data, sedangkan data testing berfungsi untuk menguji performa model yang telah dibangun. Proses pemisahan ini bertujuan agar hasil evaluasi lebih objektif, sehingga model tidak hanya mampu bekerja baik pada data latih, tetapi juga pada data baru yang belum pernah dipelajari. Ditunjukkan pada gambar 4.5.

repository.ibrahimy.ac.id



Gambar 4. 5 Split Data Training dan Data Testing

Pada proses split data, data penerima BLT dibagi agar model dapat dilatih menggunakan sebagian data, dan diuji performanya. Data BLT yang digunakan, sebanyak 117 data (79,59%) digunakan sebagai data *training*, dan 30 data (20,41%) digunakan sebagai data *testing*. Segmen program *split* data ditunjukkan pada Segmen Program 4.3.

Segmen Program 4. 3 Split Data

```
from sklearn.model_selection import train_test_split
# Bagi data menjadi data pelatihan dan pengujian
x_train, x_test, y_train, y_test =
train_test_split(df_data[['USIA', 'PEKERJAAN', 'TIDAK TERKAIT
BANTUAN LAIN']],
                                df_data['STATUS'],test_size=0.2,
random_state=42)
#simpan data latih ke file
x_text = x_train['USIA'].astype(str) + ' ' +
x_train['PEKERJAAN'].astype(str)+ ' ' + x_train['TIDAK TERKAIT
BANTUAN LAIN'].astype(str)
train_set = pd.DataFrame({'text': x_text, 'STATUS': y_train})
train_set.to_csv('train_set.csv', index=False)

#simpan data uji ke file
x_text = x_test['USIA'].astype(str) + ' ' +
x_test['PEKERJAAN'].astype(str)+ ' ' + x_test['TIDAK TERKAIT
BANTUAN LAIN'].astype(str)
test_set = pd.DataFrame({'text': x_text, 'STATUS': y_train})
test_set.to_csv('test_set.csv', index=False)

# Menampilkan informasi jumlah data
print(f'Jumlah Data Latih: {len(x_train)}')
```

repository.ibrahimy.ac.id

Segmen Program 4. 3 Split Data Lanjutan

```
print(f'Jumlah Data Uji: {len(x_test)}')

# Membuat plot
plt.figure(figsize=(6, 4))
bars = plt.bar(['Data Training', 'Data Testing'], [train_size,
test_size], color=['blue', 'orange'])

# Menambahkan label untuk setiap bar (jumlah + persentase)
for bar in bars:
    height = bar.get_height()
    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()
```

Segmen program ini digunakan untuk melakukan pemisahan dataset menjadi dua bagian, yaitu data training dan *data testing* dengan bantuan fungsi `train_test_split` dari pustaka *scikit-learn*. Atribut yang digunakan dalam proses pemisahan meliputi usia, pekerjaan, serta status tidak terkait bantuan lain, sedangkan variabel target adalah status penerima.

Setelah data terbagi, setiap baris pada data latih dan data uji digabungkan ke dalam bentuk teks dan disimpan dalam format *CSV* agar mudah diakses kembali. Program ini juga menampilkan jumlah data latih dan data uji secara langsung di konsol. Untuk memberikan gambaran visual, dibuat diagram batang yang menunjukkan perbandingan jumlah data training dan testing, lengkap dengan persentase dari keseluruhan dataset. Visualisasi ini membantu memastikan bahwa proses pemisahan data telah dilakukan secara proporsional sesuai parameter yang ditentukan.

Ibrahimy LIBRARY
Ibrahimy Repository
Universitas Ibrahimy
repository.ibrahimy.ac.id

4.5 Metode Naïve Bayes

Data Penerima bantuan BLT yang sudah dilakukan proses data *preparation*, *lebeling* data dan *split* data dibagi menjadi dua yaitu data *training* dan data *testing*. Langkah awal dalam metode *Naïve Bayes* adalah mengubah data kategorikal menjadi numerik berupa token frekuensi kata. Tujuannya adalah untuk menganalisis pola kata yang muncul dominan pada masing-masing kategori klasifikasi. hasilnya menjadi matriks frekuensi token untuk menghitung probabilitas prior.

Selanjutnya perhitungan probabilitas prior untuk setiap kelas dan dihitung jumlah data setiap kelas terhadap total data untuk menentukan hasil prediksi probabilitas prior (3.1). langkah selanjutnya adalah menghitung probabilitas kondisonal dari setiap token terhadap kelasnya. Probabilitas kondisonal ini menunjukkan seberapa besar kemungkinan sebuah kata (fitur) untuk muncul di kelas tertentu. Jika ada probabilitas nol maka menggunakan teknik Laplace smoothing untuk menambahkan nilai 1 untuk setiap frekuensi token (3.2).

Untuk membagi data menjadi data training dan data testing menggunakan train_test_split dengan raiso perbandingan 80:20. Fungsi fit() untuk model GaussianNB, digunakan untuk melakukan proses pelatihan, yang sesuai untuk data yang terdistribusi secara normal menggunakan algoritma TF-IDF (*Term Frequency - Inverse Document Frequency*). Setelah pelatihan, model digunakan untuk prediksi data testing dan hasil prediksi untuk memudahkan interpretasi hasil klasifikasi, dan untuk menentukan seberapa baik model dapat mengklasifikasikan

repository.ibrahimy.ac.id

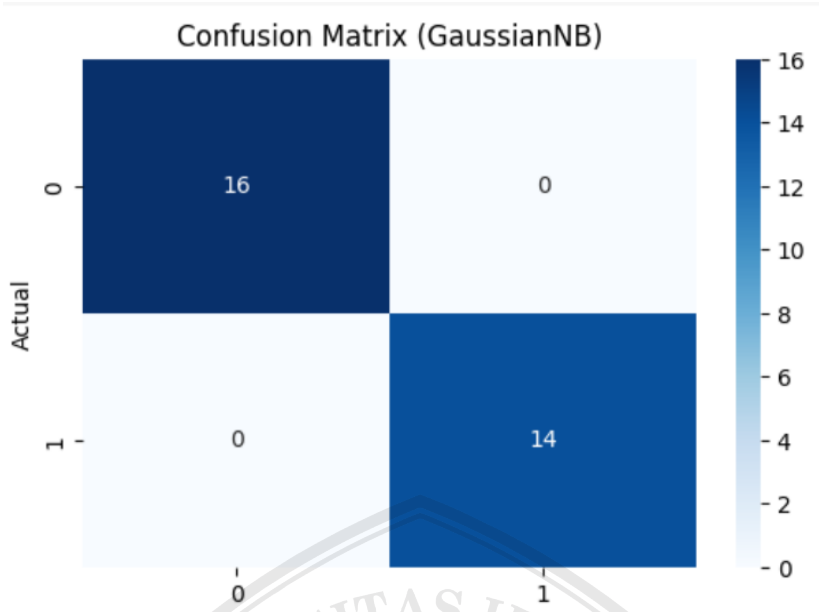
data. Dibawah ini gambar hasil klasifikasi prediksi *Naïve Bayes* menggunakan *GaussianNB* Pada gambar 4.6 .

	USIA	PEKERJAAN	TIDAK TERKAIT BANTUAN LAIN	Actual	Predicted
125	DEWASA	IBU RUMAH TANGGA	TIDAK	tidak layak	tidak layak
51	LANSIA	BURUH TANI	TIDAK	layak	layak
138	LANSIA	IBU RUMAH TANGGA	IYA	tidak layak	tidak layak
19	LANSIA	BURUH TANI	TIDAK	layak	layak
104	LANSIA	TIDAK KERJA	IYA	tidak layak	tidak layak
12	LANSIA	BURUH TANI	TIDAK	layak	layak
76	LANSIA	BURUH TANI	TIDAK	layak	layak
31	LANSIA	PETANI	IYA	tidak layak	tidak layak
81	LANSIA	PETANI	TIDAK	tidak layak	tidak layak
9	LANSIA	BURUH TANI	TIDAK	layak	layak
26	LANSIA	BURUH TANI	TIDAK	layak	layak
96	DEWASA	IBU RUMAH TANGGA	TIDAK	tidak layak	tidak layak
143	DEWASA	IBU RUMAH TANGGA	TIDAK	tidak layak	tidak layak
67	LANSIA	BURUH TANI	TIDAK	layak	layak
134	DEWASA	IBU RUMAH TANGGA	TIDAK	tidak layak	tidak layak

Gambar 4. 6 Prediksi Naïve bayes GaussianNB

Setelah pelatihan *GaussianNB* dan dihasilkan prediksi, untuk menentukan evaluasi seberapa baik dan akurat model klasifikasi dengan menentukan perbandingan hasil prediksi menggunakan *confusion matrix* pada gambar 4.8. Evaluasi *precision, recall, f1-score*, untuk menilai akurasi prediksi model terhadap setiap kelas menggunakan *Classification Report*. Untuk menguji tingkat akurasi, ketepatan prediksi model dan jumlah seluruh prediksi dibandingkan total seluruh data menggunakan *Accuracy* pada gambar 4.7. dan 4.8.

repository.ibrahimy.ac.id



Gambar 4. 7 *confusion matrix*

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

       layak       1.00      1.00      1.00        16
 tidak layak       1.00      1.00      1.00        14

    accuracy                           1.00        30
   macro avg       1.00      1.00      1.00        30
weighted avg       1.00      1.00      1.00        30

======================================
Accuracy (GaussianNB): 1.0000
======================================
```

Gambar 4. 8 Classification Report Dan Accuracy

Langkah selanjutnya menghitung *Confusion matrix*, untuk memproleh hasil prediksi yang sangat baik, di mana seluruh data uji berhasil diklasifikasikan dengan benar. Terdapat 16 data yang tergolong kategori layak berhasil diprediksi dengan tepat, serta 14 data dalam kategori tidak layak juga diprediksi dengan akurat, tanpa adanya kesalahan klasifikasi.

repository.ibrahimy.ac.id

Dengan demikian, nilai *True Positive* (TP) dan *True Negative* (TN) masing-masing adalah 16 dan 14, sementara nilai *False Positive* (FP) dan *False Negative* (FN) adalah 0, Nilai hasil perhitungan akurasi sebesar 1.00 atau 100% dengan persamaan (2.2).

$$Accuracy = \frac{16 + 14}{16 + 14 + 0 + 0} = \frac{30}{30} = 1 \text{ X } 100\% = 1.00$$

Dengan demikian, dapat disimpulkan bahwa nilai akurasi 100% ini menjadi bukti bahwa model *Naïve Bayes GaussianNB* mampu mengklasifikasikan data dengan tingkat ketelitian yang sangat tinggi. Hal ini sekaligus memperlihatkan bahwa metode yang dipakai sesuai dengan karakteristik yang sudah di tentukan.

Berikut ini perhitungan *Classification Report* pada *precision, recall, f1-score* setiap kelas:

Precision dihitung berdasarkan 16 data yang benar-benar layak dan tidak terdapat satupun yang salah diklasifikasikan sebagai layak padahal sebenarnya tidak layak, sehingga *precision*-nya adalah 16 dibagi 16, menghasilkan nilai 1.00. Dan kelas *precision* tidak layak, terdapat 14 data yang seluruhnya berhasil diprediksi dengan benar tanpa kesalahan prediksi positif, sehingga *precision*-nya menghasilkan nilai 1.00 dengan persamaan(2.3).

$$Precision\ Layak\ = \frac{16}{16 + 0} = 1.00$$

$$Precision \text{ Tidak Layak } = \frac{14}{14 + 0} = 1.00$$

repository.ibrahimy.ac.id

Recall untuk kelas Layak terdiri 16 data yang benar-benar termasuk kategori Layak, di mana seluruhnya berhasil diprediksi dengan tepat tanpa ada kesalahan klasifikasi menjadi Tidak Layak atau FN 0. Hal ini menghasilkan nilai *recall* sebesar 16 dibagi 16 + 0, menghasilkan nilai 1.00. sedangkan, untuk kelas Tidak Layak, terdapat 14 data yang seluruhnya juga berhasil diprediksi dengan benar tanpa satupun yang salah diklasifikasikan sebagai Layak, sehingga *recall*-nya adalah 14 dibagi 14 + 0, menghasilkan nilai 1.00 (2.4).

$$Recall \text{ Layak } = \frac{16}{16+0} = 1.00$$

$$Recall \text{ Tidak Layak } = \frac{14}{14+0} = 1.00$$

Perhitungan *F1-Score* dilakukan dengan cara mengombinasikan nilai *precision* dan *recall*, sehingga metrik ini mampu memberikan gambaran yang lebih menyeluruh mengenai kinerja model. Tidak seperti akurasi yang hanya menunjukkan persentase prediksi benar secara umum, *F1-Score* lebih menekankan keseimbangan antara ketepatan prediksi positif dan kemampuan model dalam mengenali seluruh data *positif* yang ada. Dengan demikian, ukuran ini sering dianggap lebih adil ketika dataset mengandung distribusi kelas yang tidak seimbang.

Berdasarkan hasil pengujian, pada kategori Layak, nilai *precision* tercatat 1.00 dan *recall* juga menunjukkan nilai 1.00. Hal yang sama terlihat pada kategori Tidak Layak, di mana kedua metrik tersebut sama-sama memperoleh nilai sempurna 1.00. Kondisi ini memperlihatkan bahwa algoritma yang digunakan

Ibrahimy LIBRARY
Ibrahimy Repository
Universitas Ibrahimy
repository.ibrahimy.ac.id

mampu mengklasifikasikan data dengan sangat baik, tanpa menghasilkan kesalahan prediksi baik pada kelas *positif* maupun *negative* dengan persamaan(2.5).

$$f1 - score \text{ Layak } = 2X\frac{1.00\ X\ 1.00}{1.00 + 1.00} = 2X\frac{1.00}{2.00} = 1.00$$

$$f1 - score \text{ Tidak Layak } = 2X\frac{1.00\ X\ 1.00}{1.00 + 1.00} = 2X\frac{1.00}{2.00} = 1.00$$

Hasil *precision* dan *recall* tersebut digabungkan melalui rumus *F1-Score*, hasil akhir yang diperoleh adalah 1.00. Nilai maksimal ini menjadi bukti bahwa model memiliki performa optimal, sebab mampu menjaga keseimbangan antara ketepatan prediksi dan kelengkapan deteksi. Dengan demikian, dapat disimpulkan bahwa metode yang diterapkan pada penelitian ini bekerja sangat efektif dalam menyelesaikan permasalahan klasifikasi yang diuji.

Pada tabel dibawah ini merupakan Segmen program *GaussianNB* berfungsi untuk melatih model dengan data *training*, kemudian melakukan prediksi pada data testing. Segmen program ini memanfaatkan pustaka *scikit-learn* untuk menerapkan algoritma *Naïve Bayes GaussianNB.* Data terlebih dahulu diolah menggunakan TF-IDF Vectorizer sehingga setiap atribut dapat direpresentasikan dalam bentuk vektor numerik. Selanjutnya, data dibagi menjadi *data training* dan data *testing* menggunakan fungsi `train_test_split`.

Model *GaussianNB* kemudian dilatih dengan data training melalui perintah `fit`, dan hasilnya digunakan untuk melakukan prediksi pada data testing. Kinerja model dievaluasi dengan menghitung *confusion matrix, classification report*, serta

repository.ibrahimy.ac.id

akurasi. Untuk mempermudah interpretasi, *confusion matrix* divisualisasikan dalam bentuk *heatmap* menggunakan pustaka Seaborn pada Segmen Program 4.4.

Segmen Program 4. 4 Algoritma Naïve Bayes GaussianNB

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB
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[['USIA', 'PEKERJAAN', 'TIDAK TERKAIT
BANTUAN LAIN']].astype(str).agg(' '.join, axis=1)

x = tfidf.fit_transform(text_data).toarray()
y = df_data['STATUS']
# 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)

# Evaluate GaussianNB
conf_matrix_gnb = confusion_matrix(y_test, y_pred_gnb,
labels=gnb.classes_)
class_report_gnb = classification_report(y_test, y_pred_gnb)
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 for GaussianNB
plt.figure(figsize=(6, 4))
sns.heatmap(conf_matrix_gnb, annot=True, fmt='d',
```

repository.ibrahimy.ac.id

Segmen Program 4. 4 Algoritma Naïve Bayes GaussianNB Lanjutan

```
cmap='Blues')
plt.title('Confusion Matrix (GaussianNB)')
plt.xlabel('Predicted')
plt.ylabel('Actual')
plt.show()
results_gnb = df_data.loc[y_test.index, ['USIA', 'PEKERJAAN',
'TIDAK TERKAIT BANTUAN LAIN']].copy()
results_gnb['Actual'] = y_test.values
results_gnb['Predicted'] = y_pred_gnb
results_gnb.to_csv('Hasil_pred_GaussianNB.csv',
encoding='utf8', index=False)
results_gnb.head(30)
```

Ibrahimy LIBRARY
Ibrahimy Repository
Universitas Ibrahimy
repository.ibrahimy.ac.id

BAB V

PENUTUP

5.1 Kesimpulan

Penelitian ini menerapkan Implementasi algoritma klasifikasi *Naïve Bayes* untuk memudahkan proses klasifikasi calon penerima Bantuan Langsung Tunai (BLT) di Desa Pesanggrahan. Dengan menggunakan data dari observasi, wawancara, dan dokumen yang diolah melalui tahap data *preparation*, *labeling*, serta pembagian *training* dan *testing*, algoritma ini mampu melakukan prediksi kelayakan penerima Bantuan Langsung Tunai lebih objektif. Penggunaan bahasa pemrograman *Python* dikombinasikan dengan metode *Naïve Bayes Gaussian* memungkinkan pengembangan model yang mampu mengklasifikasikan data dengan akurat sesuai kriteria yang ditetapkan oleh pemerintah desa, yaitu keluarga miskin, tidak terkait bantuan lain, keluarga yang berisiko terkena penyakit, dan lansia.

Hasil dari pengujiam menunjukkan performa model yang sangat baik, dengan dengan menggunakan *confusion matrix* menunjukkan bahwa semua data uji dapat diprediksi tanpa kesalahan klasifikasi, sehingga menghasilkan nilai *acuracy*, *precision*, *recall*, *f1-score* masing-masing sebesar 100%.

Berdasarkan hasil pengujian tersebut, klasifikasi penerimna Bantuan Langsung Tunai menggunakan Algoritma *Naïve Bayes* menjadi solusi untuk menyelesaikan permasalahan dan mengurangi kesalahan yang dilakukan secara manual Dengan adanya algoritma *Naïve Bayes* dapat meningkatkan akurasi prediksi dalam proses seleksi penerima bantuan BLT tepat sasaran berdasarkan

Ibrahimy LIBRARY
Ibrahimy Repository
Universitas Ibrahimy
repository.ibrahimy.ac.id

kreteria dan pengelolaan data secara *real-time* agar dapat dimanfaatkan secara berkelanjutan oleh pemerintah desa.

5.2 Saran

Klasifikasi Penerima Bantuan Langsung Tunai menggunakan Algoritma *Naïve Bayes* akurasi yang sangat baik. Oleh karena itu pada penelitian selanjutnya disarankan menggunakan algoritma klasifikasi seperti *Support Vector Machine* (SVM), *Random Forest* atau algoritma opti seperti *Particle Swarm Optimiztion* (PSO), *Genetic Algorithm* (GA), sebagai perbandingan hasil akurasi pada masing algoritma sesuai dengan karakteristik data kebutuhan desa.

Ibrahimy LIBRARY
Ibrahimy Repository
Universitas Ibrahimy
repository.ibrahimy.ac.id

DAFTAR PUSTAKA

[1] M. . Hanhan Hanafiah Solihin, S.Kom., M.Kom., Firman Noor Hasan, S.Kom., MTI, Rr. Isni Anisah Puspowati H, ST., M.Kom., Ivana Lucia Kharisma, M.Kom., Arif Himawan, S.Kom., M.M., M.Eng., Rangga Satria Perdana, S.Si., M.Kom., Dr. Ir. Dolly Indra, M.M.SI., MTA, G, *Konsep Sistem Informasi di Era Digital*. Kaizen Media Publishing, 2024.

[2] M. T. Muhammad Arhami, S.Si., M.Kom., Muhammad Nasir, S.T., *Data Mining - Algoritma dan Implementasi*. Andi Offset, 2020.

[3] M. H. I. Dr. Moh. Mufid, Lc., *Kaidah Fikih Ekonomi Dan Keuangan Kontemporer*. :Prenada Media, 2021.

[4] Helmi Yusup, L. Lukman, and G. Gusnawaty, "Pengaruh Nama Merek terhadap Citra pada Kuliner Lokal di Kota Makassar," *J. Onoma Pendidikan, Bahasa, dan Sastra*, vol. 10, no. 1, pp. 284–293, 2024, doi: 10.30605/onoma.v10i1.3206.

[5] N. A. Kilo, M. R. Katili, and I. K. Hasan, "Research in the Mathematical and Natural Sciences Perbandingan Metode K-Means dan K-Medoids Dengan Validitas Davies-Bouldin Indeks , Dunn Indeks dan Indeks Connectivity Pada Pengelompokkan Masyarakat Penerima Bantuan Langsung Tunai," vol. 4, no. 1, pp. 8–15, 2025, doi: 10.55657/rmns.v4i1.190.

[6] N. A. Setyaningrum *et al.*, "Sistem penentu penerimaan bantuan sosial (blt-dd) untuk keluarga kurang mampu," pp. 75–82, 2023.

[7] J. S. Informasi, T. Informasi, W. Dari, and A. Y. Sari, "Implementasi Data Mining Dengan Naïve Bayes Untuk Prediksi Penerima Bantuan Langsung Tunai (Blt) Warga Desa Xyz," vol. 2, no. 2, pp. 29–40, 2023.

[8] S. E. P. Rahayu Mayang Sari, Virdyra Tasril, Sri Wahyuni, *Klasifikasi Forecasting Mengunakan Algoritma Naive Bayes*. Serasi Media Teknologi, 2024.

[9] R. N. S. F. Ahmad Syafrizal Huda, Rolly Maulana Awangga, *Prediksi Penerimaan Pegawai Baru Dengan Metode Naive Bayes*. kreatif, 2020.

[10] J. B. Berlian, K. Tengah, J. B. Berlian, and K. Tengah, "Penerapan Metode Naive Bayes Dalam Mengklasifikasi Penerima BLT Pada Desa Pelangsian," vol. 14, no. 2, pp. 64–70, 2022.

[11] "Perkades BLT 2023."

[12] N. N. Saragih and R. Kurniawan, "(Journal of Computer Engineering, System and Science)," *J. Comput. Eng. Syst. Sci.*, vol. 10, no. 1, pp. 299–

repository.ibrahimy.ac.id

311, 2025.

[13] S. H. Eni Puji Estuti, Wachidah Fauziyanti, *Analisis Deskriptif dan Kuantitatif Produktivitas Garam Indonesia Studi Kasus pada Petani Garam Kabupaten Pati*. Penerbit NEM, 2021.

[14] M. M. Evellin Dewi Lusiana, *ANOVA untuk Penelitian Eksperimen Teori dan Praktik dengan R*. Universitas Brawijaya Press, 2021.

[15] M. K. L. di T. · Prof. Dr. Fahmi Rizal, M.Pd., M.T., Dr. Muhammad Ihsan, *Metodologi Penelitian Kuantitatif Pendidikan Kejuruan*. Merdeka Kreasi Group.

[16] M. S. Dr. Yusuf Tojiri, M.M., Hari Setia Putra, S.E., M.Si., Dr. Nur Faliza, S.E., *Dasar Metodologi Penelitian Teori, Desain, dan Analisis Data*. Takaza Innovatix Labs, 2023.

[17] M. P. C, *ANEKA INOVASI PEMBELAJARAN DARI STUDI KEPUSTAKAAN*. :Uwais Inspirasi Indonesia, 2024.

[18] M. P. Dr. Agus Salam, *METODE PENELITIAN KUALITATIF*. CV. AZKA PUSTAKA, 2023.

[19] Muhamad Adhi Mukti, W. Wiyanto, and Ismasari Nawangsih, "Implementasi Algoritma Naive Bayes Untuk Prediksi Calon Penerima Blt," *J. Inform. Teknol. dan Sains*, vol. 5, no. 3, pp. 383–392, 2023, doi: 10.51401/jinteks.v5i3.3106.

[20] J. Tugas, A. Manajemen, and K. Akuntansi, "IMPLEMENTASI ALGORITMA NAÏVE BAYES UNTUK PREDIKSI PENERIMA BANTUAN SOSIAL DI DESA CIGAYAM Dede Hoeriah , Bani Nurhakim , Sandy Eka Permana , Willy Prihartono ," vol. 4, no. 1, pp. 52–58, 2024.

[21] A. I Gusti Ayu Purnamawati, S.E., M.Si., *Akuntansi dan Implementasinya dalam Koperasi dan UMKM - Rajawali Pers*. :PT. RajaGrafindo Persada, 2021.

[22] I. Anggraeni, "Pengertian Implementasi dan PendapaT Ahli," *J. Chem. Inf. Model.*, vol. 53, no. 9, pp. 16–36, 2019.

[23] N. N. L. E. Z. Khairunnisa, Nurhadi Nurhadi, Andriyan Rizki Jatmiko, Legito Legito, Erwin Ardias Saputra, Fizar syafa'at, Dewi Fatmarani Surianto, Rita Komalasari, Iqbal Ramadhani Mukhlis, Sulistyowati Sulistyowati, Teguh Ansyor Lorosae, *BUKU AJAR LOGIKA & ALGORITMA*. PT. Sonpedia Publishing Indonesia, 2023.

[24] R. N. S. F. Ayu Lestari Nasution, *Klasifikasi Kondisi Peralatan Elektronik*

repository.ibrahimy.ac.id

Metode Gaussian Naïve Bayes. Penerbit Buku Pedia, 2023.

[25] H. Aulia, N. Syifa, A. Nugroho, and R. Firliana, "Perbandingan Algoritma Naïve Bayes Classifier Dan K-Nearest Neighbors Untuk Analisis Sentimen Covid-19 Di Twitter," no. 6, 2023.

[26] M. Z. Sarwani, M. Khoiron, and M. Udin, "Optimization of the Naïve Bayes Classifier Algorithm Using Cost-Sensitive Learning to Detect Lung Diseases with an Imbalanced Dataset," *J. Artif. Intell. Softw. Eng.*, vol. 5, no. 1, p. 332, 2025, doi: 10.30811/jaise.v5i1.6474.

[27] A. Y. M. Sebastianus Adi Santoso Mola, Shanty Novinia Ratu Djawa, *Text Mining Analisis Sentimen dengan Naïve Bayes*. :Kaizen Media Publishing, 2025.

[28] S. N. Fitriah, W. Rosita, and R. Rohmaniyah, "Pemanfaatan Sistem Klasifikasi dan Selving Bahan Pustaka: Upaya Memenuhi Temu Kembali Informasi di Perpustakaan Politeknik Negeri Sriwijaya," *J. El-Pustaka*, vol. 3, no. 2, pp. 83–105, 2022, doi: 10.24042/el-pustaka.v3i2.13796.

[29] M. H. I. Dr. Moh. Mufid, Lc., *Kaidah Fikih Ekonomi Dan Keuangan Kontemporer Pendekatan Tematis dan Praktis. Edisi kedua*. Prenada Media, 2021.

[30] Yuli Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4 . 5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database (KDD) . Jurnal Edik Informatika," *J. Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2019.

[31] R. S. Wahono, *Data Mining Data mining*, vol. 2, no. January 2013. 2023. [Online]. Available: https://www.cambridge.org/core/product/identifier/CBO9781139058452A007/type/book_part

[32] A. Sutanti, M. Komaruddin, P. Damayanti, and P. U. Studi Sistem Informasi Metro, "Rancang Bangun Aplikasi Perpustakaan Keliling Menggunakan Pendekatan Terstruktur," *J. Ilm. Komput. dan Inform.*, vol. 9, no. 1, 2022.

[33] E. S. Mulyadi Rusli, *Algoritma dan Pemrograman*. :Penerbit P4I, 2023.

[34] R. M. Siti Maesaroh, Afiyati, Lukman Hakim, Yunita Sartika Sari, Mohamad Yusuf, Eza Budi Perkasa, Wiranti Sri Utami, Norbertus Tri Suswanto Saptadi, Siti Mutmainah, Khairunnas, Eka Purnama Harahap, Zumhur Alamin, Inna Sabily Karima, Andi Saputra, *Bahasa Pemrograman Python*. :Sada Kurnia Pustaka, 2024.

[35] N. M. Surbakti *et al.*, "Penggunaan Bahasa Pemrograman Python dalam

repository.ibrahimy.ac.id

Pembelajaran Kalkulus Fungsi Dua Variabel," *Algoritm. J. Mat. Ilmu Pengetah. Alam, Kebumian dan Angkasa*, vol. 2, no. 3, pp. 98–107, 2024, doi: 10.62383/algoritma.v2i3.67.

[36] C. Rometdo Muzawi, M.Kom., CEH., *Mengenal Dasar Microsoft Excel Bagi Pemula*. Serasi Media Teknologi, 2024.

[37] S. M. K. Yahya, *Data Mining*. CV Jejak (Jejak Publisher), 2022.

[38] R. Gelar Guntara, "Pemanfaatan Google Colab Untuk Aplikasi Pendeteksian Masker Wajah Menggunakan Algoritma Deep Learning YOLOv7," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 5, no. 1, pp. 55–60, 2023, doi: 10.47233/jteksis.v5i1.750.

[39] I. · Dr. Ir. Rianto, M.Eng., IPM., ASEAN Eng., Prof. Ir. Paulus Insap Santosa, M.Sc., Ph.D., *Data Preparation untuk Machine Learning & Deep Learning*. penerbit andi, 2025.

[40] Suyanto, *Machine learning : tingkat dasar dan lanjut*. INFORMATIKA, 2018.

[41] D. T. Sani, R. Arifianti, J. B. Logistik, and U. Padjadjaran, "Desain Proses Untuk Pengiriman Barang Menggunakan Bagan Proses Aliran Pada Departemen Supply Chain Management PT Abhimata Citra Abadi," vol. 14, no. 2, pp. 137–151, 2024.

[42] Y. P. Astuti, A. R. Wibowo, E. Kartikadarma, E. R. Subhiyakto, N. A. Sri Winarsih, and M. S. Rohman, "Penerapan Metode Naïve Bayes Classifier Untuk Klasifikasi Sentimen Pada Judul Berita," *LogicLink*, vol. 1, no. 1, pp. 1–12, 2024, doi: 10.28918/logiclink.v1i1.7684.

Ibrahimy LIBRARY
Ibrahimy Repository
Universitas Ibrahimy
repository.ibrahimy.ac.id