**METODE *MULTI-LAYER PERCEPTRON* UNTUK KLASIFIKASI PENERIMA BANTUAN SOSIAL**

**PROGRAM KELUARGA HARAPAN (PKH)**

**DI KELURAHAN CIPAMOKOLAN**

**SKRIPSI**

**Diajukan sebagai salah satu syarat kelulusan pada Program Studi Sistem Informasi Jenjang S1 (Strata Satu) Fakultas Teknik dan Ilmu Komputer**

**Oleh :**

**DHEANDY DIVALLINO**

**10521049**

****

**PROGRAM STUDI SISTEM INFORMASI**

**FAKULTAS TEKNIK DAN ILMU KOMPUTER**

**UNIVERSITAS KOMPUTER INDONESIA**

**BANDUNG**

**2025**

# ABSTRAK

Bantuan sosial merupakan bantuan yang diberikan secara selektif kepada masyarakat dalam bentuk uang atau barang dengan tujuan meningkatkan kesejahteraan. Namun, proses penyaluran bantuan sosial masih terjadi ketidaktepatan sasaran karena masih dilakukan secara musyawarah. Penelitian ini penting dilakukan untuk meningkatkan efisiensi dan akurasi dalam proses seleksi penerima bantuan sosial Program Keluarga Harapan (PKH). Tujuan dari penelitian ini adalah untuk menerapkan metode *Multi-Layer Perceptron* (MLP) dalam mengklasifikasikan kelayakan calon penerima PKH di Kelurahan Cipamokolan serta mengetahui nilai akurasi, *precision*, *recall*, dan *f1-score* dari model yang dibangun.

Penelitian ini menggunakan metode kuantitatif dengan dataset sebanyak 422 data penerima PKH tahun 2023 dari Kelurahan Cipamokolan. Data dibagi menjadi dua bagian yaitu data latih (338 data) dan data uji (84 data). Proses pengolahan data meliputi tahapan *preprocessing*, transformasi fitur, dan pelatihan model menggunakan algoritma MLP. Model selanjutnya dievaluasi menggunakan metrik performa seperti akurasi, *precision*, *recall*, dan *f1-score*.

Hasil pengujian menunjukkan bahwa model MLP mampu memberikan performa yang baik dengan nilai akurasi sebesar 95,70%, *precision* sebesar 95,74%, *recall* sebesar 97,83%, dan *f1-score* sebesar 96,77%. Aplikasi klasifikasi dikembangkan dalam bentuk *website* menggunakan *framework streamlit* untuk memudahkan prediksi secara langsung. Penelitian ini menunjukkan bahwa penerapan metode MLP dapat membantu proses pengambilan keputusan secara optimal dan mengurangi kesalahan dalam seleksi calon penerima bantuan. Ke depannya, penelitian dapat diperluas dengan menambahkan data dari wilayah lain serta mengeksplorasi arsitektur dan parameter model secara lebih mendalam.

**Kata kunci: *Data Mining,* Program Keluarga Harapan, Klasifikasi, *Multi-Layer Perceptron*, *Streamlit*.**

# *ABSTRACT*

*Social assistance is a selective aid provided to the community in the form of money or goods with the aim of improving welfare. However, the distribution process of social assistance still suffers from mistargeting due to the selection process being conducted through deliberation. This study is important to improve the efficiency and accuracy of the selection process for recipients of the Family Hope Program (Program Keluarga Harapan / PKH). The purpose of this research is to implement the Multi-Layer Perceptron (MLP) method to classify the eligibility of PKH recipients in Cipamokolan Subdistrict and to evaluate the model’s performance in terms of accuracy, precision, recall, and F1-score.*

*This research uses a quantitative method with a dataset consisting of 422 PKH recipient records from 2023 in Cipamokolan. The data was divided into training data (338 records) and testing data (84 records). Data processing involved preprocessing, feature transformation, and model training using the MLP algorithm. The model was then evaluated using performance metrics such as accuracy, precision, recall, and F1-score.*

*The testing results show that the MLP model performs well with an accuracy of 95.70%, precision of 95.74%, recall of 97.83%, and F1-score of 96.77%. A classification application was developed in the form of a website using the Streamlit framework to enable direct predictions. This study demonstrates that the application of the MLP method can support optimal decision-making and reduce errors in the selection of social assistance recipients. Future research can be expanded by incorporating data from other regions and exploring model architectures and parameters more deeply.*

***Keywords: Data Mining, Family Hope Program, Classification, Multi-Layer Perceptron, Streamlit.***

# KATA PENGANTAR

Segala puji dan syukur saya panjatkan ke hadirat Allah SWT, yang telah melimpahkan rahmat, hidayah, dan karunia-Nya, sehingga saya dapat menyelesaikan skripsi dengan judul “**METODE *MULTI-LAYER PERCEPTRON* UNTUK KLASIFIKASI PENERIMA BANTUAN SOSIAL PROGRAM KELUARGA HARAPAN (PKH) DI KELURAHAN CIPAMOKOLAN**” tepat pada waktunya. Terima kasih saya ucapkan kepada semua pihak yang terlibat memberikan bantuan,motivasi, bimbingan, dan saran-saran, serta nasehat yang tidak ternilai pada penelitian ini.

1. Bapak Prof. Dr. Ir. H. Eddy Soeryanto Soegoto, MT. selaku Rektor Universitas Komputer Indonesia yang telah memberikan izin dan fasilitas untuk penyusunan skripsi ini.
2. Bapak Assoc. Prof. Dr. Ir.Herman Soegoto., MBA. selaku Dekan Fakultas Teknik & Ilmu Komputer Universitas Komputer Indonesia yang telah memberikan izin dalam penyusunan skripsi ini.
3. Ibu Dr. Wartika, S.Kom., MT. selaku Ketua Program Studi Sistem Informasi Universitas Komputer Indonesia.
4. Bapak Rangga Sidik, S.Kom., M.Kom., M.Eng, selaku Dosen Wali Kelas SI-2 Program Studi Sistem Informasi yang telah memberikan arahan kepada penulis dalam mengikuti dan menyelesaikan studi di Fakultas Teknik dan Ilmu Komputer Universitas Komputer Indonesia.
5. Bapak Dr. Assoc. Prof. Agus Nursikuwagus, ST., MT., MM., MOS., MTA. selaku Dosen Pembimbing yang telah memberikan bimbingan, arahan, dan saran-saran sehingga skripsi ini dapat diselesaikan dengan baik.
6. Bapak Muhammad Rajab Fachrizal, S.Kom., M.Kom. dan Ibu Dr. Wartika, S.Kom., M.T. selaku Dosen Penguji yang telah memberikan masukin dan saran sehingga skripsi ini bisa diselesaikan dengan baik.
7. Seluruh Dosen Pengajar dan Staf Universitas Komputer Indonesia khusunya di Jurusan Sistem Informasi yang telah memberikan ilmu pengetahuan selama penulis menuntut ilmu di Fakultas Teknik dan Ilmu Komputer.
8. Seksi Kesejahteraan Sosial Kantor Kecamatan Rancasari dan Koordinator Lapangan Progam Keluarga Harapan (PKH) beserta staf yang telah meluangkan waktu dan memberikan kesempatan kepada penulis untuk melakukan penelitian di Kantor Kecamatan Rancasari sehingga skripsi ini dapat diselesaikan dengan baik.
9. Kedua orang tua dan saudara kandung penulis tersayang, Ayahanda Darmadi, S.Pt., Ibunda Dessy Zulfia, Kakak Dhea Eudora, dan Adik Dheanjiela Nataghsya yang telah menjadi orang tua dan saudara kandung terhebat. Terimakasih yang tiada terhingga atas limpahan kasih sayang dan cinta yang tulus, do`a yang tak pernah putus, materi, motivasi, nasehat, perhatian, pengorbanan, dan semangat yang diberikan selalu membuat penulis selalu bersyukur telah memiliki keluarga yang luar biasa.
10. Teman-teman penulis dari grup SIU ESPORT yaitu Mochamad Adnan Quartelio, Muhammad Fakhri Dimastoro, Syahrul Choliq Mawardi, Sahlan Putra Pratama, Aloysius Antony Edward Woda, Muhammad Chaidar Aulia, Muhammad Fachrizal Rachman, Bilal Alamsyach, dan Tahfid Asrori serta teman-teman kelas IS-2 Sistem Informasi Universitas Komputer Indonesia yang telah menemani penulis selama masa perkuliahan. Terimakasih atas doa dan dukungan yang telah diberikan kepada penulis selama ini.
11. Semua pihak yang turut andil memberikan kontribusi penyusunan skripsi ini, yang tidak bisa disebutkan satu persatu. Tentu dukungan dan motivasi dari keluarga dan teman-teman sangat berarti bagi penulis.

Semoga rasa terima kasih ini dapat sampai kepada semua pihak yang telah membantu dalam menyelesaikan skripsi ini. Semoga Allah SWT membalas segala kebaikan dan usaha mereka dengan berlipat ganda. berharap skripsi ini dapat bermanfaat bagi pembaca.

Akhir kata, saya berharap dapat terus belajar dan berkembang serta menjadi seseorang yang bermanfaat bagi masyarakat dan bangsa. Semoga segala upaya yang telah dilakukan dapat menjadi amal jariyah bagi kehidupan di dunia dan akhirat.

Bandung, 25 Agustus 2025

Dheandy Divallino

# DAFTAR ISI

LEMBAR PENGESAHAN

LEMBAR PERNYATAAN KEASLIAN

[ABSTRAK i](#_Toc207124771)

[*ABSTRACT* ii](#_Toc207124772)

[KATA PENGANTAR iii](#_Toc207124773)

[DAFTAR ISI vi](#_Toc207124774)

[DAFTAR GAMBAR x](#_Toc207124775)

[DAFTAR TABEL xii](#_Toc207124776)

[DAFTAR SIMBOL xiii](#_Toc207124777)

[BAB I PENDAHULUAN 1](#_Toc207124778)

[1.1. Latar Belakang Penelitian 1](#_Toc207124779)

[1.2. Identifikasi dan Rumusan Masalah 6](#_Toc207124780)

[1.2.1. Identifikasi Masalah 6](#_Toc207124781)

[1.2.2. Rumusan Masalah 7](#_Toc207124782)

[1.3. Maksud dan Tujuan Penelitian 7](#_Toc207124783)

[1.3.1. Maksud Penelitian 7](#_Toc207124784)

[1.3.2. Tujuan Penelitian 8](#_Toc207124785)

[1.4. Kegunaan Penelitian 8](#_Toc207124786)

[1.5. Batasan Masalah 9](#_Toc207124787)

[1.6. Lokasi dan Waktu Penelitian 10](#_Toc207124788)

[1.6.1. Lokasi Penelitian 10](#_Toc207124789)

[1.6.2. Waktu Penelitian 10](#_Toc207124790)

[BAB II LANDASAN TEORI 12](#_Toc207124791)

[2.1. Penelitian Terdahulu 12](#_Toc207124792)

[2.2. Teori Pendukung 14](#_Toc207124793)

[2.2.1. Program Keluarga Harapan 14](#_Toc207124794)

[2.2.2. *Data Mining* 15](#_Toc207124795)

[2.2.2.1. Operasi *Data Mining* 16](#_Toc207124796)

[2.2.2.2. Pengelompokan *Data Mining* 17](#_Toc207124797)

[2.2.3. *Knowledge Discovery in Database* 20](#_Toc207124798)

[2.2.4. Klasifikasi 23](#_Toc207124799)

[2.2.5. *Deep Learning* 23](#_Toc207124800)

[2.2.6. *Multi-Layer Perceptron* 24](#_Toc207124801)

[2.2.6.1. Definisi Algoritma 24](#_Toc207124802)

[2.2.6.2. Pengertian Algoritma *Multi-Layer Perceptron* 24](#_Toc207124803)

[2.3. Piranti Pendukung 26](#_Toc207124804)

[2.3.1. *Python* 26](#_Toc207124805)

[2.3.2. *Jupyter Notebook* 26](#_Toc207124806)

[2.3.3. *Visual Studio Code* 27](#_Toc207124807)

[2.3.4. *Website* 28](#_Toc207124808)

[2.3.5. *Streamlit* 28](#_Toc207124809)

[BAB III OBJEK DAN METODE PENELITIAN 29](#_Toc207124810)

[3.1. Objek Penelitian 29](#_Toc207124811)

[3.1.1. Sejarah Singkat Kecamatan Rancasari 29](#_Toc207124812)

[3.1.2. Visi dan Misi Kecamatan Rancasari 30](#_Toc207124813)

[3.1.2.1. Visi Kecamatan Rancasari 30](#_Toc207124814)

[3.1.2.2. Misi Kecamatan Rancasari 30](#_Toc207124815)

[3.1.3. Struktur Organisasi Kantor Kecamatan Rancasari 31](#_Toc207124816)

[3.1.4. Deskripsi Tugas 31](#_Toc207124817)

[3.2. Metode Penelitian 33](#_Toc207124818)

[3.2.1. Desain Penelitian 34](#_Toc207124819)

[3.2.2. Jenis dan Metode Pengumpulan Data 34](#_Toc207124820)

[3.2.2.1. Sumber Data Primer 34](#_Toc207124821)

[3.2.2.2. Sumber Data Sekunder 35](#_Toc207124822)

[3.2.3. Tahapan Penelitian 36](#_Toc207124823)

[3.2.3.1. *Business Understanding* 36](#_Toc207124824)

[3.2.3.2. *Data Understanding* 37](#_Toc207124825)

[3.2.3.3. *Data Preparation* 37](#_Toc207124826)

[3.2.3.4. *Modelling* 41](#_Toc207124827)

[3.2.3.5. *Evaluation* 44](#_Toc207124828)

[3.2.3.6. *Deployment* 46](#_Toc207124829)

[3.3. Analisis Sistem yang Berjalan 46](#_Toc207124830)

[3.3.1. *Business Understanding* 47](#_Toc207124831)

[3.3.2. *Data Understanding* 55](#_Toc207124832)

[3.3.2.1. Sumber 55](#_Toc207124833)

[3.3.2.2. Validator 56](#_Toc207124834)

[3.3.2.3. Keterikatan Antar Fitur 57](#_Toc207124835)

[3.3.2.4. Statistik Deskriptif 59](#_Toc207124836)

[3.3.3. *Data Preparation* 64](#_Toc207124837)

[3.3.3.1. *Data Cleansing* 64](#_Toc207124838)

[3.3.3.2. *Data Transformation* 65](#_Toc207124839)

[3.3.3.3. *Outlier* 68](#_Toc207124840)

[3.3.3.4. *Data Selection* 70](#_Toc207124841)

[3.3.3.5. *Data Balancing* 73](#_Toc207124842)

[3.3.3.6. *Data Splitting* 75](#_Toc207124843)

[BAB IV HASIL DAN PEMBAHASAN 76](#_Toc207124844)

[4.1. Hasil 76](#_Toc207124845)

[4.1.1. *Modelling* 76](#_Toc207124846)

[4.1.1.1. Perhitungan Manual *Multi-Layer Perceptron* 77](#_Toc207124847)

[4.1.2. *Metrik* 79](#_Toc207124848)

[4.1.2.1. *Test Set* 80](#_Toc207124849)

[4.1.2.2. *Train Set* 83](#_Toc207124850)

[4.1.3. *Deployment* 85](#_Toc207124851)

[4.2. Pembahasan dan Rekomendasi 88](#_Toc207124852)

[4.2.1. Pembahasan 89](#_Toc207124853)

[4.2.1.1. Perbandingan Kinerja MLP dengan Algoritma Lain 89](#_Toc207124854)

[4.2.2. Rekomendasi 93](#_Toc207124855)

[BAB V KESIMPULAN DAN SARAN 95](#_Toc207124856)

[5.1. Kesimpulan 95](#_Toc207124857)

[5.2. Saran 96](#_Toc207124858)

[DAFTAR PUSTAKA 98](#_Toc207124859)

[LAMPIRAN 101](#_Toc207124860)

# DAFTAR GAMBAR

[Gambar 1. 1 Lokasi Penelitian 10](#_Toc197299054)

[Gambar 2. 1 Model atau Pengetahuan Merupakan Output *Data Mining* 16](#_Toc197299043)

[Gambar 2. 2 Strategi *Data Mining* 17](#_Toc197299044)

[Gambar 2. 3 Pengelompokan *Data Mining* 17](#_Toc197299045)

[Gambar 2. 4 Proses *Knowledge Discovery in Database* (KDD) 21](#_Toc197299046)

[Gambar 3. 1 Logo Kecamatan Rancasari 29](#_Toc204887304)

[Gambar 3. 2 Struktur Organisasi Kantor Kecamatan Rancasari 31](#_Toc204887305)

[Gambar 3. 3 Alir Penelitian 36](#_Toc204887306)

[Gambar 3. 4 Alir Penerapan Feature Selection 38](#_Toc204887307)

[Gambar 3. 5 Alir PKH 49](#_Toc204887308)

[Gambar 3. 6 Alir Proses Justifikasi Data 53](#_Toc204887309)

[Gambar 3. 7 Data PKH Tahun 2023 56](#_Toc204887310)

[Gambar 3. 8 *Chart* Status Kelayakan 62](#_Toc204887311)

[Gambar 3. 9 *Bar Chart* Pekerjaan 63](#_Toc204887312)

[Gambar 3. 10 Penghapusan dan Pemisahan Fitur 64](#_Toc204887313)

[Gambar 3. 11 Fitur Target 64](#_Toc204887314)

[Gambar 3. 12 *Missing Value* 65](#_Toc204887315)

[Gambar 3. 13 *One-Hot Encoding* 66](#_Toc204887316)

[Gambar 3. 14 Hasil Transformasi Fitur Kategori 66](#_Toc204887317)

[Gambar 3. 15 Transformasi Fitur Target Dengan *Label Binarizer* 67](#_Toc204887318)

[Gambar 3. 16 *Outlier* 68](#_Toc204887319)

[Gambar 3. 17 Visualisasi *Outlier* 69](#_Toc204887320)

[Gambar 3. 18 *Data Selection* Menggunakan CFS 70](#_Toc204887321)

[Gambar 3. 19 *Correlation Matrix* 71](#_Toc204887322)

[Gambar 3. 20 *Feature Selection* 72](#_Toc204887323)

[Gambar 3. 21 Diagram *Imbalanced Class* 73](#_Toc204887324)

[Gambar 3. 22 *Data Balancing* Menggunakan SMOTE 74](#_Toc204887325)

[Gambar 3. 23 Diagram Status Kelayakan Setelah Proses SMOTE 74](#_Toc204887326)

[Gambar 4. 1 *Modelling* dengan MLP 76](#_Toc204887295)

[Gambar 4. 2 Dataset Setelah *Preparation* Data 78](#_Toc204887296)

[Gambar 4. 3 *Confusion Matrix Test Set* *Heatmap* 80](#_Toc204887297)

[Gambar 4. 4 *Confusion Matrix Train Set Heatmap* 83](#_Toc204887298)

[Gambar 4. 5 Tampilan *Input* Data Tunggal 86](#_Toc204887299)

[Gambar 4. 6 Hasil Prediksi *Input* Data Tunggal 86](#_Toc204887300)

[Gambar 4. 7 Tampilan *Input* Multi Data 87](#_Toc204887301)

[Gambar 4. 8 Hasil Prediksi *Input* Multi Data 88](#_Toc204887302)

[Gambar 4. 10 Visualisasi Hasil Klasifikasi 93](#_Toc204887303)

# DAFTAR TABEL

[**Tabel 1. 1 Waktu Penelitian 11**](#_Toc197299844)

[**Tabel 3. 1 Confusion Matrix 42**](#_Toc197566511)

[**Tabel 3. 2 Statistik Deskriptif Fitur Numerik 59**](#_Toc197566512)

[**Tabel 3. 3 Statistik Deskriptif Fitur Kategorikal 60**](#_Toc197566513)

[**Tabel 3. 4 Statistik Deskriptif Fitur Kategorikal Lanjutan 60**](#_Toc197566514)

[**Tabel 3. 5 Statistik Deskriptif Hubungan Antar Pekerjaan Dengan Fitur Numerik 61**](#_Toc197566515)

[**Tabel 3. 6 Hasil Pengecekan Missing Value 65**](#_Toc197566516)

[**Tabel 3. 7 Perbedaan Fitur Target Sebelum dan Sesudah Transformasi 68**](#_Toc197566517)

[**Tabel 3. 8 Pembagian Proporsi Data 75**](#_Toc197566518)

[**Tabel 4. 1 Data Uji yang Dipilih 77**](#_Toc204887153)

[**Tabel 4. 2 *Confusion Matrix Test Size* 80**](#_Toc204887154)

[**Tabel 4. 3 *Confusion Matrix Train Size* 83**](#_Toc204887155)

[**Tabel 4. 4 Hasil Klasifikasi pada Proporsi Data 89**](#_Toc204887156)

[**Tabel 4. 5 Hasil Klasifikasi Perbandingan Algoritma 90**](#_Toc204887157)

[**Tabel 4. 6 Klasifikasi Data Kelayakan Penerima Bantuan Sosial PKH 92**](#_Toc204887158)

# DAFTAR SIMBOL

1. **Simbol Flowchart**

**(Sumber : Jurnal Ilmiah Komputer dan Informatika** [1]**)**

|  |  |  |
| --- | --- | --- |
| **Simbol** | **Nama** | **Keterangan** |
|  | ***Flow Direction Symbol*** | **Untuk menghubungkan antara simbol yang satu dengan simbol yang lain. Simbol ini disebut juga *connecting line.*** |
|  | ***Terminator Symbol*** | **Untuk permulaan (*start)* atau akhir (*stop*) dari suatu kegiatan.** |
|  | ***Connector Symbol*** | **Untuk keluar-masuk atau penyambungan proses dalam lembar/halaman yang sama.** |
|  | ***Processing Symbol*** | **Menunjukkan pengolahan yang dilakukan oleh komputer.** |
|  | **Simbol Input-Output** | **Menyatakan proses input dan output tanpa tergantung dengan jenis peralatannya.** |

# PENDAHULUAN

## Latar Belakang Penelitian

Kemiskinan menjadi salah satu penyebab terhambatnya pembangunan di suatu wilayah, karena permasalahannya yang bersifat kompleks dan sulit diselesaikan secara mandiri oleh daerah tersebut. Kemiskinan dapat diartikan sebagai kondisi dimana suatu keluarga tidak mampu mencukupi kebutuhan dasar untuk menjalani kehidupan yang layak sesuai dengan standar lingkungan sekitar. Untuk mengatasi permasalahan ini, dilakukan berbagai upaya melalui bantuan dan perlindungan sosial, seperti pelaksanaan Program Keluarga Hatapan (PKH) yang bertujuan mengurangi angka kemiskinan [2]. PKH merupakan program bantuan sosial bersyarat yang diberikan pemerintah kepada keluarga miskin dengan tujuan meningkatkan kesejahteraan masyarakat dalam aspek pendidikan dan kesehatan [3].

Cipamokolan adalah kelurahan di Kecamatan Rancasari, Kota Bandung, Provinsi Jawa Barat, yang melaksanakan program PKH. Berdasarkan data Badan Pusat Statistik Kota Bandung dan dinas sosial tahun 2023, jumlah penerima program PKH di Kelurahan Cipamakolan sebanyak 236 KK dari total 6.339 KK dengan total jumlah penduduk 25.668 jiwa. Melalui wawancara dengan Seksi Kesejahteraan Sosial Kantor Kecamatan Rancasari, pengelolaan data penerima PKH selama ini telah berjalan sesuai prosedur: data warga dimiliki oleh masing-masing kelurahan, diteruskan ke kecamatan, kemudian dilaporkan ke Dinas Sosial untuk proses penyaluran bantuan. Program PKH sendiri dilakukan sebanyak tiga kali dalam setahun. Meskipun proses ini berjalan cukup baik, proses verifikasi dan validasi data masih dilakukan secara manual. Misalnya, apabila terjadi perubahan kondisi keluarga seperti peningkatan pendapatan, perpindahan jenjang pendidikan anak, atau keluarga sudah tidak memenuhi kriteria, maka proses pembaruannya memerlukan musyawarah oleh kelurahan, Rukun Warga (RW), dan Rukun Tetangga (RT) yang memakan waktu dan sumber daya. Pendekatan manual ini berpotensi menimbulkan keterlambatan pembaruan data, yang pada akhirnya dapat berdampak pada ketidaktepatan dalam penyaluran bantuan.

Permasalahan ini sejalan dengan temuan yang dilansir dari Bisnis.com (2024) bahwa situasi di lapangan menunjukkan masih banyak permasalahan dalam pendataan dan verifikasi penerima bantuan. Seperti yang ditemukan oleh BPK, masih terdapat penerima manfaat yang telah meninggal dunia, memiliki penghasilan di atas UMP/UMK, bahkan memiliki badan usaha terdaftar namun tetap menerima bantuan. Di Kelurahan Cipamokolan, permasalahan serupa dimungkinkan terjadi, mengingat proses pendataan penerima bantuan masih dilakukan secara konvensional. Hal ini menunjukkan bahwa masalah ini layak untuk diteliti guna memberikan solusi berbasis teknologi, khususnya dalam proses *data mining* untuk klasifikasi data penerima bantuan.

*Data mining* merupakan proses yang mencakup pengumpulan serta pemakaian data historis untuk menemukan pola, keteraturan, atau hubungan tertentu dalam kumpulan data yang besar. Hasil dari proses ini dapat digunakan sebagai dasar untuk meningkatkan kualitas pengambilan keputusan di masa depan. Klasifikasi merupakan proses membangun model atau fungsi yang dapat menggambarkan serta membedakan berbagai kelas data atau konsep, dengan tujuan untuk memprediksi kelas dari suatu objek yang belum diketahui labelnya. Metode klasifikasi ini bertujuan untuk menentukan nilai dari suatu variable yang belum diketahui berdasarkan variable-variabel lain yang telah ditetapkan sebelumnya [4]. Fungsi klasifikasi berperan dalam mempermudah proses identifikasi antara individu yang layak dan tidak layak menerima bantuan sosial PKH. Teknik klasifikasi sendiri merupakan metode pembelajaran yang digunakan untuk memperkirakan nilai dari fitur target yang bersifat kategorikal [5].

*Artificial Neural Network* (ANN) merupakan salah satu algoritma yang banyak digunakan dalam teknik klasifikasi, ANN dirancang menyerupai cara kerja otak manusia dalam mengolah informasi [6]. ANN terdiri dari beberapa lapisan neuron yang saling terhubung, yakni lapisan *input*, lapisan tersembunyi, dan lapisan *output* [7]. Struktur jaringan ini memungkinkan model untuk belajar dari data melalui proses pelatihan yang melibatkan penyesuaian bobot antar neuron, sehingga mampu mengenali pola yang kompleks dalam data yang mungkin sulit dideteksi oleh algoritma lainnya [7]. Kelebihan utama ANN dalam klasifikasi adalah kemampuannya dalam menangani data bersifat non-linear dan menyesuaikan struktur model terhadap tingkat kompleksitas data, yang sangat berguna untuk memprediksi penerima bantuan dengan tingkat akurasi yang tinggi [8]. ANN memiliki berbagai bentuk yang dirancang untuk menangani berbagai kebutuhan khusus. Dari berbagai bentuk tersebut, *Multi-Layer Perceptron* (MLP) menjadi salah satu bentuk ANN yang paling umum digunakan [9].

MLP adalah jenis arsitektur jaringan saraf tiruan dengan beberapa lapisan neuron, dimana setidaknya satu lapisan tersembunyi di antara lapisan *input* dan *output*. MLP termasuk dalam jenis jaringan saraf *feedforward*, dimana algoritmanya bekerja dengan menghitung bobot dari *input* dan menerapkan fungsi aktivasi untuk menghasilkan *output*. Metode ini efektif dalam memodelkan fungsi yang kompleks serta dalam menangani tugas-tugas yang melibatkan data terstruktur [9]. MLP memiliki keunggulan dalam memodelkan hubungan yang kompleks dan non-linear dalam data. Selain itu, MLP juga mampu melakukan ekstraksi fitur secara otomatis dan menyelesaikan permasalahan yang tidak dapat ditangani oleh model linear. Karena keunggulan tersebut, MLP menjadi salah satu algoritma yang cukup populer dalam berbagai penerapan pembelajaran mesin [10]. Dalam hal klasifikasi data, arsitektur MLP yang terdiri dari jaringan neuron yang saling terhubung dapat mempermudah dalam mengenali pola yang kompleks pada dataset [11].

Penelitian ini bertujuan untuk menghasilkan sebuah model klasifikasi yang akurat dan andal untuk memprediksi penerima bantuan sosial PKH yang benar-benar layak menggunakan metode *Multi-Layer Perceptron* (MLP). Dalam penelitian ini, fitur yang digunakan merujuk pada kriteria penilaian penerima PKH sebagaimana diatur dalam Peraturan Menteri Sosial (Permensos) Republik Indonesia Nomor 1 tahun 2018 yang terdiri dari 8 kriteria. Kriteria tersebut mencakup: Ibu hamil atau menyusui, balita, anak prasekolah, anak sekolah (tingkat SD/sederajat, SMP/sederajat, dan SMA/sederajat), lanjut usia, serta penyandang disabilitas [3].

Beberapa penelitian sebelumnya telah menggunakan pendekatan *data mining* dalam penentuan penerima bantuan sosial. Namun, Sebagian besar masih menggunakan algoritma seperti *Decision Tree*, *Naïve Bayes*, atau *K-Nearest Neighbor.* Oleh karena itu, penelitian ini mencoba menerapkan model *deep learning* dengan menggunakan metode *Multi-Layer Perceptron* (MLP). Penelitian ini akan membandingkan efektifitas model MLP terhadap pendekatan lain yang sudah ada.

Secara teoritis, penelitian ini memperkuat literatur mengenai penerapan *deep learning* dalam klasifikasi data sosial. Secara praktis, sistem klasifikasi yang dibangun dapat digunakan sebagai alat bantu dalam pengambilan keputusan bagi pemerintah daerah, khususnya dalam menyaring data keluarga calon penerima bantuan sosial dengan lebih tepat. Dengan begitu, bantuan sosial PKH akan lebih tepat sasaran dan membawa dampak yang lebih signifikan terhadap peningkatan kesejahteraan masyarakat.

Berdasarkan uraian permasalahan diatas, penelitian ini dilakukan untuk mengklasifikasi calon keluarga penerima bantuan sosial PKH Kelurahan Cipamokolan. Oleh karena itu dalam penelitian ini akan berfokus terhadap metode MLP dalam melakukan klasifikasi terhadap masyarakat yang layak dan tidak layak menerima bantuan PKH. Peneliti akan melakukan penelitian di Kantor Kecamatan Rancasari yang berlokasi di Jl. Santosa Asih No.17 Bandung dengan judul “**METODE MULTI-LAYER PERCEPTRON UNTUK KLASIFIKASI PENERIMA BANTUAN SOSIAL PROGRAM KELUARGA HARAPAN (PKH) DI KELURAHAN CIPAMOKOLAN**”.

## Identifikasi dan Rumusan Masalah

Dalam mengidentifikasi dan merumuskan masalah, penulis telah melakukan observasi langsung dilokasi penelitian untuk mengumpulkan data yang dibutuhkan mengenai permasalahan yang sedang dihadapi.

### Identifikasi Masalah

Berdasarkan uraian latar belakang, maka dapat diidentifikasikan permasalahan sebagai berikut:

1. Masih terdapat risiko ketidaktepat sasaran dalam menentukan penerima bantuan sosial PKH, dimana adanya kemungkinan beberapa penerima yang tidak lagi memenuhi kriteria begitu juga sebaliknya karena terjadi perubahan kondisi keluarga.
2. Proses verifikasi dan validasi data penerima PKH masih belum berjalan secara optimal, sehingga masih terdapat penerima yang sebenarnya tidak memenuhi syarat karena perubahan kondisi keluarga namun tetap memperoleh bantuan, dan berpotensi menyebabkan pemborosan anggaran negara.
3. Kurangnya pemanfaatan teknologi prediktif dalam proses pengambilan keputusan terkait penyaluran bantuan sosial, seperti pemanfaatan model *deep learning* dengan menggunakan metode klasifikasi *Multi-Layer Perceptron* (MLP), yang sebenarnya memiliki potensi tinggi untuk meningkatkan akurasi prediksi dalam membantu menentukan siapa yang benar-benar berhak menerima bantuan berdasarkan pola data historis.

### Rumusan Masalah

Berdasarkan identifikasi masalah di atas, maka rumusan masalah pada penelitian ini adalah:

1. Bagaimana penerapan metode *Multi-Layer Perceptron* (MLP) dalam mengklasifikasikan penerima bantuan sosial PKH di Kelurahan Cipamokolan.
2. Bagaimana alir *preprocessing* data dalam memprediksi penerima bantuan sosial PKH di Kelurahan Cipamokolan.
3. Bagaimana model prediksi yang digunakan untuk menentukan penerima bantuan sosial PKH di Kelurahan Cipamokolan.
4. Bagaimana tingkat akurasi data dalam proses prediksi penerima bantuan sosial PKH di Kelurahan Ciapmokolan.
5. Bagaimana *deployment* model berbasis web algoritma MLP dalam memprediksi penerima bantuan sosial PKH di Kelurahan Cipamokolan.

## Maksud dan Tujuan Penelitian

Berdasarkan uraian rumusan masalah di atas, berikut maksud dan tujuan yang ingin dicapai dalam mengatasi permasalahan yang terjadi.

### Maksud Penelitian

Maksud dari penelitian ini adalah untuk menerapkan teknik *data mining* dengan model *deep learning* menggunakan metode *Multi-Layer Perceptron* (MLP) untuk mengklasifikasikan penerima bantuan sosial PKH di Kelurahan Cipamokolan, sehingga dapat meminimalkan potensi kesalahan dalam proses penyaluran bantuan tersebut.

### Tujuan Penelitian

Adapun tujuan dari penelitian ini, antara lain:

1. Menentukan kelayakan keluarga penerima bantuan sosial PKH berdasarkan hasil pengklasifikasian menggunakan metode *Multi-Layer Perceptron.*
2. Mengetahui *preprocessing* data dalam memprediksi penerima bantuan sosial PKH di Kelurahan Cipamokolan.
3. Mengetahui model prediksi yang digunakan untuk menentukan penerima bantuan sosial PKH di Kelurahan Cipamokolan.
4. Mengetahui tingkat akurasi data dalam memprediksi penerima bantuan sosial PKH di Kelurahan Cipamokolan.
5. Mengetahui *deployment* model berbasis web algoritma MLP dalam memprediksi penerima bantuan sosial PKH di Kelurahan Cipamokolan

## Kegunaan Penelitian

Penelitian ini diharapkan dapat memberikan manfaat dari dua sisi, yaitu secara teoritis dan secara praktis sebagai berikut:

1. Kegunaan Teoritis

Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan ilmu pengetahuan di bidang teknologi informasi, khususnya dalam penerapan *data mining* untuk klasifikasi pada data sosial. Dengan menggunakan metode Mulit-Layer Perceptron (MLP), penelitian ini juga diharapkan dapat menambah referensi ilmiah terkait implementasi deep learning pada data tabular, terutama pada klasifikasi penerima bantuan sosial. Selain itu, penelitian ini dapat menjadi pedoman bagi penelitian selanjutnya yang ingin mengeksplorasi penggunaan MLP dalam permasalahan sosial lainnya.

1. Kegunaan Praktis

Hasil penelitian ini diharapkan dapat membantu instansi terkait dalam meningkatkan akurasi dan efektivitas proses klasifikasi penerima bantuan sosial PKH. Dengan adanya model klasifikasi berbasis MLP, penyaluran bantuan dapat lebih tepat sasaran, menguransi risiko kesalahan data penerima, dan meminimalkan potensi penyimpangan dalam distribusi bantuan.

## Batasan Masalah

Klasifikasi ini memiliki cakupan yang luas, untuk itu agar penelitian lebih fokus, maka peneliti membuat batasan masalah yaitu:

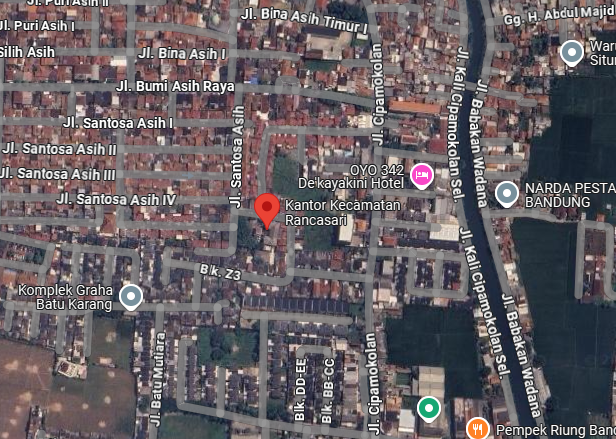
1. Menggunakan metode *Multi-Layer Perceptron* (MLP) untuk klasifikasi kelayakan masyarakat dalam mendapatkan bantuan sosial PKH;
2. Klasifikasi penerima bantuan sosial PKH dilakukan berdasarkan data penerima bantuan sosial PKH 2023 yang diperoleh dari Kantor Kecamatan Rancasari dengan total data sampel sebanyak 422 data;
3. Data yang digunakan dalam penelitian ini diperoleh melalui hasil observasi dan wawancara langsung di Kantor Kecamatan Rancasari;
4. Menggunakan *Confusion Matrix* sebagai pengujian terhadap hasil klasifikasi;
5. Fitur yang digunakan adalah usia, penghasilan, pekerjaan, status pernikahan, jumlah tanggungan, pendidikan, status rumah, jenis kelamin, dan kriteria penerima PKH;
6. Mengunakan Bahasa pemrograman *Python* dengan *tools Visual Studio Code* dan *Jupyter Notebook*.

## Lokasi dan Waktu Penelitian

Penelitian ini dilakukan pada Kantor Kecamatan dengan waktu yang telah direncanakan sebagai berikut.

### Lokasi Penelitian

Penelitian ini dilakukan di Kantor Kecamatan Rancasari yang beralamat di Jalan Santosa Asih No.17, Cipamokolan, Kec. Rancasari, Kota Bandung, Jawa Barat 40292.



Gambar 1. 1 Lokasi Penelitian  
(Sumber: https://maps.google.com/maps)

### Waktu Penelitian

Penelitian ini dilaksanakan selama 6 bulan mulai dari Bulan Februari 2025 sampai dengan Juli 2025.

Tabel 1. 1 Waktu Penelitian

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **No.** | **Aktivitas** | **Waktu** | | | | | | | | | | | | | | | | | | | | | | | |
| Februari | | | | Maret | | | | April | | | | Mei | | | | Juni | | | | Juli | | | |
| 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| **1.** | **Identifikasi Kebutuhan** | | | | | | | | | | | | | | | | | | | | | | | | |
| a. | Studi Literatur |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| b. | Observasi |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| c. | Wawancara |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| d. | Pengumpulan Data |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **2.** | **Perancangan** | | | | | | | | | | | | | | | | | | | | | | | | |
| a. | *Preprocessing* data |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| b. | *Splitting* data |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **3.** | **Implementasi dan *Deployment*** | | | | | | | | | | | | | | | | | | | | | | | | |
| a. | *Modelling* data |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| b. | *Deployment Website* |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

# LANDASAN TEORI

## Penelitian Terdahulu

Penelitian yang dilakukan oleh Agus Nursikuwagus, dkk. pada tahun 2023 yang berjudul “*Hyperparameter Tuning of Multilayer Convolutional and Augmentation Method for Classification Motive of Batik*” bertujuan untuk membuat sistem klasifikasi gambar motif batik untuk memudahkan masyarakat dalam mengenali nama jenis motif batik. Penelitian tersebut menggunakan metode kuantitatif dengan tujuh jenis motif batik yang terlebih dahulu dilakukan augmentasi. Hasil penelitian menunjukkan bahwa akurasi dan presisi mencapai 98,5% dengan menggunakan metode *Machine Learning Convolutional Neural Network* (CNN). Persamaan penelitian ini dengan penelitian yang penulis lakukan terletak pada penggunaan model *deep learning* dengan algoritma *neural network* untuk menyelesaikan permasalahan klasifikasi. Perbedaannya terdapat pada jenis dataset dan algoritma yang digunakan, di mana penelitian yang dilakukan oleh Agus Nursikuwagus, dkk. menggunakan data gambar dan algoritma CNN yang khusus untuk pengolahan citra, sedangkan penelitian ini menggunakan data tabular dan menerapkan metode MLP yang lebih sesuai untuk data numerik dan kategorikal [12].

Penelitian yang dilakukan oleh Senang Hati Gulo dan Andre Hasudungan Lubis pada tahun 2024 yang berjudul “Pengembangan *Multi-Layer Perceptron* untuk Mengklasifikasi Penduduk Kurang Mampu” mengaplikasikan metode MLP untuk mengklasifikasikan penduduk kurang mampu dengan menggunakan dataset sebanyak 386 sampel. Penelitian ini membagi data menjadi data latih dan uji dengan proporsi 70:30, 80:20, dan 90:10. Hasil penelitian menunjukkan bahwa arsitektur MLP 3-5-1 (dengan 3 *input layer*, 5 *hidden layer*, dan 1 *output layer*) memberikan performa terbaik dengan akurasi 96,9%, recall 92%, precision 98,5%, dan F-Score 94,9% pada proporsi data 70:30. Persamaan penelitian ini dengan penelitian yang penulis lakukan terletak pada penggunaan algoritma MLP untuk klasifikasi data penduduk yang berkaitan dengan status sosial ekonomi serta teknik *encoding*. Perbedaannya terletak pada fokus penelitian, di mana penelitian yang dilakukan oleh Senang Hati Gulo dan Andre Hasudungan Lubis berfokus pada klasifikasi penduduk dalam tiga kategori, sedangkan penelitian ini lebih spesifik mengklasifikasikan penerima bantuan sosial sebagai layak atau tidak layak [9].

Penelitian yang dilakukan oleh I Wayan Supriana, dkk. pada tahun 2023 yang berjudul “Pengembangan Sistem Prediksi Bantuan Program Keluarga Harapan (PKH) Berbasis *Machine Learning*” mengembangkan sistem prediksi bantuan PKH berbasis *Artificial Neural Network* (ANN) dengan algoritma *backpropagation*. Penelitian tersebut menggunakan data penerima PKH sebanyak 6792 baris dengan 8 kriteria penilaian. Hasil penelitian menunjukkan bahwa model ANN dengan jumlah *hidden layer* sebanyak 500 neuron dan 375 *epoch* mampu mencapai nilai koefisien determinasi sebanyak 69,85%. Persamaan penelitian ini dengan penelitian yang penulis lakukan terletak pada penggunaan pendekatan jaringan saraf tiruan untuk membantu proses seleksi penerima bantuan PKH, termasuk tahap *preprocessing data* seperti *encoding* serta pengujian model menggunakan data latih dan uji. Perbedaannya terletak pada fokus penelitian, di mana penelitian yang dilakukan oleh I Wayan Supriana, dkk. berfokus pada prediksi nilai atau jumlah bantuan yang akan diberikan, sedangkan penelitian penulis berfokus pada klasifikasi status penerima sebagai layak atau tidak layak menerima bantuan [13].

## Teori Pendukung

Teori pendukung merupakan istilah dalam penelitian yang dapat digunakan sebagai acuan untuk memahami penelitian yang dilakukan penulis.

### Program Keluarga Harapan

Program Keluarga Harapan (PKH) merupakan salah satu bentuk bantuan tunai bersyarat (*Conditional Cash Transfer*/CCT) yang dikenal didunia sebagai strategi efektif dalam menanggulangi kemiskinan kronis. Di Indonesia, PKH menjadi bagian dari upaya percepatan pengentasan kemiskinan dan perlindungan sosial bagi keluarga yang paling rentan. Pelaksanaan PKH didasari oleh ketentuan hukum, salah satunya tercantum dalam Peraturan Menteri Sosial Nomor 1 Tahun 2018 tentang Program Keluarga Harapan, yang menggantikan regulasi sebelumnya yakni Peraturan Menteri Sosial Nomor 10 Tahun 2017 (Berita Negara Republik Indonesia Tahun 2017 Nomor 940) [14].

Program ini menargetkan Keluarga Penerima Manfaat (KPM) yang memiliki tanggungan di bidang pendidikan dan kesehatan, sebagai syarat utama untuk memperoleh bantuan. Tujuan utama PKH adalah meningkatkan kualitas hidup keluarga miskin melalui kemudahan akses terhadap layanan pendidikan, layanan kesehatan, serta pelayanan kesejahteraan sosial. Sasaran penerima PKH adalah individu atau keluarga miskin serta kelompok rentan yang terdaftar dalam Data Terpadu Program Penanganan Fakir Miskin. Kelompok sasaran ini dikategorikan berdasarkan komponen yang menjadi prioritas bantuan, yakni komponen kesehatan (seperti ibu hamil dan balita), komponen pendidikan (anak usia sekolah), serta komponen kesejahteraan sosial (seperti penyandang disabilitas berat dan lanjut usia) [14].

### *Data Mining*

*Data mining* adalah sebuah proses sistematis untuk menggali informasi dan pola tersembunyi yang bernilai dari kumpulan data berukuran besar. Proses ini mencakup serangkaian tahapan, mulai dari pengumpulan data, ekstraksi data, analisis data, dan statistik data. *Data mining* juga dikenal dengan sebagai *knowledge discovery, knowledge extraction*, analisis pola/data, hingga pemanfaatan informasi (*information harvesting*) [15].

*Data mining* juga merupakan proses logis yang bertujuan menemukan informasi yang berguna. Pola dan informasi yang berhasil ditemukan dapat dimanfaatkan sebagai alat pendukung dalam proses pengambilan keputusan strategis, khususnya dalam konteks pengembangan bisnis. Alat data mining mampu menjawab pertanyaan-pertanyaan yang terkait dengan bisnis dan sering kali terlalu sulit diselesaikan hanya dengan analisis konvensional. *Data mining* juga memiliki peran penting dalam meramalkan tren di masa depan, memungkinkan para pelaku bisnis untuk menyusun keputusan yang lebih efektif, proaktif, serta dinamis. Dengan mengolah data menggunakan teknik *data mining* yang tepat, organisasi dan perusahaan dapat menghasilkan pengetahuan yang akurat dan relevan, sehingga sesuai dengan kebutuhan dan harapan dalam menghadapi tantangan bisnis yang terus berkembang. Secara sederhana, *data mining* dapat digambarkan sebagai suatu pola atau model atau kaidah atau pengetahuan yang dihasilkan dari *data mining* [15], seperti yang ditunjukkan pada Gambar 2.1.

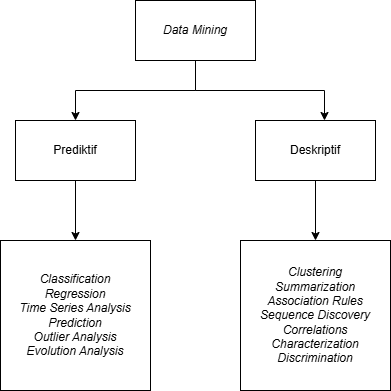


Gambar 2. 1 Model atau Pengetahuan Merupakan *Output* *Data Mining*  
(Sumber: *Data Mining*-Algoritma dan Implementasi [15])

Definisi lain dari *data mining* adalah suatu tahapan penting dalam proses analisis data yang bertujuan untuk mengekstraksi informasi berharga dari kumpulan database yang sangat besar. Dalam penerapannya, *data mining* memanfaatkan teknik khusus untuk membangun sebuah model berdasarkan data yang telah tersedia. Model ini kemudian digunakan untuk mengidentifikasi dan mengenali pola-pola tersembunyi dalam data lain yang sebelumnya belum dianalisis atau tidak terdapat dalam basis data yang ada [16].

#### Operasi *Data Mining*

*Data mining* memiliki berbagai macam fungsi, namun secara umum dapat dikelompokkan menjadi dua tujuan utama, yaitu fungsi deskriptif dan fungsi prediktif Fungsi deskriptif dalam *data mining* berfokus pada pemahaman mendalam terhadap data yang diamati dan bertujuan untuk menemukan pola, relasi, atau anomali dalam data. Strategi deskriptif dalam *data mining* mencakup teknik seperti *Clustering, Summarization, Association Rules, Sequence Discovery, Correlations, Characterization, dan Discrimination*. Sementara itu, fungsi prediktif berperan dalam membangun model yang mampu meramalkan atau memprediksi nilai dari variabel lain yang belum diketahui. Strategi prediktif dalam *data mining* mencakup teknik seperti *Classification, Regression, Time Series Analysis, Prediction, Outlier Analysis, dan Evolution Analysis* [16]. Pembagian teknik pada strategi prediktif dan deskriptif dapat dilihat pada Gambar 2.2.



Gambar 2. 2 Strategi *Data Mining*

#### Pengelompokan *Data Mining*

*Data mining* dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu



Gambar 2. 3 Pengelompokan *Data Mining*  
(Sumber: CBIS Journal [17])

1. Deskripsi

Proses deskripsi bertujuan untuk menemukan atau mengenali pola yang sering muncul lalu mengubah pola tersebut menjadi aturan yang dapat dimanfaatkan untuk mempermudah suatu aktivitas. Contoh pada sebuah toko, jika pelanggan sering membeli produk A dan produk B secara bersamaan dan terjadi secara berulang, maka pihak manajemen toko dapat menyesuaikan penataan katalog dengan menempatkan produk A dan B dalam posisi yang berdekatan, sehingga pelanggan akan lebih mudah saat ingin membeli kembali kedua produk tersebut. Salah satu algoritma yang digunakan dalam proses deskripsi adalah algoritma *Apriori*.

1. Estimasi

Estimasi memiliki kemiripan dengan klasifikasi, namun perbedaannya terletak pada variabel target yang berupa numerik, bukan kategorikal. Proses pembuatan model estimasi diawali dengan menggunakan data lengkap yang sudah mencakup nilai variabel target untuk tujuan prediksi. Pada tahap berikutnya, nilai variabel target akan diperkirakan berdasarkan data dari variabel-variabel prediksi. Sebagai ilustrasi, misalnya dilakukan estimasi terhadap tekanan darah sistolik pasien rumah sakit dengan menggunakan informasi seperti usia, jenis kelamin, indeks massa tubuh, dan kadar natrium dalam darah. Pola hubungan antara tekanan darah sistolik dengan variabel-variabel prediksi ini selama proses pelatihan akan membentuk sebuah model estimasi. Model yang dihasilkan ini nantinya dapat diterapkan untuk memprediksi kasus serupa pada data pasien yang baru. Algoritma yang termasuk dalam estimasi adalah regresi linear sederhana, regresi linear berganda, dan lain-lain.

1. Prediksi

Prediksi memiliki kemiripan dengan proses klasifikasi dan estimasi, namun perbedaannya terletak pada hasil yang diproyeksikan untuk waktu yang akan datang. Beberapa metode dan teknik yang diterapkan dalam klasifikasi maupun estimasi juga dapat dimanfaatkan dalam prediksi, asalkan digunakan dalam kondisi yang sesuai. Algoritma yang termasuk dalam prediksi adalah *Rough Set, Cart*, ID3, C4.5, J48, dan C.50.

1. Klasifikasi

Pada proses klasifikasi, variabel target yang digunakan bersifat kategorikal. Sebagai ilustrasi, klasifikasi tingkat pendapatan dapat dibedakan menjadi tiga kelompok, yaitu pendapatan tinggi, pendapatan menengah, dan pendapatan rendah. Algoritma yang termasuk dalam klasifikasi adalah *Cart*, ID3, C4.5, H48, C5.0, *nearest neighbor, naïve bayes*, MLP, dan lain-lain.

1. *Clustering*

Klasterisasi adalah proses mengelompokkan data, pengamatan, atau objek-objek ke dalam kelas berdasarkan kemiripan yang dimiliki. Sebuah klaster terdiri dari sekumpulan data yang serupa satu sama lain, namun berbeda dengan data yang terdapat pada klaster lain. Berbeda dengan klasifikasi, klasterisasi tidak melibatkan variabel target. Tujuan dari klasterisasi bukan untuk mengklasifikasikan, memperkirakan, atau memprediksi nilai variabel tertentu, melainkan untuk membagi seluruh kumpulan data menjadi kelompok-kelompok homogen, di mana kemiripan antar data dalam satu kelompok diupayakan setinggi mungkin, sementara kemiripan dengan data di kelompok lain dibuat serendah mungkin. Contohnya pengelompokan siswa yang *eligible* dan yang tidak *eligible* berdasarkan nilai rapor dan prestasi akademis maupun non akademis. Algoritma yang termasuk dalam *Clustering* adalah *K-Means, K-Medoids, K-Nearest Neighbour*, dan lain-lain.

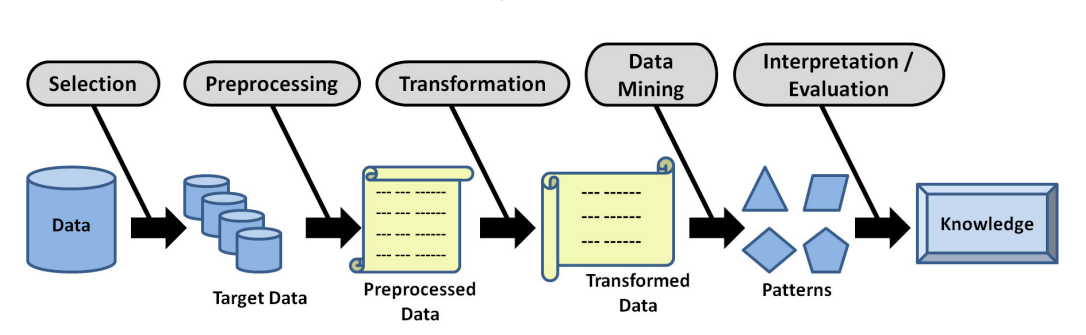
1. Asosiasi

Tugas asosiasi dalam *data mining* bertujuan untuk mengidentifikasi fitur-fitur yang muncul secara bersamaan dalam satu kejadian. Dalam konteks bisnis, proses ini lebih dikenal dengan istilah market basket analysis atau analisis keranjang belanja. Contoh ketika dibeli produk A maka dibeli produk B, ketika dibeli produk B maka dibeli produk A, ketika dibeli produk A dan B, maka dibeli produk C, dan seterusnya. Algoritma yang termasuk dalam asosiasi adalah *Association Rule* [17].

### *Knowledge Discovery in Database*

*Data mining* secara umum selalu dikaitkan dengan bagaimana cara menyelesaikan permasalahan melalui analisis data yang terdapat dalam basis data. *Data mining* juga bisa dikategorikan sebagai bidang ilmu dan teknologi karena kemampuannya dalam menghadirkan solusi baru terhadap permasalahan, melalui beragam teknik penyelesaian yang dimiliki. Seiring dengan pemanfaatan data historis maupun informasi yang tersimpan dalam basis data, banyak peneliti dan pakar literatur menyamakan *data mining* dengan proses *Knowledge Discovery in Databases* (KDD). Bahkan, sebagian menyatakan bahwa *data mining* merupakan inti dari keseluruhan proses KDD [15].

Definisi KDD cukup beragam, namun terdapat satu kesamaan utama yang menjadi ciri khasnya, yaitu bahwa proses ini terdiri dari sejumlah tahapan yang telah disepakati oleh para ahli terkemuka di bidang *data mining*. Berbagai metode telah dikembangkan untuk membagi tahapan ini, masing-masing dengan keunggulan dan kekurangannya. Dalam pembahasan kali ini, proses KDD yang digunakan merupakan hasil adopsi dari berbagai pendekatan yang telah diterapkan sebelumnya, sehingga dapat dirangkum ke dalam beberapa tahap berikut:



Gambar 2. 4 Proses *Knowledge Discovery in Database* (KDD)  
(Sumber: Physics Procedia [18])

1. Identifikasi Tujuan, merupakan langkah awal untuk memahami domain permasalahan yang akan dianalisis serta menetapkan target yang ingin dicapai secara jelas.
2. Spesifikasi Permasalahan, pada tahap ini yang dilakukan adalah merencanakan, merancang, dan mengatur domain aplikasi yang akan digunakan untuk menyelesaikan masalah, sekaligus mendokumentasikan pengetahuan relevan dari para ahli dan menentukan tujuan akhir yang ditujukan untuk pengguna.
3. Seleksi Data, dilakukan untuk memilih data yang relevan dari basis data, dengan mempertimbangkan masukan dari para pakar agar hasil yang diperoleh akurat dan mendalam.
4. *Preprocessing* data, tahap ini meliputi serangkaian operasi seperti pembersihan data dari gangguan dan inkonsistensi, integrasi data dari berbagai sumber, transformasi data, serta reduksi data yang melibatkan pemilihan fitur, ekstraksi fitur, dan penyaringan sampel dalam basis data.
5. *Data mining*, merupakan proses penting di mana metode-metode tertentu digunakan untuk mengekstraksi pola-pola yang valid dari data. Pada tahap ini juga dilakukan pemilihan strategi pemodelan data yang paling sesuai, baik itu klasifikasi, regresi, pengelompokan, maupun asosiasi, serta pemilihan algoritma yang digunakan untuk membangun model tersebut.
6. Pola atau Model, diperoleh dari proses *data mining* yang telah dilakukan melalui strategi-strategi *data mining* yang sesuai dengan permasalahan yang ada.
7. Interpretasi atau evaluasi, merupakan tahapan untuk menilai, mengidentifikasi, dan menafsirkan pola-pola yang benar-benar signifikan dan mencerminkan pengetahuan yang dapat diandalkan, berdasarkan ukuran atau derajat kepentingan tertentu.
8. Pengetahuan, merupakan tahapan terakhir di mana hasil yang diperoleh dapat langsung dimanfaatkan, diintegrasikan ke dalam sistem lain untuk keperluan lanjutan, atau divisualisasikan dan direpresentasikan guna menyampaikan pengetahuan yang ditemukan kepada pengguna secara efektif [15].

### Klasifikasi

Klasifikasi adalah metode dalam *data mining* yang mengelompokkan data ke dalam kategori atau kelas yang telah ditentukan. Klasifikasi termasuk metode pembelajaran supervised learning karena memerlukan data pelatihan untuk membuat aturan yang dapat mengklasifikasikan data uji ke dalam kategori tertentu. Teknik klasifikasi adalah pendekatan terstruktur untuk membangun model klasifikasi dari sekumpulan data masukan. Beberapa contoh teknik klasifikasi meliputi pohon keputusan (*decision tree*), *rule-based classifiers*, jaringan saraf (*neural networks*), *Support Vector Machines*, dan *Naïve Bayes Classifiers* [19].

Model dalam klasifikasi dapat mempunyai arti yang sama dengan *blackbox*, di mana model menerima input, kemudian memprosesnya dan menghasilkan output sebagai hasil dari proses tersebut. Model yang telah dibangun selama pelatihan dapat digunakan untuk memprediksi label kelas dari data baru yang belum diketahui label kelasnya. Dalam proses pembangunan model ini, diperlukan suatu algoritma yang berfungsi untuk membangunnya, yang dikenal dengan istilah algoritma pelatihan [20].

### *Deep Learning*

*Deep Learning* merupakan salah satu cabang dari *Machine Learning* yang menggunakan jaringan saraf tiruan (*Artificial Neural Network* atau ANN) dan dapat dianggap sebagai pengembangan lebih lanjut dari ANN. Topik ini belakangan mendapat perhatian besar di kalangan ilmuwan teknologi informasi. *Deep Learning* berbeda dengan jaringan saraf tiruan tradisional karena menggabungkan beberapa lapisan tersembunyi yang dirancang untuk meningkatkan akurasi *output*. Dalam beberapa tahun terakhir, *Deep Learning* menjadi metodologi yang umum digunakan untuk mengimplementasikan *Machine Learning*, dengan dasar arsitektur yang berbasis jaringan saraf tiruan [21].

### *Multi-Layer Perceptron*

*Multi-Layer Perceptron* (MLP) merupakan salah satu algoritma dari *Artificial Neural Network* (ANN) yang populer dalam aplikasi *machine learning*, terutama dalam tugas klasifikasi dan prediksi. Berikut definisi dari algoritma, pengertian algoritma MLP dan persamaannya.

#### Definisi Algoritma

Algoritma adalah suatu metode atau prosedur yang diterapkan untuk menyelesaikan sebuah masalah. Secara lebih spesifik, algoritma terdiri dari serangkaian langkah-langkah yang terstruktur dan berurutan, yang dirancang untuk mencapai solusi yang diinginkan. Langkah-langkah ini dapat dituangkan dalam bentuk tertulis, baik dalam bentuk catatan tertulis (*hard copy*) maupun dalam bentuk kode program untuk aplikasi komputer. Algoritma memberikan panduan sistematis dalam menyelesaikan masalah dengan cara yang efisien, terorganisir, dan dapat diulang. Setiap langkah dalam algoritma biasanya mencakup operasi atau perhitungan tertentu yang, ketika dijalankan secara berurutan, akan menghasilkan solusi yang tepat sesuai dengan tujuan yang diinginkan [22].

#### Pengertian Algoritma *Multi-Layer Perceptron*

*Multi-Layer Perceptron* (MLP) adalah arsitektur jaringan saraf yang banyak digunakan dan sangat cocok untuk tugas *supervised learning*, termasuk klasifikasi dan regresi. Efektivitasnya terletak pada kemampuannya untuk memodelkan hubungan non-linear yang kompleks dalam data, menjadikannya pilihan ideal untuk tugas-tugas yang memerlukan pemahaman pola yang rumit dan membuat prediksi yang akurat. Neuron dalam sebuah *Perceptron* harus menggunakan fungsi aktivasi seperti ReLU atau sigmoid. Namun, neuron dalam *Multi-Layer Perceptron* dapat menggunakan fungsi aktivasi apa saja. Persamaan fungsi aktivasi ditulis sebagai berikut [11].

Hasil z kemudian dilewatkan ke fungsi aktivasi f(z):

Umumnya fungsi aktivasi yang digunakan adalah ReLu (*Rectified Linear Unit*):

Untuk *output* *layer* (klasifikasi biner), digunakan fungsi sigmoid:

Secara umum, proses per *layer* dituliskan:

Keterangan:

Z : Nilai input total sebelum aktivasi

: Bobot tantara input input I dan neuron ini

: *Input* ke neuron

b : Bias

a (l) : *Output* (aktivasi) pada *layer* ke-l

W(l) : Bobot matriks di *layer* ke-l

B(l) : Bias di *layer* ke-l

a (l-1) : *Output* dari *layer* sebelumnya

## Piranti Pendukung

Piranti pendukung adalah alat yang digunakan untuk membantu penulis dalam menyelesaikan penelitian.

### *Python*

*Python* adalah bahasa pemrograman tingkat tinggi yang mampu mengeksekusi berbagai instruksi secara langsung (interpretatif) dengan pendekatan *Object-Oriented Programming* (OOP). Python juga menggunakan semantik dinamis yang membuat sintaksnya lebih mudah dibaca dan dipahami. Sebagai bahasa pemrograman tingkat tinggi, *Python* dikenal ramah bagi pemula karena didukung dengan fitur manajemen memori otomatis.

Bahasa pemrograman ini pertama kali diperkenalkan pada tahun 1991 oleh Guido van Rossum. Hingga kini, pengembangan *Python* terus berlanjut di bawah naungan *Python Software Foundation*. Python mendukung hampir semua sistem operasi, termasuk Linux, di mana sebagian besar distribusinya sudah menyertakan *Python* secara *default* [5].

### *Jupyter Notebook*

*Jupyter Notebook* adalah salah satu *tools* populer untuk pengolahan data menggunakan bahasa pemrograman *Python*. *Platform* ini memungkinkan integrasi antara kode program dan *output* secara interaktif dalam satu dokumen. *Jupyter* sendiri merupakan organisasi *non-profit* yang fokus mengembangkan perangkat lunak interaktif untuk berbagai bahasa pemrograman. Salah satu produk utamanya adalah *Jupyter Notebook*, sebuah aplikasi *web* *open-source* yang memungkinkan pengguna untuk membuat dan berbagi dokumen interaktif berisi kode yang dapat dijalankan langsung, persamaan matematis, visualisasi, serta teks naratif yang informatif.

*Jupyter Notebook* yang diluncurkan pada tahun 2015, kini telah menjadi salah satu aplikasi andalan dalam bidang analisis data, khususnya di kalangan *data scientist*. Di era saat ini, data memiliki peranan yang sangat penting. Banyak perusahaan membuka peluang karir di bidang pengolahan dan penyajian data untuk mendukung branding perusahaan serta pengambilan keputusan strategis. Arah kebijakan suatu perusahaan kerap kali ditentukan berdasarkan analisis dari data yang dikumpulkan. Hal ini juga berlaku di dunia akademis, di mana publikasi artikel yang disertai data dan visualisasi yang valid menjadi nilai penting dalam keilmuan, termasuk untuk mencegah plagiarisme.

Nama *Jupyter* sendiri berasal dari tiga bahasa pemrograman yang didukungnya, yaitu *Julia* (Ju), *Python* (Py), dan R. Aplikasi ini dapat digunakan secara gratis secara online, memungkinkan pengguna untuk menulis kode, melakukan perhitungan, menyajikan hasil analisis dalam bentuk visualisasi, serta menyusun narasi komputasi secara komprehensif dalam satu *platform* [23].

### *Visual Studio Code*

Visual Studio Code adalah teks editor ringan namun handal yang dikembangkan oleh Microsoft dan mendukung berbagai sistem operasi, termasuk Linux, macOS, dan Windows. Teks editor ini secara langsung mendukung bahasa pemrograman seperti JavaScript, TypeScript, dan Node.js. Selain itu, pengguna dapat memperluas fungsionalitasnya dengan memasang berbagai plugin melalui Visual Studio Code Marketplace, sehingga dapat digunakan untuk bahasa pemrograman lain seperti C++, C#, Python, Go, Java, PHP, dan masih banyak lagi [24].

### *Website*

*Website* merupakan kumpulan halaman yang memuat berbagai informasi digital, seperti teks, gambar, animasi, suara, video, atau kombinasi dari semuanya. Informasi ini disajikan melalui koneksi internet sehingga dapat diakses dan dilihat oleh siapa saja di seluruh dunia. Setiap halaman dalam website dibuat menggunakan bahasa standar yaitu *HyperText Markup Language* (HTML), yang kemudian diterjemahkan oleh *web browser* menjadi tampilan informasi yang mudah dibaca dan dipahami oleh pengguna [25].

### *Streamlit*

*Streamlit* adalah *framework* *Python* yang digunakan untuk membangun aplikasi web dengan antarmuka pengguna interaktid untuk proyek-proyek *data science* dan *machine learning*. Ini memungkinkan pengembang untuk membuat aplikasi *web* khusus dan tampilan data dengan sedikit usaha dan keahlian dalam pengembangan, *web* *Streamlit* dibangun di atas Pustaka ilmu data yang populer seperti *NumPy*, *Pandas*, dan *Matpolib*, sehingga memudahkan untuk membuat visualisasi dan komponen interaktif dalam aplikasi. *Streamlit* juga menyediakan API sederhana untuk *input*  dan *output* data, sehingga mudah untuk terhubung ke berbagai sumber data dan API [26].

# OBJEK DAN METODE PENELITIAN

## Objek Penelitian

Lokasi objek penelitian adalah suatu sarana ilmiah dengan tujuan dan kegunaan tertentu untuk mendapatkan data tertentu. Objek yang diangkat pada penelitian ini beralamat di Jalan Santosa Asih No.17, Cipamokolan, Kec. Rancasari, Kota Bandung, Jawa Barat 40292.



Gambar 3. 1 Logo Kecamatan Rancasari  
(Sumber: https://multisite.bandung.go.id/kecamatan-rancasari/)

Penelitian tersebut dilakukan selama 6 bulan, adapun yang akan dibahas pada objek penelitian ini yaitu mengenai sejarah singkat perusahaan, visi dan misi perusahaan, struktur organisasi perusahaan dan deskripsi tugas di Kantor Kecamatan Rancasari Kota Bandung.

### Sejarah Singkat Kecamatan Rancasari

Kecamatan Rancasari merupakan salah satu kecamatan yang berada di wilayah pemerintahan Kota Bandung, di bentuk berdasarkan PP No. 16 Tahun 1987 Tentang perubahan batas wilayah Kabupaten Bandung dan Kota Madya Bandung. Pada awalnya, wilayah Rancasari merupakan bagian dari Kecamatan Buah Batu. Kemudian, pada tahun 1997, sebagian wilayah Kecamatan Rancasari dimekarkan Kembali menjadi Kecamatan Gedebage. Terhitung dari tanggal 5 April 2007 Kecamatan Rancasari dimekarkan menjadi empat Kelurahan yang terdiri dari Kelurahan Cipamokolan, Derwati, Mekarjaya, dan Manjahlega.

Kecamatan Rancasari memiliki total 52 Rukun Warga (RW) dan 349 Rukun Tetangga (RT). Secara geografis, Kecamatan Rancasari berbatasan dengan Kecamatan Arcamanik di sebelah utara, Kabupaten Bandung di sebelah selatan, Kecamatan Buah Batu di sebelah barat, serta Kecamatan Gedebage di sebelah timur. Kantor Kecamatan Rancasari sendiri dibangun pada tahun 1989 dengan luas tanah sekitar 2.080 m2 dan luas bangunan 695 m2, serta direnovasi pada tahun 2013-2014.

### Visi dan Misi Kecamatan Rancasari

Visi dan misi adalah gambaran arah tujuan dan langkah-langkah yang akan diambil untuk mencapai tujuan tersebut dalam menjalankan tugas dan fungsi pemerintahan di tingkat kecamatan. Berikut adalah visi dan misi Kecamatan Rancasari Kota Bandung.

#### Visi Kecamatan Rancasari

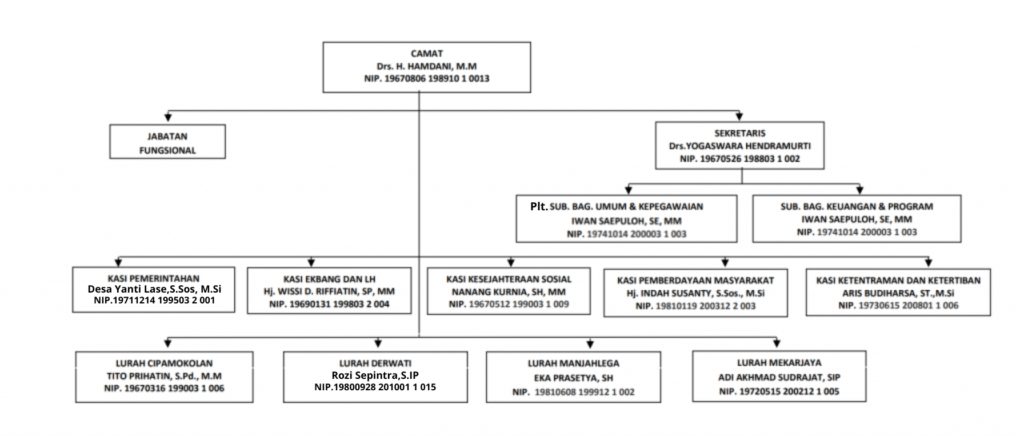
“Mandiri, Agamis, Damai, Aman, Nyaman, Indah”

#### Misi Kecamatan Rancasari

1. Mewujudkan masyarakat yang MANDIRI dalam melaksanakan kegiatan pembangunan di segala bidang melalui pemberdayaan dan swadaya.
2. Mewujudkan masyarakat yang taat menjalankan agama yang pelakunya dengan penuh keimanan.
3. Mewujudkan lingkungan masyarakat yang DAMAI, AMAN, NYAMAN, dan INDAH.

### Struktur Organisasi Kantor Kecamatan Rancasari

Adanya struktur organisasi adalah untuk mengetahui hubungan kerja antar individu dalam sebuah organisasi sehingga dapat menjalankan fungsi bisnisnya dengan baik. Sebuah struktur dapat berisi jabatan, divisi, fungsi, dan lainnya. Berikut struktur organisasi Kantor Kecamatan Rancasari Kota Bandung.



Gambar 3. 2 Struktur Organisasi Kantor Kecamatan Rancasari  
(Sumber: https://multisite.bandung.go.id/kecamatan-rancasari/)

### Deskripsi Tugas

Camat mempunyai tugas melaksanakan kewenangan pemerintahan yang dilimpahkan oleh Walikota untuk menangani sebagian urusan otonomi daerah. Untuk melaksanakan tugas sebagaimana dimaksud, Camat mempunyai fungsi:

1. Pengorganisasian kegiatan pemberdayaan masyarakat;
2. Pengorganisasian upaya penyelenggaraan ketentraman dan ketertiban umum;
3. Pengorganisasian penerapan dan penegakan peraturan perundang-undangan;
4. Pengorganisasian pemeliharaan prasarana dan fasilitas pelayanan umum;
5. Pengorganisasian penyelenggaraan kegiatan pemerintahan di tingkat kecamatan;
6. Pembinaan penyelenggaraan pemerintahan kelurahan;
7. Pelaksanaan pelayanan masyarakat yang menjadi ruang lingkup tugasnya dan atau yang belum dapat dilaksanakan pemerintahan kelurahan;
8. Pengelolaan urusan ketatausahaan;
9. Pelaksanaan tugas lain yang diberikan oleh Walikota sesuai bidang tugas dan fungsinya;
10. Pelaporan dan pertanggungjawaban atas pelaksanaan tugas dan fungsinya kepada Walikota melalui Sekretaris Daerah sesuai standar yang ditetapkan.

Untuk melaksanakan tugas dan fungsi sebagaimana dimaksud, Camat dibantu oleh:

1. Sekretaris Camat

Tugas pokok Sekretaris Camat adalah membantu dalam melaksanakan tugas pemerintahan dan memberikan pelayanan administrasi kepada seluruh perangkat kecamatan.

1. Kepala Seksi Pemerintahan

Tugas pokok Kepala Seksi Pemerintahan adalah membantu Camat dalam menjalankan fungsi tata pemerintahan, pembinaan pemerintahan kelurahan, dan lembaga kemasyarakatan.

1. Kepala Seksi Ketentraman dan Ketertiban

Tugas pokok Kepala Seksi Ketentraman dan Ketertiban adalah membantu Camat dalam menjaga ketentraman dan ketertiban masyarakat, serta melaksanakan tugas-tugas pemerintahan terkait dengan bidang tersebut.

1. Kepala Seksi Kesejahteraan Sosial

Tugas pokok Kepala Seksi Kesejahteraan Sosial adalah membantu Camat dalam perencanaan, pelaksanaan, evaluasi, dan pelaporan urusan kesejahteraan sosial.

1. Kepala Seksi Pembangunan dan Pemberdayaan Masyarakat

Tugas pokok Kepala Seksi Pembangunan dan Pemberdayaan Masyarakat adalah memimpin dan mengkoordinasikan seluruh kegiatan pembangunan dan pemberdayaan masyarakat di wilayah kecamatan.

1. Kelompok Jabatan Fungsional

Tugas pokok kelompok jabatan fungsional adalah melaksanakan sebagian tugas kecamatan secara professional sesuai dengan keahlian dan kebutuhan.

## Metode Penelitian

Metode penelitian adalah kegiatan untuk memperoleh fakta-fakta atau prinsip-prinsip dari suatu pengetahuan dengan cara mengumpulkan, mencatat, dan menganalisa data yang dikerjakan secara sistematis berdasarkan ilmu pengetahuan (metode ilmiah). Metode yang digunakan dalam melaksanakan penelitian merupakan dasar penyusun rancangan secara umum.

### Desain Penelitian

Desain penelitian merupakan kerangka kerja yang digunakan sebagai pedoman dalam melakukan penelitian untuk membantu mengatasi permasalahan yang ada. Dalam penelitian ini, metode penelitian yang digunakan adalah metode kuantitatif. Metode ini digunakan karena melibatkan pengolahan data numerik dan penggunaan algoritma *machine learning* untuk mengklasifikasikan data penerima bantuan sosial PKH.

Desain penelitian dilakukan dengan langkah-langkah terstruktur, guna memprediksi calon penerima bantuan, sehingga dapat mengetahui kepala keluarga yang layak dan tidak layak menerima bantuan sosial PKH. Dengan demikian, proses ini dapat membantu instansi terkait dalam menyalurkan bantuan secara lebih cepat, tepat, dan efisien.

### Jenis dan Metode Pengumpulan Data

Dalam penelitian ini, menggunakan metode pengumpulan data yang meliputi sumber data primer dan data sekunder. Berikut adalah penjelasan mengenai kedua metode tersebut.

#### Sumber Data Primer

Sumber data primer merupakan sumber data yang diperoleh secara langsung dari objek penelitian dan memiliki karakteristik data mentah. Untuk mendapatkan data yang diperlukan, beberapa metode pengumpulan data yang digunakan antar lain:

1. Observasi

Observasi merupakan metode pengamatan yang melibatkan pemusatan pada kegiatan yang sedang berlangsung. Dalam penelitian ini, observasi dilakukan dengan cara pengamatan secara langsung ke lokasi penelitian di Kantor Kecamatan Rancasari untuk mendapatkan data secara umum dan memperoleh informasi yang diperlukan.

1. Wawancara

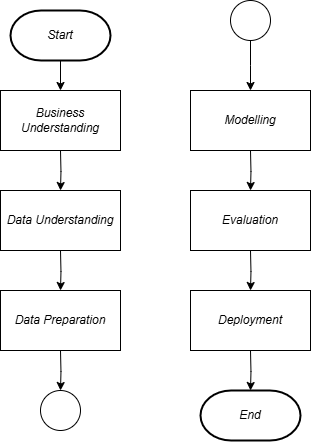
Wawancara adalah teknik pengumpulan data yang melibatkan interaksi tatap muka antara peneliti dan narasumber dari instansi terkait dengan melakukan tanya jawab secara langsung. Dalam penelitian ini, melakukan aksi tanya jawab langsung pada Seksi Kesejahteraan Sosial Kecamatan Rancasari. Hasil dari wawancara dengan narasumber yang didapatkan yaitu sejarah singkat Kecamatan Rancasari, struktur organisasi, visi dan misi, deskripsi tugas masing-masing bagian, serta dataset penerima bantuan sosial PKH tahun 2023.

#### Sumber Data Sekunder

Sumber data sekunder adalah informasi tak langsung dari objek penelitian, untuk metode pengumpulan data sekunder diperoleh dari buku, jurnal, dan dokumen relevan dengan cara memahami, membaca, mempelajari, dan melakukan analisis pengetahuan yang ada.

### Tahapan Penelitian

Dalam pengerjaan tugas skripsi ini, diperlukan tahapan kegiatan penelitian untuk mendapatkan hasil yang maksimal. Berikut tahapan penelitian yang digambarkan melalui diagram sebagai berikut:



Gambar 3. 3 Alir Penelitian  
(Sumber: Jurnal Teknologi Informasi dan Ilmu Komputer [27])

Tahapan penelitian pada diagram diatas menggambarkan alir penelitian secara keseluruhan. Berikut penjelasan dari tahapan-tahapan yang dilakukan.

#### *Business Understanding*

Tahap awal ini berfokus pada pemahaman tujuan dan kebutuhan dari sisi bisnis, yang kemudian di terjemahkan menjadi pengetahuan untuk merumuskan masalah utama yang dapat diatasi melalui teknik *data mining* [27]. Berdasarkan objek penelitian yang telah ditentukan, tahapan pertama dalam penelitian ini dimulai dengan melakukan kegiatan wawancara. Wawancara ini dilakukan baik secara langsung maupun tidak langsung untuk mengidentifikasi permasalahan yang ada serta menemukan solusi yang tepat. Dalam penelitian ini, wawancara dilakukan kepada salah satu pegawai yang bertugas di Seksi Kesejahteraan Sosial Kecamatan Rancasari,

#### *Data Understanding*

Pada tahap ini, peneliti akan melakukan proses memahami karakteristik data yang digunakan dalam penelitian. Tahapan-tahapan yang dilakukan meliputi identifikasi data, memahami kualitas data, mendapatkan wawasan dan memperoleh beberapa hipotesis untuk menemukan informasi tersembunyi di dalam data. Pada tahap ini, beberapa metode akan diterapkan untuk mendukung pemahaman data penerima bantuan sosial PKH Kelurahan Cipamokolan tahun 2023 secara menyeluruh. Statistik deskriptif digunakan untuk menggambarkan karakteristik umum data, seperti nilai rata-rata (*mean*), dan standar deviasi dari setiap fitur. Visualisasi data juga dilakukan untuk mempermudah pemahaman terhadap distribusi data dan membantu dalam proses pembersihan data yang hilang atau menghapus fitur yang bermasalah untuk menghasilkan model pembelajaran mesin yang lebih baik dan lebih umum [27].

#### *Data Preparation*

Pada tahap ini, semua kegiatan difokuskan pada penyusunan dataset akhir penerima bantuan sosial PKH Kelurahan Cipamokolan tahun 2023 yang akan dimasukkan ke dalam model yang dibuat. Persiapan data mencakup berbagai tahapan yang diperlukan untuk membangun kumpulan data yang siap diproses oleh model [27]. Data mentah yang telah diperoleh pada tahap sebelumnya perlu melalui proses *data* *preparation* agar memenuhi standar yang dibutuhkan. Adapun tahapan pengolahan data yang akan dilakukan dalam penelitian ini meliputi beberapa proses sebagai berikut:

1. *Data Selection*

*Data Selection* merupakan proses pemilihan data yang akan digunakan dalam penelitian dengan menerapkan teknik *Feature Subset Selection*. *Feature Subset Selection* adalah metode dalam *feature selection* yang bertujuan untuk menemukan subset fitur terbaik dari seluruh fitur yang tersedia. Subset terbaik adalah kumpulan fitur yang memiliki kontribusi paling signifikan terhadap meningkatan akurasi model, sekaligus mengurangi jumlah dimensi data data yang tidak relevan [28].



Gambar 3. 4 Alir Penerapan Feature Selection  
(Sumber: Jurnal Sistem dan Informatika [28])

Seleksi fitur adalah salah satu teknik *data mining* yang umum digunakan pada tahapan *preprocessing*. Teknik ini digunakan untuk mengurangi kompleksitas fitur yang akan dikelola pada *processing* dan analisis. Teknik ini dilakukan untuk mengetahui *subset* fitur yang paling signifikan dari dataset [28]. Pada penelitian ini, akan digunakan salah satu teknik seleksi fitur yaitu *Correlation Based.*

Teknik *Correlation Based* yang populer digunakan untuk melakukan seleksi fitur yang paling relevan dalam dataset adalah teknik *Corelation Feature Selection* (CFS). Teknik ini menghitung korelasi antara masing-masing fitur dan variabel *output*, lalu memilih fitur yang mempunyai nilai korelasi menengah ke atas (mendekati 1) dan membuang fitur yang memiliki nilai korelasi rendah (mendekati 0). Berdasarkan berbagai penelitian, CFS terbukti dapat menurunkan dimensi dataset secara drastis dengan tetap menjaga atau meningkatkan kinerja dari *learning algorithm* [28].

1. *Data Cleansing*

*Data Cleansing* merupakan proses pembersihan data untuk mengatasi adanya data yang tidak valid, kosong, fitur kurang relevan, *missing value*, maupun *outlier* pada dataset. Proses ini dilakukan menggunakan fungsi “*isna()*” untuk mendeteksi data kosong dan “*fillna()*” untuk mengisi nilai kosong dengan nilai rata-rata (*mean*), median, modus, dan strategi lain yang sesuai dengan karakteristik fitur. Selain itu, proses *data cleansing* juga mencakup penanganan *outlier*. *Outlier* adalah nilai yang jauh berbeda dari nilai lainnya dalam kumpulan data. Pada penelitian ini, proses deteksi *outlier* akan menggunakan metode *Z-Score. Z-Score* merupakan ukuran statistik yang menunjukkan penyimpangan suatu titik data dari rata-rata dalam satuan standar deviasi. Penanganan Outlier dapat dilakukan melalui metode seperti penambahan atau penghapusan nilai agar distribusi data menjadi lebih bersih dan representatif [29].

1. *Data Transformation*

*Data Transformation* merupakan proses mengubah data ke dalam format yang lebih sesuai agar dapat di proses secara optimal oleh sistem menggunakan teknik *data mining*. Pada penelitian ini, proses transformasi dilakukan dengan mengubah fitur yang bersifat kategorikal menjadi bentuk numerik menggunakan metode *Laber Binarizer* dan *One-Hot Encoding*. *Laber Binarizer* merupakan metode yang digunakan untuk mengubah label kategori menjadi nilai biner (0 dan 1). Metode ini digunakan untuk kasus klasifikasi biner atau mengubah target/label yang kategorikal menjadi format yang bisa dibaca oleh algoritma pembelajaran mesin. Sementara itu, *One-Hot Encoding* digunakan untuk mengubah fitur kategori (bukan label) yang memiliki lebih dari dua kategori menjadi representasi numerik yang bisa diproses model. Setiap kategori diubah menjadi vektor biner di mana hanya satu posisi yang bernilai 1 dan sisanya 0[29].

1. *Data Balancing*

*Data Balancing* merupakan proses untuk menyeimbangkan distribusi kelas dalam sebuah dataset. Ketidakseimbangan data terjadi ketika jumlah data untuk satu kelas jauh lebih banyak atau lebih sedikit dibandingkan kelas lainnya [29]. Salah satu metode yang dapat digunakan untuk mengatasi masalah *imbalance class* dalam klasifikasi adalah *Synthetic Minority Oversampling Technique* (SMOTE), SMOTE merupakan teknik *oversampling* yang digunakan untuk menangani ketidakseimbangan distirbusi kelas dalam data. Apabila ketidakseimbangan distribusi kelas tidak ditangani atau diabaikan, maka model yang dibangun cenderung memiliki bias yang tinggi terhadap kelas mayoritas. Hal ini dapat menyebabkan model menjadi tidak peka terhadap prediksi pada kelas minoritas yang mungkin memiliki nilai prediktif yang penting. Akibatnya, akurasi model dapat sangat tinggi karena dominasi kelas mayoritas, tetapi hasil prediksi pada kelas minoritas akan menjadi sangat rendah [30].

1. *Data Splitting*

Pada tahap *Data Splitting*, dataset dibagi menjadi dua bagian, yaitu data latih (*data training*) dan data uji (*data testing*). Pembagian umumnya dilakukan dengan perbandingan 80:20 atau 70:30 [9]. Tujuan utama dari proses ini adalah untuk memastikan bahwa model *machine learning* yang akan dibangun memiliki performa yang baik serta memiliki kemampuan generalisasi yang optimal terhadap data yang belum pernah dilihat sebelumnya [29].

#### *Modelling*

Pada tahap *Modelling,* dilakukan pemilihan dan penentuan model yang akan digunakan dalam penelitian. Tahap ini akan melibatkan kajian terhadap literatur terdahulu untuk mengidentifikasi model prediksi yang berhasil diterapkan [27]. Proses pemilihan model dalam penelitian ini dilakukan dengan mempertimbangkan karakteristik dataset yang telah dikumpulkan. Dataset penerima bantuan sosial PKH Kelurahan Cipamokolan tahun 2023 bersifat tabular dengan label kelas yang sudah ditentukan, sehingga pendekatan yang sesuai adalah menggunakan algoritma dalam kelompok *supervised learning*. Algoritma *supervised learning* dipilih karena proses pelatihannya menggunakan data yang sudah dilabeli, dengan tujuan untuk membangun model yang mampu memprediksi atau mengklasifikasikan data baru berdasarkan pola yang telah dipelajari [19].

Berdasarkan permasalahan yang diangkat dalam penelitian ini, yaitu untuk melakukan klasifikasi terhadap penerima bantuan sosial Program Keluarga Harapan (PKH) ke dalam kategori layak atau tidak layak, maka teknik yang digunakan adalah klasifikasi. Dalam berbagai literatur yang dikaji, metode *Multi-Layer Perceptron* (MLP) diketahui memiliki performa yang baik dalam menyelesaikan masalah klasifikasi, khususnya pada data tabular. MLP merupakan bagian dari algoritma *neural network* yang mampu menangkap hubungan non-linear antar fitur dalam dataset sehingga dapat menghasilkan prediksi yang lebih akurat [9]. Selain itu, MLP juga fleksibel dan telah terbukti memberikan hasil yang optimal pada tugas klasifikasi dalam berbagai penelitian terdahulu. Dengan pertimbangan tersebut, penelitian ini memilih untuk menggunakan metode Multi-Layer Perceptron (MLP) sebagai algoritma utama dalam membangun model klasifikasi penerima bantuan sosial.

Tahap pemodelan bertujuan untuk menghasilkan *Confusion Matrix* untuk mengevaluasi kinerja model dalam masalah klasifikasi. *Confusion Matrix* merupakan sebuah metode untuk evaluasi yang menggunakan tabel matrix seperti pada Tabel 3.1 [31].

Tabel 3. 1 Confusion Matrix  
(Sumber: Infotech Journal [31])

|  |  |  |
| --- | --- | --- |
|  | **Prediksi** | |
| **Aktual** | *True* | *False* |
| *True* | *TP* | *FP* |
| *False* | *FN* | *TN* |

*Confusion Matrix* merupakan salah satu alat yang umum digunakan untuk mengukur kinerja model, khususnya dalam tugas klasifikasi. Selain sebagai metrik evaluasi, *confusion matrix* juga berfungsi sebagai alat visual yang membantu dalam menilai hasil klasifikasi yang dilakukan oleh model. Matriks merupakan indikator performa untuk masalah klasifikasi yang outputnya bisa lebih dari satu kelas [31].

*Confusion Matrix* disusun dalam bentuk tabel yang berisi empat kombinasi dari hasil prediksi model dan kondisi aktual data. Beberapa istilah utama yang terdapat dalam confusion matrix meliputi:

1. *True Positive* (TP)

Menyatakan jumlah data yang sebenarnya termasuk dalam kelas positif dan berhasil diprediksi dengan benar oleh model sebagai positif. Dengan kata lain, ini adalah kondisi ideal di mana model mampu mengenali data positif secara akurat.

1. *True Negative* (TN)

Menyatakan jumlah data yang benar-benar termasuk dalam kelas negatif dan juga diprediksi dengan tepat sebagai negatif oleh model. Hal ini menunjukkan bahwa model mampu menghindari kesalahan klasifikasi terhadap data negatif.

1. *False Positive* (FP)

Disebut sebagai kesalahan tipe I, di mana data yang seharusnya masuk dalam kelas negatif malah diprediksi sebagai kelas positif oleh model. Misalnya, pada klasifikasi penerima bantuan, kesalahan ini berarti seseorang yang tidak layak malah dinyatakan layak.

1. *False Negative* (FN)

Disebut sebagai kesalahan tipe II, di mana data yang seharusnya termasuk dalam kelas positif malah diprediksi sebagai negatif.. Misalnya pada klasifikasi penerima bantuan, kesalahan ini berarti orang yang sebenarnya layak justru tidak mendapatkan bantuan.

1. Nilai Prediksi

Merupakan hasil keluaran dari model klasifikasi yang mengelompokkan data ke dalam kategori tertentu, dalam hal ini ke dalam kelas positif atau negatif.

1. Nilai Aktual

Merupakan nilai sebenarnya dari data yang mencerminkan kondisi sesungguhnya. Nilai ini digunakan sebagai acuan untuk mengevaluasi sejauh mana prediksi model sesuai dengan kenyataan. Pada *confusion matrix*, nilai aktual ini dihadapkan dengan nilai prediksi untuk mengukur tingkat akurasi dan kesalahan model [31].

*Confusion Matrix* berperan dalam memberikan pemahaman tentang sejauh mana model mampu mengidentifikasi dan mengklasifikasikan setiap kasus dengan tepat maupun keliru, serta menjadi acuan dalam melakukan penyesuaian dan penyempurnaan model berdasarkan hasil evaluasi.

#### *Evaluation*

Pada tahap *Evaluation*, model *Multi-Layer Perceptron* (MLP) untuk klasifikasi penerima bantuan sosial PKH di Kelurahan Cipamokolan akan dievaluasi dan dianalisis untuk memastikan bahwa model yang dipilih memiliki kualitas terbaik dalam mencapai tujuan dari masalah bisnis yang ada. Evaluasi didasarkan pada nilai akurasi, *recall*, presisi, dan *f1-score* yang dihasilkan oleh model [27]. Pendekatan ini memungkinkan analisis meneyeluruh terhadap efektivitas berbagai konfigurasi dan fungsi aktivasi dalam memprediksi penerima bantuan [32].

*Confussion Matrix* yang merupakanhasil dari tahapan pemodelan digunakan untuk menghitung nilai akurasi, presisi, *recall*, dan *F1-Score* [33].

1. *Accuracy*

Akurasi digunakan untuk mengukur proporsi prediksi yang benar dibandingkan dengan total prediksi, dihitung dengan rumus:

Nilai akurasi menggambarkan seberapa akurat model dalam melakukan klasifikasi dengan benar [33].

1. *Precision*

Presisi digunakan untuk mengukur proporsi prediksi positif yang benar dibandingkan dengan total prediksi positif, dihitung dengan rumus:

Presisi menggambarkan akurasi antara data yang diminta dengan hasil prediksi yang diberikan oleh model [33].

1. *Recall*

*Recall* atau sensitivitas digunakan untuk mengukur proporsi kasus positif yang benar-benar terdeteksi dibandingkan dengan total kasus positif yang sebenarnya, dihitung dengan rumus:

*Recall* menggambarkan keberhasilan model dalam menemukan kembali sebuah informasi [33].

1. *F1-Score*

*F1-Score* merupakan perbandingan rata-rata presisi dan *recall* yang dibobotkan [33].

Dengan menggunakan *accuracy, precision, recall,* dan *F1-Score* dapat melakukan evaluasi terhadap performa model dalam mengklasifikasikan penerima bantuan sosial PKH di Kelurahan Cipamokolan.

#### *Deployment*

Pada tahap *Deployment*, model Multi-Layer Perceptron (MLP) untuk klasifikasi penerima bantuan sosial PKH di Kelurahan Cipamokolan yang telah dibangun dan dilatih akan diintegrasikan ke dalam aplikasi *web* untuk memberikan layanan prediksi secara langsung kepada pengguna [27]. Dalam penelitian ini, model yang telah diterapkan untuk klasifikasi penerima bantuan sosial PKH akan diimplementasikan ke dalam sebuah aplikasi berbasis *web* menggunakan *framework Streamlit*. Tujuan utama dari aplikasi ini adalah untuk membantu pengguna untuk menginput data calon penerima bantuan sosial dan mendapatkan prediksi apakah mereka layak atau tidak layak menerima bantuan berdasarkan kriteria yang telah ditetapkan.

## Analisis Sistem yang Berjalan

Analisis terhadap sistem yang berjalan bertujuan untuk memahami tahapan-tahapan yang dilakukan.

### *Business Understanding*

Berdasarkan hasil wawancara yang dilakukan dengan salah satu pegawai Seksi Kesejahteraan Sosial Kecamatan Rancasari, diperoleh pemahaman lebih mendalam mengenai proses pelaksanaan program Bantuan Sosial Program Keluarga Harapan (PKH) di lapangan. Wawancara ini bertujuan untuk mengidentifikasi permasalahan yang terjadi, memahami alur data, serta mengkaji kriteria yang digunakan dalam menentukan keluarga yang layak menerima bantuan sosial. Berikut merupakan rangkuman dari pertanyaan-pertanyaan yang diajukan beserta jawaban dari narasumber:

1. Apa permasalahan yang terjadi dalam klasifikasi calon penerima bantuan?

Narasumber menjelaskan bahwa proses klasifikasi calon penerima bantuan sosial PKH hingga saat ini masih dilakukan secara manual, yaitu melalui forum musyawarah di tingkat RT, RW, kelurahan, dan kecamatan. Proses ini belum memanfaatkan teknologi prediktif atau sistem klasifikasi berbasis data, sehingga rawan keterlambatan pembaruan data dan potensi ketidaktepatan dalam penentuan penerima bantuan.

1. Bagaimana sumber data didapatkan?

Data penerima bantuan sosial PKH didapatkan melalui sensus penduduk yang dilakukan pemerintah, dan data tersebut dikelola oleh pihak kelurahan. Untuk keperluan penelitian ini, dataset penerima bantuan sosial PKH di Kelurahan Cipamokolan tahun 2023 diperoleh melalui koordinator PKH Kecamatan Rancasari yang memiliki data resmi terkait keluarga penerima manfaat di wilayah tersebut.

1. Bagaimana jika ada perubahan data, misalnya perubahan pendidikan dari SD menjadi SMP?

Perubahan data dilakukan dengan cara keluarga mengajukan permohonan perubahan kepada ketua RT atau RW setempat. Selanjutnya, perubahan tersebut akan dimusyawarahkan bersama pihak kelurahan, kemudian dilaporkan ke kecamatan untuk diteruskan ke Dinas Sosial guna pembaruan data secara resmi.

1. Apa saja kriteria dalam menentukan penerima bantuan?

Kriteria penerima bantuan sosial PKH mengacu pada Peraturan Menteri Sosial (Permensos) Nomor 1 Tahun 2018, yang menetapkan 8 kelompok sasaran utama, yaitu: ibu hamil atau menyusui, balita, anak prasekolah, anak sekolah (tingkat SD, SMP, dan SMA), lanjut usia, serta penyandang disabilitas berat.

1. Bagaimana hubungan antar kriteria dengan fitur lainnya agar memenuhi syarat untuk menerima bantuan?

Narasumber menjelaskan bahwa kriteria penerima bantuan PKH memiliki hubungan erat dengan sejumlah fitur atau fitur lain seperti usia, penghasilan, jenis pekerjaan, status pernikahan, jumlah tanggungan, tingkat pendidikan, dan status kepemilikan rumah. Misalnya, usia berpengaruh dalam menentukan kategori penerima, seperti lansia yang berusia 60 tahun ke atas. Penghasilan dan jenis pekerjaan juga menjadi indikator tingkat kesejahteraan ekonomi, di mana pekerjaan tidak tetap seperti buruh harian dikategorikan sebagai keluarga miskin atau rentan miskin. Status pernikahan mempengaruhi beban tanggungan keluarga. Mereka yang sudah menikah biasanya memiliki anak atau tanggungan lain yang menambah kebutuhan hidup. Selain itu, status seperti belum menikah, cerai hidup, atau cerai mati juga mempengaruhi kondisi ekonomi keluarga. Jumlah tanggungan menjadi pertimbangan langsung, karena semakin banyak anggota keluarga yang ditanggung, semakin tinggi kebutuhan hidup keluarga tersebut. Tingkat pendidikan berhubungan dengan potensi pendapatan, di mana keluarga dengan pendidikan rendah lebih sulit mendapatkan pekerjaan yang layak dengan penghasilan tinggi. Sementara itu, status kepemilikan rumah mencerminkan kondisi ekonomi keluarga.

1. Apa hubungan antara jenis pekerjaan dengan kelayakan menerima bantuan?

Sebagai contoh, kategori buruh yang dimaksud mencakup buruh harian, buruh bangunan, dan buruh lepas. Faktor ekonomi dari pekerjaan ini menjadi pertimbangan dalam penentuan penerima bantuan.

1. Apa saja jenis pekerjaan yang termasuk dalam kategori pedagang dan pekerjaan lainnya?

Jenis pekerjaan dalam kategori pedagang antara lain pedagang sayur dan pedagang bakso, sedangkan untuk kategori pekerjaan lainnya seperti buruh, terdiri dari buruh harian, buruh bangunan, dan buruh lepas.

Alir penerimaan bantuan sosial PKH yang teridentifikasi dari hasil wawancara ini dapat dilihat pada Gambar 3.5 berikut:



Gambar 3. 5 Alir PKH  
(Sumber: Jurnal Teknologi Informasi dan Ilmu Komputer [29])

Berdasarkan Gambar 3.5 terdapat alir dalam penerimaan PKH yaitu sebagai berikut:

1. Musyarawah RT dan RW

Mengadakan musyawarah untuk mengidentifikasi keluarga keluraga yang layak mendapatkan bantuan sosial PKH sesuai dengan syarat kriteria penerima bantuan sosial PKH sebagaimana diatur dalam Peraturan Menteri Sosial Republik Indonesia Nomor 1 tahun 2018. Data hasil musyawarah berupa daftar usulan calon penerima.

1. Musyawarah Kelurahan

Data dari RT/RW dibawa ke tingkat kelurahan untuk dibahas lebih lanjut untuk memastikan bahwa calon penerima yang diusulkan sudah tepat sasaran berdasarkan kriteria nasional. Hasilnya berupa daftar calon penerima yang sudah disepakati bersama.

1. Verifikasi Kecamatan

Pihak Kecamatan akan memeriksa kelengkapan administrasi dan kebenaran data penerima. Kecamatan bertugas memastikan bahwa data yang diajukan memang sesuai dengan ketentuan program PKH.

1. Verifikasi Dinas Sosial

Dinas Sosial melakukan validasi akhir terhadap data hasil verifikasi kecamatan dengan mengecek apakah calon penerima sudah sesuai dengan Data Terpadu Kesejahteraan Sosial (DTKS) dan kriteria lain yang berlaku. Setelah verifikasi, calon penerima masuk sebagai peserta PKH dan berhak mendapatkan bantuan.

Proses musyawarah dalam menentukan kelayakan calon penerima bantuan sosial PKH dilakukan dengan mempertimbangkan berbagai fitur yang merujuk pada teori serta aturan yang berlaku. Berikut penjelasannya dari masing-masing fitur:

1. Usia

Usia kepala keluarga menjadi salah satu pertimbangan dalam proses musyawarah. Berdasarkan UU No. 11 Tahun 2009 tentang Kesejahteraan Sosial, kepala keluarga yang telah berusia 60 tahun ke atas dikategorikan rentan dan menjadi prioritas dalam menerima bantuan. Sementara itu, usia produktif antara 18 – 59 tahun juga dapat dipertimbangkan layak menerima bantuan apabila termasuk dalam kelompok miskin atau rentan miskin.

1. Penghasilan

Menurut Peraturan Menteri Sosial No. 20 Tahun 2019 tentang Penetapan Keluarga Miskin, salah satu kriteria utama kelayakan adalah penghasilan bulanan kepala keluarga yang berada di bawah garis kemiskinan. Penghasilan dibawah Rp500.000 hingga Rp1.000.000 per bulan menjadi indikator utama dalam menentukan kelayakan.

1. Jenis Pekerjaan

Jenis pekerjaan kepala keluarga juga menjadi salah satu acuan dalam menentukan penerima PKH. Calon penerima yang bekerja di sektor informal seperti buruh harian, petani kecil, atau bahkan tidak memiliki pekerjaan tetap, cenderung di prioritaskan. Sementara itu, mereka yang bekerja sebagai ASN, TNI/Polri, atau memiliki pekerjaan dengan penghasilan tetap, tidak termasuk dalam kategori yang layak menerima bantuan.

1. Status Pernikahan

Menurut Badan Pusat Statistik (BPS), keluarga dengan kepala keluarga yang menikah dan memiliki anak, ataupun kepala keluarga yang berstatus janda atau duda dengan tanggungan, dianggap berada dalam kondisi sosial ekonomi yang lebih rentan sehingga lebih membutuhkan bantuan.

1. Jumlah Tanggungan

Jumlah anggota keluarga yang menjadi tanggungan berperan penting dalam menentukan tingkat kelayakan. Berdasarkan Survei Sosial Ekonomi Nasional (Susenas) dan Badan Pusat Statistik (BPS), keluarga dengan tanggungan lebih banyak terutama anak usia sekolah, lansia, atau penyandang disabilitas memiliki peluang lebih besar untuk memenuhi kriteria sebagai penerima bantuan.

1. Tingkat Pendidikan

Pendidikan terakhir kepala keluarga mencerminkan kondisi ekonomi keluarga. Berdasarkan UU No. 20 Tahun 2003 tentang Sistem Pendidikan Nasional, kepala keluarga yang hanya menyelesaikan pendidikan hingga tingkat SD atau SMP, atau bahkan tidak menyelesaikan pendidikan, cenderung lebih rentan terhadap kemiskinan.

1. Status Rumah

Berdasarkan Peraturan Pemerintah No. 25 Tahun 2000 tentang Program Nasional Pemberdayaan Masyarakat (PNPM), keluarga yang tidak memiliki rumah sendiri, menumpang, atau menyewa dikategorikan sebagai prioritas untuk menerima bantuan.

Alir proses justifikasi data penerima bantuan sosial PKH merupakan alir proses yang dibuat berdasarkan alir PKH yang teridentifikasi dari hasil wawancara.



Gambar 3. 6 Alir Proses Justifikasi Data  
(Sumber: Jurnal Teknologi Informasi dan Ilmu Komputer [29])

Berikut penjelasan dari Alir Proses Justifikasi Data berdasarkan Gambar 3.6 diatas:

1. Identifikasi Data Calon Penerima

Data calon penerima dikumpulkan dari berbagai sumber, seperti data dari pemerintah desa/kelurahan dan data terpadu kesejahteraan sosial (DTKS). Data yang terkumpul meliputi informasi demografi (nama, alamat, tanggal lahir, NIK), informasi sosial ekonomi (pekerjaan dan penghasilan), dan informasi terkait kriteria PKH (ibu hamil/menyusui balita, anak prasekolah, anak sekolah, lansia, dan anggota keluarga disabilitas).

1. Verifikasi Awal Data

Melakukan pengecekan awal terhadap kelengkapan dan kevalidan data calon penerima, data yang tidak lengkap atau terindikasi tidak valid akan dikembalikan untuk dilengkapi.

1. Validasi Data Lapangan

Tim verifikasi melakukan kunjungan langsung ke lapangan (rumah calon penerima). Tujuan verifikasi lapangan adalah untuk memastikan kondisi calon penerima sesuai dengan data yang telah dikumpulkan. Proses ini melibatkan wawancara dengan calon penerima dan observasi kondisi tempat tinggal serta aset yang dimiliki.

1. Analisis dan Penilaian Data

Data hasil verifikasi lapangan dianalisis dan dibandingkan dengan kriteria penerima PKH yang telah ditetapkan oleh pemerintah saat musyawarah. Penilaian dilakukan untuk menentukan apakah calon penerima memenuhi syarat untuk menerima bantuan PKH.

1. Justifikasi dan Penetapan Penerima

Hasil analisis dan penilaian data dari musyawarah diajukan untuk proses justifikasi. Tim justifikasi (yang terdiri dari pihak terkait, seperti perwakilan dinas sosial, pemerintah daerah, dan pendamping PKH) melakukan peninjauan akhir terhadap data calon penerima yang lolos verifikasi. Justifikasi memastikan bahwa calon penerima yang ditetapkan benar-benar memenuhi kriteria dan layak menerima bantuan. Setelah proses justifikasi selesai, daftar penerima bantuan PKH ditetapkan secara resmi.

1. Pengumuman

Daftar penerima bantuan PKH yang telah ditetapkan diumumkan kepada masyarakat, melalui pemerintah desa/kelurahan.

1. Pemutakhiran Data Berkala

Data penerima PKH tidak bersifat tetap. Pemutakhiran data secara berkala dilakukan untuk memastikan bantuan tepat sasaran. Proses pemutakhiran dapat melibatkan verifikasi ulang, laporan perubahan kondisi penerima, atau mekanisme umpan balik dari masyarakat. Jika ditemukan penerima yang tidak lagi memenuhi kriteria, status kepesertaannya dapat ditinjau kembali.

Berdasarkan hasil wawancara, dapat disimpulkan bahwa adanya proses manual dalam verifikasi data calon penerima bantuan berpotensi menimbulkan keterlambatan dan ketidaktepatan dalam penyaluran bantuan. Oleh karena itu, diperlukan suatu sistem klasifikasi yang dapat membantu menentukan keluarga yang berhak menerima bantuan sosial PKH.

### *Data Understanding*

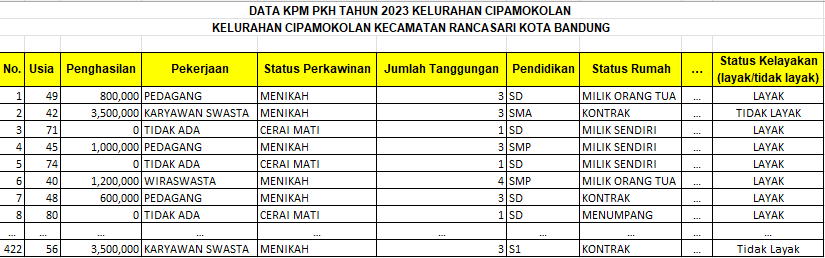
Pada tahap ini, dilakukan proses pemahaman terhadap data yang digunakan dalam penelitian.

#### Sumber

Sumber data primer yang digunakan dalam penelitian ini merupakan data penerima bantuan sosial Program Keluarga Harapan (PKH) tahun 2023 yang ada di Kelurahan Cipamokolan. Data diperoleh secara langsung melalui proses wawancara dan koordinasi dengan koordinator PKH Kecamatan Rancasari. Wawancara tersebut dilakukan bersama salah satu pegawai yang bertugas di Seksi Kesejahteraan Sosial Kecamatan Rancasari.

#### Validator

Data yang digunakan dalam penelitian ini merupakan data yang valid dan telah melewati proses verifikasi serta validasi resmi. Data tersebut diberikan langsung oleh Koordinator PKH Kecamatan Rancasari, yang sebelumnya diperoleh dari hasil pendataan yang dimiliki oleh pihak Kelurahan Cipamokolan. Kemudian data tersebut dimusyawarahkan oleh RT, RW, dan pihak Kelurahan untuk memastikan akurasi dan kelayakan data calon penerima bantuan. Selanjutnya, data tersebut diverifikasi dan divalidasi lebih lanjut oleh pihak Kecamatan Rancasari dan Dinas Sosial. Adapun data yang diperoleh dapat dilihat pada Gambar 3.7.



Gambar 3. 7 Data PKH Tahun 2023  
(Sumber: Kantor Kelurahan Cipamokolan)

Data pada Gambar 3.7 merupakan data yang masih mentah yang berjumlah sebanyak 422 *record* dan terdiri dari 15 fitur yaitu Usia, Penghasilan, Pekerjaan, Status Pernikahan, Jumlah Tanggungan, Pendidikan, Status Rumah, Jenis Kelamin, Kriteria PKH, dan Status Kelayakan. Untuk kriteria PKH terdiri dari 6 fitur yaitu Ibu Hamil/Menyusui, Balita, Anak Prasekolah, Anak Sekolah (SD/Sederajat, SMP/Sederajat, SMA/Sederajat), Lansia, Penyandang Disabilitas.

#### Keterikatan Antar Fitur

Berdasarkan data yang diperoleh, setiap fitur yang digunakan memiliki keterikatan atau hubungan yang saling mempengaruhi dengan kriteria PKH dalam menentukan kelayakan penerima bantuan sosial PKH. Berikut penjelasan keterikatan antar fitur berdasarkan hasil wawancara:

1. Usia

Usia kepala keluarga mempengaruhi kemampuan dalam mencari penghasilan. Oleh karena itu, usia menjadi salah satu indikator penting dalam menilai kelayakan penerima bantuan.

1. Penghasilan

Penghasilan merupakan salah satu indikator utama dalam penilaian kesejahteraan ekonomi masyarakat. Pada data yang diperoleh, penghasilan merepresentasikan jumlah pendapatan yang diterima oleh masing-masing kepala keluarga.

1. Pekerjaan

Jenis pekerjaan berhubungan erat dengan tingkat ekonomi keluarga. Pada data yang diperoleh, kategori pekerjaan meliputi asisten rumah tangga, karyawan swasta, bertani, buruh, pedagang, wiraswasta, pegawai, dan peternak. Pekerjaan seperti buruh dan asisten rumah tangga umumnya masuk ke dalam kategori masyarakat berpenghasilan rendah sehingga menjadi faktor kelayakan penerima bantuan sosial.

1. Status Pernikahan

Status pernikahan berpengaruh terhadap jumlah tanggungan dalam keluarga. Misalnya, keluarga dengan cerai hidup atau cerai mati memiliki beban ekonomi yang lebih besar sehingga menjadi faktor kelayakan penerima bantuan. Pada penelitian ini, kategori status pernikahan meliputi menikah, cerai mati, cerai hidup, dan belum menikah.

1. Jumlah Tanggungan

Semakin banyak jumlah tanggungan, maka semakin tinggi kebutuhan ekonomi sehari-hari. Oleh karena itu, keluarga dengan tanggungan lebih banyak menjadi salah satu faktor kelayakan penerima bantuan.

1. Pendidikan

Tingkat pendidikan kepala keluarga mempengaruhi status sosial ekonomi. Pada data yang diperoleh, kategori pendidikan meliputi SD, SMP, SMA, dan S1. Tingkat pendidikan menjadi salah satu faktor perhitungan kelayakan penerima bantuan.

1. Status Rumah

Kategori status rumah meliputi milik sendiri, kontrak, menumpang, bebas sewa, dan milik orang tua dapat mencerminkan tingkat kesejahteraan keluarga. Oleh karena itu, Status rumah menjadi salah satu faktor dalam menentukan kelayakan penerima bantuan.

Fitur-fitur yang sudah dijelaskan saling mempengaruhi dengan kriteria PKH. Oleh karena itu, keterikatan antar fitur menjadi dasar dalam proses klasifikasi untuk menentukan apakah sebuah keluarga layak atau tidak menerima bantuan sosial PKH.

#### Statistik Deskriptif

Statistik deskriptif dibuat untuk memberikan gambaran umum mengenai karakteristik data, serta hubungan antar fitur yang relevan. Untuk fitur numerik, ditampilkan ringkasan statistik yang berisi nilai rata-rata (*mean*), standar deviasi, nilai ninimum, dan nilai maksimum. Ringkasan ini bertujuan untuk mengetahui sebaran dan distribusi data numerik usia, penghasilan, dan jumlah tanggungan.

Tabel 3. 2 Statistik Deskriptif Fitur Numerik

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Usia** | **Penghasilan (Rp.)** | **Jumlah Tanggungan** |
| ***Count*** | 422 | 422 | 422 |
| ***Mean*** | 55,72 | 1.117.606 | 2,33 |
| ***STD*** | 15,18 | 818.362 | 1,11 |
| ***Min*** | 22 | 0 | 1 |
| ***Max*** | 89 | 4.500.000 | 6 |

Berikut perhitungan mencari nilai rata-rata (x̄) pada fitur usia, penghasilan, dan jumlah tanggungan:

Usia (x̄) =

Penghasilan (x̄) =

Jumlah tanggungan (x̄) =

Berikut perhitungan mencari nilai standar deviasi (σ) pada fitur usia, penghasilan, dan jumlah tanggungan:

Usia (σ) = 2 = 2 = 15,18

Penghasilan (σ) = 2

Penghasilan (σ) = 2 = 818.362

Jumlah tanggungan (σ) = 2

Jumlah tanggungan (σ) = 2 = 1,11

Berdasarkan Tabel 3.2 dapat dilihat bahwa rata-rata usia kepala keluarga adalah 55 tahun, dengan rata-rata penghasilan Rp1.117.606 dan jumlah tanggungan 2 orang. Standar deviasi 15 tahun menunjukkan bahwa usia kepala keluarga tersebar dalam rentang 40 sampai 70 tahun.

Tabel 3. 3 Statistik Deskriptif Fitur Kategorikal

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Pekerjaan** | **Status Pernikahan** | **Pendidikan** | **Status Rumah** | **L/P** | **Status Kelayakan** |
| ***Count*** | 422 | 422 | 422 | 422 | 422 | 422 |
| ***Unique*** | 9 | 4 | 4 | 5 | 2 | 2 |
| ***Top*** | Tidak ada | Menikah | SMA | Milik Sendiri | L | Layak |
| ***Freq*** | 142 | 254 | 138 | 105 | 217 | 236 |

Tabel 3. 4 Statistik Deskriptif Fitur Kategorikal Lanjutan

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Ibu Hamil /Menyusui** | **Balita** | **Anak Prasekolah** | **Anak Sekolah** | **Lansia** | **Penyandang Disabilitas** |
| ***Count*** | 422 | 422 | 422 | 422 | 422 | 422 |
| ***Unique*** | 2 | 2 | 2 | 2 | 2 | 2 |
| ***Top*** | Tidak | Tidak | Tidak | Tidak | Tidak | Tidak |
| ***Freq*** | 248 | 292 | 323 | 362 | 319 | 259 |

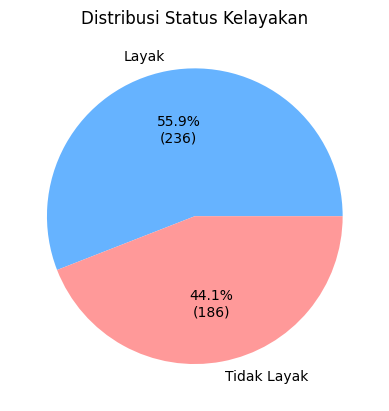
Berdasarkan Tabel 3.3 dan 3.4 dapat diketahui bahwa statistik deskriptif fitur kategorikal menghitung frekuensi atau jumlah kemunculan setiap kategori. Nilai *unique* menunjukkan banyaknya kategori yang terdapat dalam sebuah fitur. Pada fitur pekerjaan terdapat 9 kategori, yaitu asisten rumah tangga, karyawan swasta, pedagang, bertani, buruh, wiraswasta, peternak, pegawai, dan belum memiliki pekerjaan. Dari data tersebut, kepala keluarga yang tidak memilik pekerjaan tercatat sebagai jumlah yang paling banyak, dengan total 142 kepala keluarga. Pada fitur status pernikahan terdapat 4 kategori, yaitu menikah, cerai mati, cerai hidup, dan belum menikah. Mayoritas kepala keluarga berstatus menikah, dengan jumlah 254 kepala keluarga. Untuk fitur tingkat pendidikan memiliki 4 kategori, yaitu SD, SMP, SMA, dan S1. Tingkat pendidikan SMA menjadi jumlah yang paling banyak dengan total 138 kepala keluarga. Sementara itu, pada fitur status rumah terdapat 5 kategori, yaitu milik orang tua, kontrak, milik sendiri, menumpang, dan bebas sewa. Kepala keluarga dengan status rumah milik sendiri memiliki jumlah yang paling banyak dengan total 105 kepala keluarga.

Selanjutnya, fitur kriteria PKH yaitu ibu hamil/menyusui, balita, anak prasekolah, anak sekolah, lansia, dan penyandang disabilitas masing-masing terdapat 2 kategori, yaitu kategori Ya dan kategori Tidak. Untuk fitur status kelayakan juga terdapat 2 kategori, yaitu status layak dan tidak layak, di mana dari 422 *record* data sebanyak 236 *record* merupakan layak menerima bantuan. Dengan mengetahui distribusi tiap kategori, dapat diketahui dominasi atau proporsi antar kelompok dalam dataset.

Tabel 3. 5 Statistik Deskriptif Hubungan Antar Pekerjaan Dengan Fitur Numerik

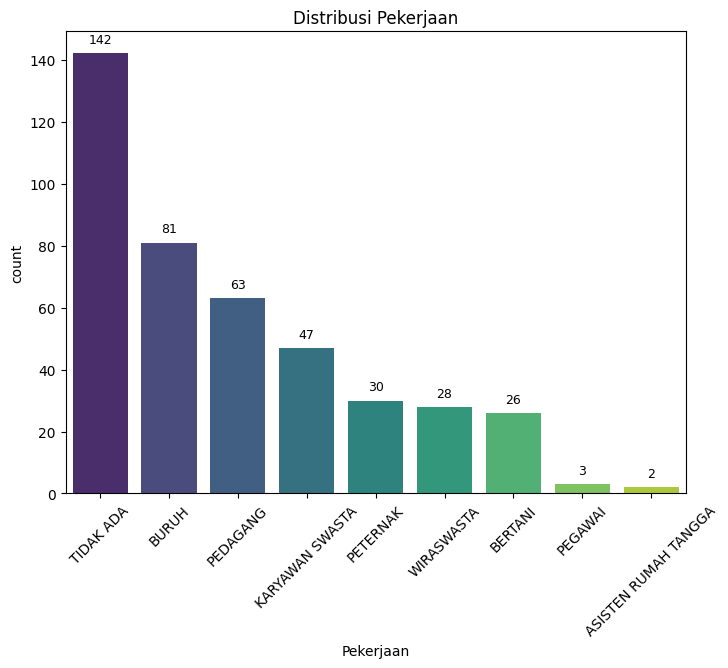
| Status Kelayakan | Pekerjaan | Jumlah | Usia Rata-rata | Std Usia | Min Usia | Max Usia | Rata Rata Tanggungan |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Layak | Asistem Rumah Tangga | 2 | 51.5 | 2.12 | 50 | 53 | 2.5 |
| Layak | Bertani | 5 | 69.4 | 8.76 | 58 | 82 | 2.0 |
| Layak | Buruh | 81 | 53.65 | 14.14 | 22 | 89 | 2.35 |
| Layak | Karywan Swasta | 8 | 54.25 | 8.08 | 39 | 62 | 3.0 |
| Layak | Pedagang | 8 | 50.75 | 12.94 | 40 | 80 | 3.0 |
| Layak | Pegawai | 3 | 47.33 | 7.64 | 39 | 54 | 3.0 |
| Layak | Peternak | 1 | 78.0 | 0 | 78 | 78 | 3.0 |
| Layak | Tidak Ada | 117 | 61.11 | 13.43 | 27 | 89 | 2.32 |
| Layak | Wiraswasta | 17 | 48.12 | 14.21 | 26 | 72 | 2.45 |
| Tidak Layak | Bertani | 21 | 55.33 | 15.79 | 27 | 79 | 2.43 |
| Tidak Layak | Karyawan Swasta | 39 | 51.74 | 17.43 | 26 | 80 | 2.13 |
| Tidak Layak | Pedagang | 55 | 52.87 | 14.9 | 25 | 78 | 2.55 |
| Tidak Layak | Peternak | 29 | 57.24 | 18.6 | 27 | 80 | 2.28 |
| Tidak Layak | Tidak Ada | 25 | 51.56 | 17.15 | 25 | 77 | 2.08 |
| Tidak Layak | Wiraswasta | 17 | 48.12 | 14.21 | 26 | 72 | 2.47 |

Berdasarkan Tabel 3.5 terlihat bahwa pada fitur pekerjaan dengan kategori buruh dan tidak memiliki pekerjaan dengan status kelayakan layak memiliki jumlah terbanyak. Untuk kategori buruh yang layak menerima bantuan tercatat sebanyak 81 kepala keluarga. Sementara itu, kategori tidak memiliki pekerjaan yang layak menerima bantuan sebanyak 117 kepala keluarga. Analisis ini bertujuan untuk menggambarkan karakteristik demografis tiap kategori pekerjaan yang menjadi penerima bantuan sosial PKH.



Gambar 3. 8 *Chart* Status Kelayakan

Visualisasi seperti diagram batang pada Gambar 3.8 disajikan untuk memperjelas distribusi data. Diagram batang tersebut merupakan diagram status kelayakan yang menunjukkan jumlah data pada kategori layak dan tidak layak. Berdasarkan visualisasi tersebut terlihat bahwa 55.9% dari total kepala keluarga pada data memiliki status layak. Visualisasi ini membantu dalam mengidentifikasi apakah terdapat ketidakseimbangan antara kelas-kelas tersebut.



Gambar 3. 9 *Bar Chart* Pekerjaan

Gambar 3.9 merupakan grafik batang yang menampilkan distribusi pekerjaan kepala keluarga penerima bantuan sosial PKH. Berdasarakan visualisasi tersebut, terlihat bahwa kepala keluarga yang tercatat tidak memiliki pekerjaan merupakan kategori terbanyak, yaitu 142 kepala keluarga. Selanjutnya, kategori buruh menempati urutan kedua dengan jumlah 81 kepala keluarga. Melalui grafik ini, penyebaran fitur pekerjaan dalam data penerima bantuan sosial PKH dapat dipahami dengan lebih baik dan jelas.

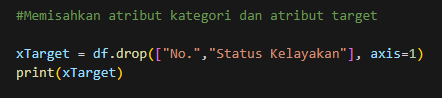
Secara keseluruhan, analisis deksriptif memberikan fondasi awal untuk memahami pola hubungan antar fitur dalam dataset sebelum masuk ke tahap pemodelan klasifikasi.

### *Data Preparation*

Data mentah yang diperoleh pada tahapan sebelumnya perlu dilakukan *Data Preparation*. Proses yang dilakukan dalam pengolahan data pada penelitian ini adalah sebagai berikut.

#### *Data Cleansing*

Pada tahap ini dilakukan persiapan data diantaranya menghapus fitur yang tidak digunakan, memisahkan fitur kategori dengan fitur target, dan mencari nilai yang hilang atau *missing value*.



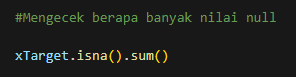
Gambar 3. 10 Penghapusan dan Pemisahan Fitur

Cuplikan kode pada Gambar 3.10 digunakan untuk menghapus fitur No. yang tidak berpengaruh terhadap proses analisis karena hanya merupakan nomor urut data. Selain itu, fitur Status Kelayakan yang merupakan target juga dihapus dari variabel fitur. Dengan demikian, fitur-fitur predictor yang tersisa disimpan dalam variabel xTarget.



Gambar 3. 11 Fitur Target

Cuplikan kode pada Gambar 3.11 digunakan untuk menyimpan fitur Status Kelayakan yang merupakan target ke dalam variabel yTarget.



Gambar 3. 12 *Missing Value*

Cuplikan kode pada Gambar 3.12 dilakukan pengecekan terhadap data yang memiliki nilai kosong atau *missing value* menggunakan fungsi “*isna()*”. Berikut hasil dari pengecekan tersebut:

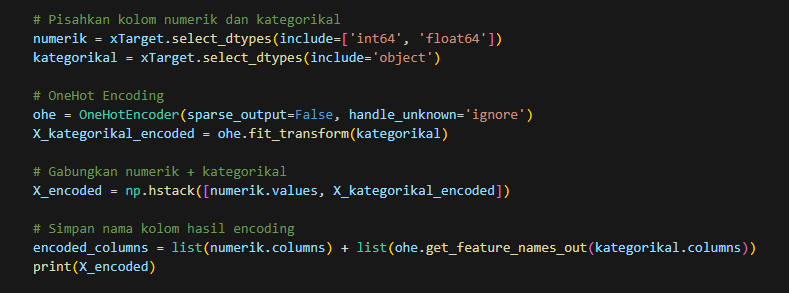
Tabel 3. 6 Hasil Pengecekan Missing Value

|  |  |
| --- | --- |
| **Nama Fitur** | **Jumlah Nilai Null** |
| Usia | 0 |
| Penghasilan | 0 |
| Pekerjaan | 0 |
| Status Pernikahan | 0 |
| Jumlah Tanggungan | 0 |
| Pendidikan | 0 |
| Status Rumah | 0 |
| L/P | 0 |
| Ibu Hamil/Menyusui | 0 |
| Balita | 0 |
| Anak Prasekolah | 0 |
| Anak Sekolah | 0 |
| Lansia | 0 |
| Penyandang Disabilitas | 0 |

Berdasarkan Tabel 3.6 terlihat bahwa setelah dilakukan pengecekan pada masing-masing fitur, tidak ditemukan nilai kosong. Oleh karena itu, proses dapat dilanjutkan ke tahap berikutnya.

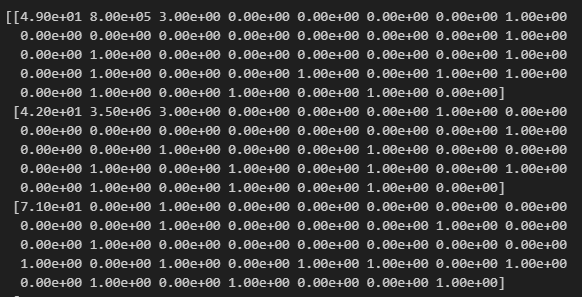
#### *Data Transformation*

Pada *Data Transformation* dilakukan perubahan data untuk pemodelan. Metode yang digunakan adalah metode *Label Binarizer* untuk fitur target dan metode *One-Hot Encoding* untuk fitur kategori.



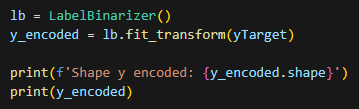
Gambar 3. 13 *One-Hot Encoding*

Pada dataset, terdapat fitur numerik yaitu Usia, Penghasilan, dan Jumlah Tanggungan. Oleh karena itu, seperti yang ditampilkan pada Gambar 3.13, fitur numerik tersebut dipisahkan terlebih dahulu sebelum dilakukan proses transformasi pada fitur kategorikal menggunakan metode *One-Hot Encoding*. Pemisahan ini dilakukan karena fitur numerik tidak memerlukan transformasi seperti pada fitur kategorikal. Setelah proses transformasi selesai, fitur kategorikal tersebut digabungkan kembali dengan fitur numerik sehingga membentuk dataset yang siap digunakan untuk proses pemodelan.



Gambar 3. 14 Hasil Transformasi Fitur Kategori

Pada Gambar 3.14 dapat dilihat bahwa hasil transformasi menggunakan *One-Hot Encoding* pada fitur kategorikal menyebabkan penambahan jumlah fitur. Awalnya dataset memiliki 14 fitur, namun setelah proses transformasi jumlahnya bertambah menjadi 39 fitur. Fitur-fitur tersebut meliputi: Usia, Penghasilan, Pekerjaan\_BURUH, Pekerjaan\_WIRASWASTA, Pekerjaan\_PEGAWAI, Pekerjaan\_KARYAWAN SWASTA, Pekerjaan\_PEDAGANG, Pekerjaan\_TIDAK ADA, Pekerjaan\_BERTANI, Pekerjaan\_ASISTEN RUMAH TANGGA, Pekerjaan\_PETERNAK, Status Pernikahan\_MENIKAH, Status Pernikahan\_CERAI MATI, Status Pernikahan\_CERAI HIDUP, Status Pernikahan\_BELUM MENIKAH, Jumlah Tanggungan, Pendidikan\_SD, Pendidikan\_SMP, Pendidikan\_SMA, Pendidikan\_S1, Status Rumah\_MILIK ORANG TUA, Status Rumah\_KONTRAK, Status Rumah\_MILIK SENDIRI, Status Rumah\_MENUMPANG, Status Rumah\_BEBAS SEWA, L/P\_L, L/P\_P, Ibu Hamil/Menyusui\_Ya, Ibu Hamil Menyusui\_Tidak, Balita\_Ya, Balita\_Tidak, Anak Prasekolah\_Ya, Anak Prasekolah\_Tidak, Anak Sekolah\_Ya, Anak Sekolah\_Tidak, Lansia\_Ya, Lansia\_Tidak, Penyandang Disabilitas\_Ya, dan Penyandang Disabilitas\_Tidak.



Gambar 3. 15 Transformasi Fitur Target Dengan *Label Binarizer*

Cuplikan kode pada Gambar 3.15 menggunakan *Label Binarizer* untuk mengubah format fitur target yang semula berupa string, yaitu “Layak “ dan “Tidak Layak”, menjadi format numerik.

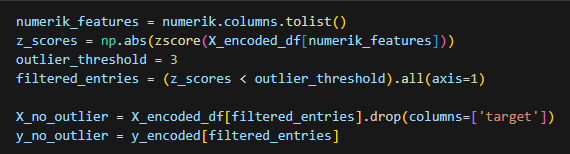
Tabel 3. 7 Perbedaan Fitur Target Sebelum dan Sesudah Transformasi

|  |  |
| --- | --- |
| **Fitur Status Kelayakan Sebelum Transformasi** | **Fitur Status Kelayakan Sesudah Transformasi** |
| Layak | 0 |
| Tidak Layak | 1 |

Berdasarkan tabel 3.7, fitur target menjadi format numerik berupa 0 dan 1. Proses ini dilakukan agar fitur target dapat diproses oleh model untuk melakukan proses pelatihan dan prediksi dengan optimal.

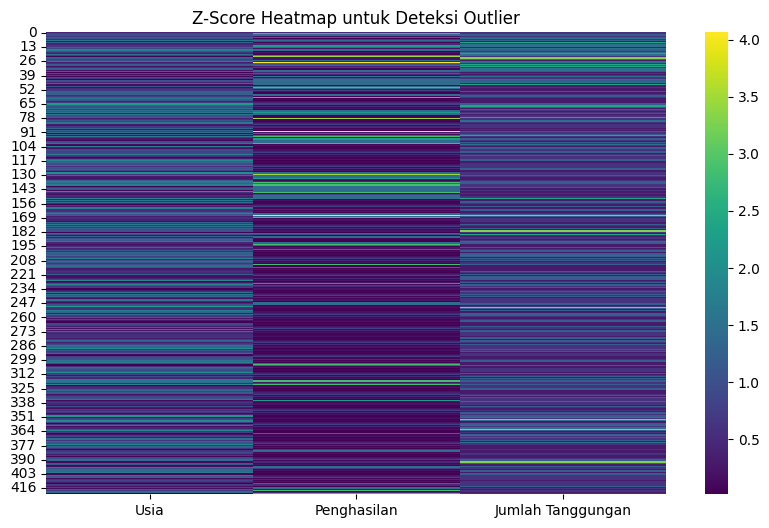
#### *Outlier*

*Outlier* adalah nilai yang jauh berbeda dari nilai lainnya dalam kumpulan data. Pada penelitian ini, Tahap pengecekan *outlier* dilakukan setelah tahap transformasi dan seleksi fitur.



Gambar 3. 16 *Outlier*

Cuplikan kode pada Gambar 3.16 merupakan proses mendeteksi *outlier* yang dilakukan pada fitur numerik menggunakan metode *Z-Score*. Pemilihan ambang batas *Z-Score* sebesar 3 berdasarkan aturan empiris yang menyatakan bahwa dalam distribusi normal, sekitar 99.7% data berada dalam 3 standar deviasi dari nilai rata-rata [29]. Oleh karena itu, nilai yang memiliki *Z-Score* lebih besar dari 3 berpotensi menjadi *outlier* yang dapat mendistorsi analisis model statistik.



Gambar 3. 17 Visualisasi *Outlier*

Berikut perhitungan mencari nilai *Z-Score* pada fitur usia, penghasilan, dan jumlah tanggungan:

Z-Score usia = = → 0,44

Z-Score penghasilan = =

*Z-Score* penghasilan = -0,39 → 0,39

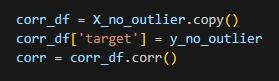
Z-Score jumlah tanggungan =

Z-Score jumlah tanggungan =

Berdasarkan hasil visualisasi pada Gambar 3.17 terlihat bahwa warna ungu atau gelap menunjukkan nilai *Z-Score* yang mendeteksi nol atau nilai yang tidak jauh dari rata-rata fitur tersebut. Warna kuning atau terang menunjukkan nilai *Z-Score* yang tinggi. Nilai *Z-Score* yang jauh dari nol menandakan sebagai potensi outlier [29]. Dengan demikian, data yang memiliki nilai *Z-Score* lebih dari 3 akan dihapus dari dataset. Hal ini bertujuan untuk menjaga kualitas data agar model yang dibangun tidak dipengaruhi oleh data yang menyimpang.

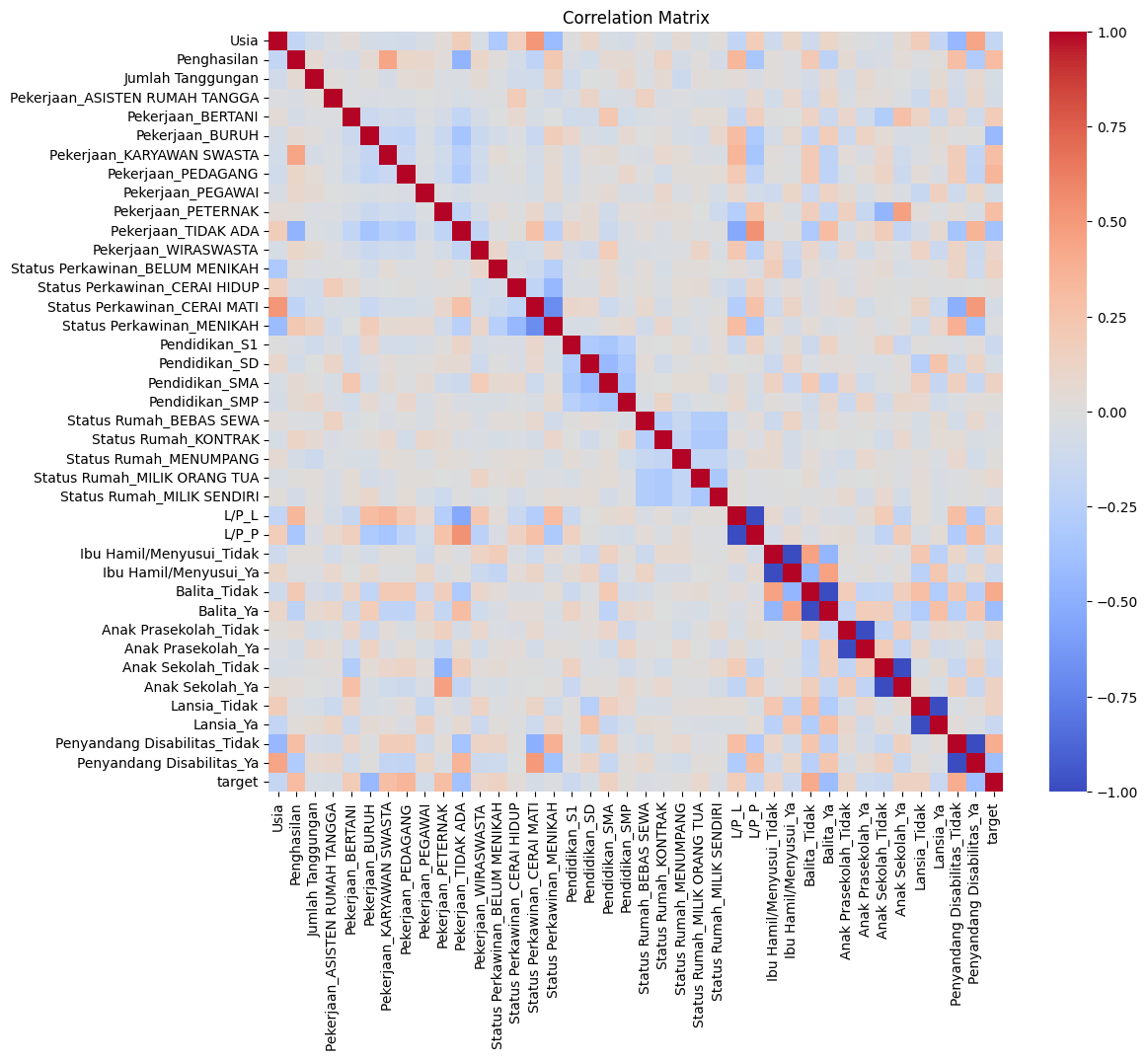
#### *Data Selection*

Tahap *Data Selection* dilakukan setelah *Data Transformation* dan mengatasi *outlier*, Hal ini disebabkan karena dalam menggunakan teknik Correlation Feature Selection (CFS) semua fitur dalam dataset harus berada dalam format numerik agar korelasi antar fitur dapat dihitung dengan akurat dan data sudah bersih dari *outlier* [28]. Oleh karena itu, seluruh fitur kategorikal pada dataset terlebih dahulu ditransformasikan menjadi bentuk numerik. Setelah seluruh data berada dalam format numerik, proses seleksi fitur dengan CFS dapat dijalankan untuk memilih fitur-fitur yang memiliki hubungan signifikan dengan target klasifikasi.



Gambar 3. 18 *Data Selection* Menggunakan CFS

Cuplikan kode pada Gambar 3.18 diatas dilakukan seleksi fitur menggunakan teknik *Corelation Feature Selection* (CFS). Dengan menerapkan CFS, fitur-fitur yang kurang relevan atau redundan dapat dieliminasi, sehingga model yang dihasilkan menjadi lebih efisien dan memiliki performa yang lebih baik.



Gambar 3. 19 *Correlation Matrix*

Berikut dibuat salah satu perhitungan korelasi, yakni antara fitur usia dengan fitur target:

Nilai korelasi usia dengan target =

Nilai korelasi usia dengan target =

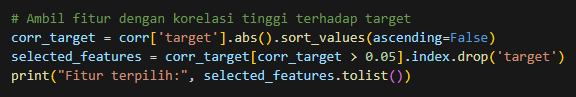
Keterangan:

Xi = Usia

Yi = Status Kelayakan

x̄ = Nilai Rata-rata

Gambar 3.19 menampilkan *Correlation Matrix* yang merupakan visualiasi dari hasil perhitungan korelasi antar fitur-fitur kategori terhadap fitur target. Nilai korelasi berkisar antara -1 sampai 1, dimana nilai yang mendekati 1 menyatakan korelasi positif yang kuat, artinya kedua fitur cenderung meningkat bersama. Nilai yang mendekati -1 menyatakan korelasi negative yang kuat, yang berarti ketika satu fitur meningkat, fitur lainnya menurun. Sementara nilai yang mendekati 0 menyatakan bahwa tidak ada hubungan atau korelasi lemah antara fitur-fitur tersebut. Selain itu, diagonal atau nilai 1 menunjukkan korelasi fitur terhadap dirinya sendiri [28].



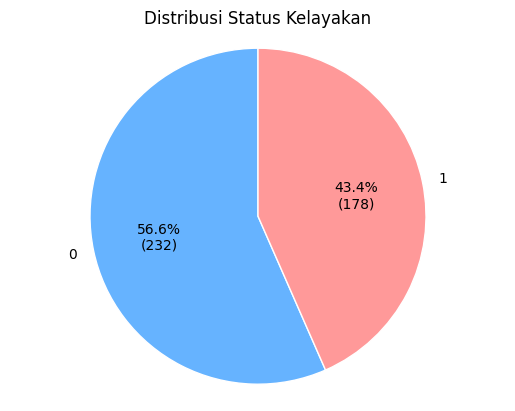
Gambar 3. 20 *Feature Selection*

Cuplikan kode pada Gambar 3.20 merupakan tahap seleksi fitur dengan menggunakan nilai *corr\_target* yang dipilih lebih dari 0.05, sesuai dengan hasil pada *Correlation Matrix*. Pemilihan ambang batas nilai korelasi ini didasarkan pada standar umum dalam analisis korelasi, dimana nilai tersebut menunjukkan hubungan yang cukup kuat antara fitur dengan target [28].

Pada tahap ini, terpilih sebanyak 30 fitur dari 39 yang tersedia. Adapun fitur-fitur tersebut adalah Pekerjaan\_BURUH, Balita\_Tidak, Balita\_Ya, Penyandang Disabilitas\_Tidak, Penyandang Disabilitas\_Ya, Pekerjaan\_TIDAK ADA, Pekerjaan\_PEDAGANG, Penghasilan, Pekerjaan\_KARYAWAN SWASTA, Pekerjaan\_PETERNAK, Pekerjaan\_BERTANI, L/P\_P, L/P\_L, Usia, Lansia\_Tidak, Lansia\_Ya, Pendidikan\_SMA, Status Perkawinan\_BELUM MENIKAH, Anak Sekolah\_Ya, Anak Sekolah\_Tidak, Pendidikan\_S1, Ibu Hamil/Menyusui\_Ya, Ibu Hamil/Menyusui\_Tidak, Anak Prasekolah\_Tidak, Anak Prasekolah\_Ya, Pekerjaan\_WIRASWASTA, Status Rumah\_MILIK ORANG TUA, Pekerjaan\_PEGAWAI, Pekerjaan\_ASISTEN RUMAH TANGGA, dan Status Perkawinan\_CERAI MATI.

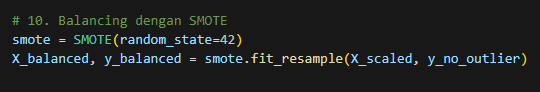
#### *Data Balancing*

Pada tahap *data balancing* menggunakan metode *Synthetic Minority Over-sampling Technique* (SMOTE). Proses ini dilakukan karena terdapat *imbalanced class* pada kelas Layak dan Tidak Layak. Sehingga diperlukan proses penyeimbangan data. Adapun presentasi *class* yang Layak dan Tidak Layak dapat dilihat pada Gambar 3.21 berikut:



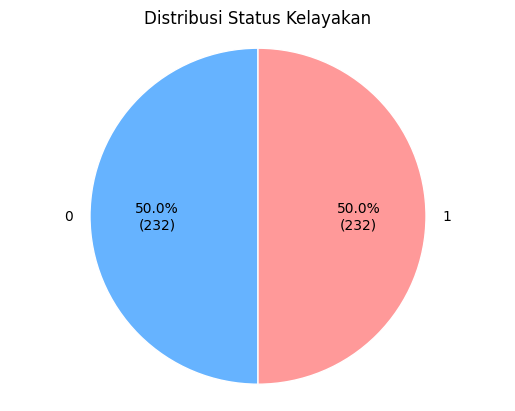
Gambar 3. 21 Diagram *Imbalanced Class*

Berdasarkan gambar diagram diatas, terjadi *imbalanced class* pada data tersebut yang memiliki perbandingan yaitu 56.6% yang layak dan 43.4% tidak layak. Secara jumlah, data kelas layak (dengan label 0) sebanyak 232 data, sedangkan kelas tidak layak (dengan label 1) sebanyak 178 data. Untuk mengatasi kondisi ini, dibutuhkan proses penyeimbangan data *class* tersebut. Teknik yang digunakan untuk menyeimbangkan data tersebut adalah SMOTE.



Gambar 3. 22 *Data Balancing* Menggunakan SMOTE

Cuplikan kode pada Gambar 3.22 menunjukkan proses *Data Balancing* menggunakan teknik SMOTE. Teknik ini bekerja dengan cara membuat data sintetis baru pada kelas minoritas hingga jumlahnya seimbang dengan kelas mayoritas. Dalam kode tersebut, objek SMOTE diinisialisasi dengan parameter *random\_state* untuk memastikan hasil yang konsisten.



Gambar 3. 23 Diagram Status Kelayakan Setelah Proses SMOTE

Berdasarkan Gambar 3.23 dapat dilihat bahwa bahwa jumlah kelas Layak dan Tidak Layak sudah seimbang setelah proses *balancing*. Keseimbangan ini penting untuk memastikan bahwa model yang dibangun nantinya tidak bias terhadap satu kelas. Dengan demikian, proses dapat dilanjutkan ke tahap berikunya.

#### *Data Splitting*

Pada tahap ini, data dibagi menjadi dua, yakni data latih dan data uji. Data latih digunakan untuk melatih model atau algoritma, sedangkan data uji digunakan untuk menguji seberapa baik kinerja model setelah dilatih. Adapun pembagian data dapat dilihat pada tabel

Tabel 3. 8 Pembagian Proporsi Data

|  |  |  |
| --- | --- | --- |
| **Pembagian** | **Jumlah Data Latih** | **Jumlah Data Uji** |
| 70:30 | 295 | 127 |
| 80:20 | 338 | 84 |

Berdasarkan Tabel 3.8, proses data splitting pada penelitian ini dilakukan dengan dua skenario, yaitu dengan rasio 70:30 dan 80:20. Pada rasio 70:30, sebanyak 70% data digunakan sebagai data latih (training) untuk membangun model, sedangkan 30% sisanya digunakan sebagai data uji (testing) untuk mengevaluasi kinerja model. Sementara itu, pada rasio 80:20, sebanyak 80% data digunakan untuk pelatihan model, dan 20% digunakan untuk pengujian. Penggunaan kedua rasio ini bertujuan untuk membandingkan performa model pada proporsi data latih dan data uji yang berbeda, sehingga dapat diketahui pengaruh pembagian data terhadap akurasi dan generalisasi model yang dihasilkan.

# HASIL DAN PEMBAHASAN

## Hasil

Hasil dan pembahasan dari penelitian ini merupakan penerapan dari metode *Multi-Layer Perceptron* (MLP) untuk melakukan klasifikasi penerima bantuan sosial Program Keluarga Harapan (PKH) di Kelurahan Cipamokolan. Pada pembahasan ini dilakukan beberapa tahap, yaitu *Modelling*, *Evaluation*, dan *Deployment*.

### *Modelling*

Pada tahap ini, menerapkan teknik pemodelan *data* mining, alat bantu *data* mining, algoritma atau metode yang digunakan dan menyesuaikan aturan model untuk mendapatkan nilai yang optimal. Multi*-Layer Perceptron* (MLP) akan diimplementasikan untuk mengklasifikasi menggunakan fungsi *Multi-Layer Perceptron Classifier* dari *library scikit-learn*.



Gambar 4. 1 *Modelling* dengan MLP

Pada Gambar 4.1, model yang dibangun menggunakan konfigurasi arsitektur jaringan yang terdiri dari dua *hidden layer* yang masing-masing memiliki 100 dan 50 neuron. Pemilihan jumlah neuron ini bertujuan untuk memberikan kapasitas pembelajaran yang cukup agar model mampu mengenali pola-pola kompleks dalam data. Model dilatih dengan parameter ‘*max\_iter=300’*, yang berarti proses pelatihan akan berlangsung hingga maksimal 300 iterasi atau hingga model mencapai titik stabil lebih awal. Selain itu, digunakan parameter ‘*random\_state=42’* agar hasil pelatihan dapat diulang dengan hasil yang konsisten. Proses pelatihan dilakukan dengan memanggil fungsi ‘*fit(X\_train, y\_train)*, di mana ‘*X\_train’* merupakan data fitur dan ‘*y\_train’* adalah label target. Setelah proses pelatihan selesai, model akan memiliki bobot-bobot yang telah disesuaikan berdasarkan data latih dan siap digunakan untuk melakukan prediksi terhadap data uji [9].

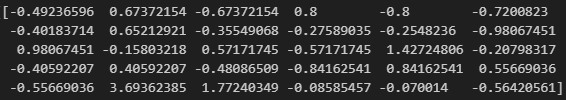
#### Perhitungan Manual *Multi-Layer Perceptron*

Perhitungan manual *Multi-Layer Perceptron* (MLP) dilakukan dengan menggunakan satu data sebagai sampel dari dataset yang telah melalui tahap *preprocessing*. Data tersebut dipilih untuk dijadikan uji coba dalam proses perhitungan manual untuk klasifikasi penerima bantuan sosial Program Keluarga Harapan (PKH) menggunakan metode MLP.

Tabel 4. 1 Data Uji yang Dipilih

|  |  |
| --- | --- |
| **Atribut** | **Parameter** |
| Usia | 49 |
| Penghasilan | Rp800.000,- |
| Pekerjaan | Pedagang |
| Status Perkawinan | Menikah |
| Jumlah Tanggungan | 3 |
| Pendidikan | SD |
| Status Rumah | Milik Orang Tua |
| Jenis Kelamin | Laki-laki |
| Ibu Hamil/Menyusui | Ya |
| Balita | Ya |
| Anak Prasekolah | Tidak |
| Anak Sekolah | Tidak |
| Lansia | Tidak |
| Penyandang Disabilitas | Tidak |

Setelah melalui tahap *preparation* data, dataset memiliki sebanyak 30 fitur. Oleh karena itu, jumlah neuron pada *input layer* adalah ditentukan sebanyak 30 neuron, sesuai dengan jumlah fitur sebagai *input* ke dalam jaringan MLP.



Gambar 4. 2 Dataset Setelah *Preparation* Data

Untuk memahami proses kerja model *Multi-Layer Perceptron* (MLP), dilakukan perhitungan manual berdasarkan bobot dan bias yang ditentukan secara otomatis oleh model selama proses pelatihan. Perhitungan ini dilakukan terhadap satu data sebagai sampel dengan asumsi struktur jaringan MLP terdiri dari:

1. *Input layer*: 30 neuron
2. *Hidden layer* 1: 100 neuron
3. *Hidden layer* 2: 50 neuron
4. *Output layer*: 1 neuron (klasifikasi biner “Layak” atau “Tidak Layak”)

Pada tahap pertama, nilai aktivasi masing-masing neuron pada *hidden layer* 1 dihitung dengan rumus:

Sebagai contoh, untuk satu neuron pada *hidden layer* 1:

Hasil ini kemudian dihitung menggunakan fungsi aktivasi *ReLu*:

Nilai aktivasi dari *hidden layer* 2 dihitung dengan menjadikan hasil aktivasi dari 100 neuron di *hidden layer* 1 sebagai *input*. Misalnya, untuk salah satu neuron pada *hidden layer* 2:

Kemudian diaplikasikan fungsi aktivasi *ReLu*:

*Output layer* menerima *input* dari 50 neuron pada *hidden layer* 2. Misalnya:

Nilai ini kemudian diolah dengan fungsi aktivasi *sigmoid*:

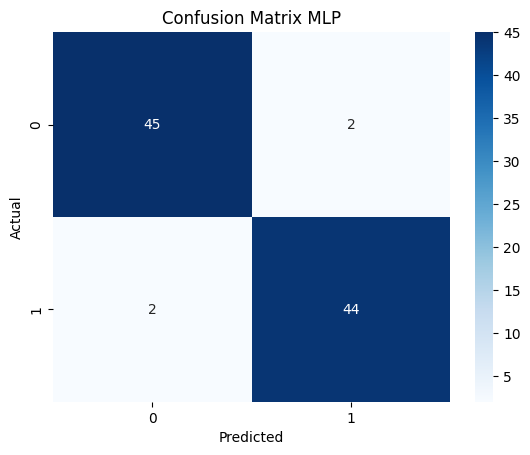
Jika nilai probabilitas hasil *sigmoid* lebih besar dari 0.5 maka data diklasifikasikan sebagai “Layak” menerima bantuan sosial PKH. Batas 0.5 dipilih karena merupakan titik tengah dari rentang probabilitas, dan secara umum digunakan sebagai ambang batas klasifikasi biner [9].

### *Metrik*

Evaluasi dilakukan pada dua bagian, yaitu data latih (*train set*) dan data uji (*test set*). Pada data latih, evaluasi bertujuan untuk melihat seberapa baik model mempelajari pola dari data yang tersedia. Sedangkan pada data uji, evaluasi digunakan untuk mengukur kemampuan generaliasasi model terhadap data baru yang belum pernah dilihat sebelumnya. Perbandingan hasil evaluasi pada kedua data tersebut dapat memberikan gambaran mengenai potensi *overfitting* atau *underfitting* yang terjadi pada model [5].

#### *Test Set*

Evaluasi kinerja model dilakukan pada data uji untuk mengukur kemampuan model dalam melakukan generaliasi terhadap data yang belum pernah dilihat sebelumnya [5]. Kinerja diperoleh dengan memberikan nilai pada *Confusion Matrix* untuk menghitung nilai *Accuracy*, *Precision*, *Recall*, dan *F1-Score* dari hasil pengujian.



Gambar 4. 3 *Confusion Matrix Test Set* *Heatmap*

Berdasarkan Gambar 4.3, dijelaskan bahwa terdapat 45 *True Positif* (TP), 2 *False Negatif* (FN), 2 *False Positif* (FP), dan 44 *True Negatif* (TN).

Tabel 4. 2 *Confusion Matrix Test Size*

|  |  |  |
| --- | --- | --- |
|  | **Prediksi** | |
| **Aktual** | *True* | *False* |
| *True* | TP=45 | FN=2 |
| *False* | FP=2 | TN=44 |

Kinerja yang diukur dengan menggunakan *Confusion Matrix* memiliki empat keluaran yang merepresentasikan hasil dari proses klasifikasi. Berikut penjelasannya:

1. *True Positive* (TP), jumlah data yang memang layak dan berhasil diprediksi oleh model, yaitu sebanyak 45.
2. *False Negatif* (FN), jumlah data yang seharusnya layak tetapi diprediksi tidak layak oleh model, yaitu sebanyak 2.
3. *False Positive* (FP), jumlah data yang seharusnya tidak layak tetapi diprediksi layak oleh model, yaitu sebanyak 2.
4. *True Negatif* (TN), jumlah data yang benar-benar tidak layak dan diprediksi tidak layak oleh model, yaitu sebanyak 44.

Dengan menggunakan metode evaluasi ini, tingkat akurasi keseluruhan kinerja *Multi-Layer Perceptron* (MLP) dapat dihitung sebagai berikut:

1. Untuk menghitung nilai *Accuracy* menggunakan rumus persamaan (6), hasil perhitungan sebagai berikut:

Akurasi mengukur seberapa banyak prediksi yang benar secara keseluruhan. Dalam hal ini, model memberikan prediksi yang benar untuk 95,70% dari semua *instance* dalam data.

1. Untuk menghitung nilai *Precision* menggunakan rumus persamaan (7), hasil perhitungan sebagai berikut:

*Precision* mengukur sejauh mana prediksi positif yang dibuat adalah benar. Dalam konteks ini, dari semua prediksi positif yang dibuat, 95,74% di antaranya adalah benar-benar positif, sementara sisanya mungkin adalah kesalahan.

1. Untuk menghitung nilai *Recall* menggunakan rumus persamaan (8), hasil perhitungan sebagai berikut:

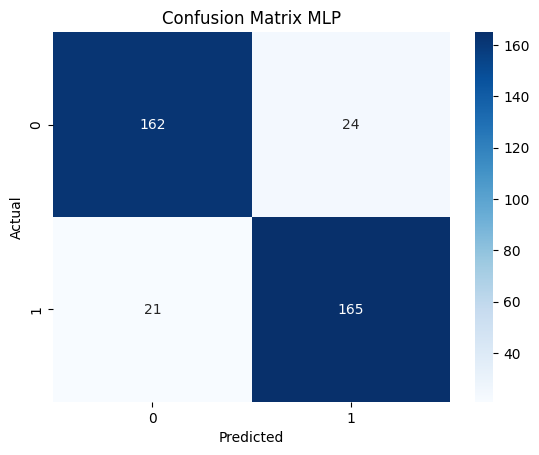
*Recall* mengukur sejauh mana model mampu mendeteksi semua *instance* positif yang sebenarnya. Dalam kasus ini, model berhasil mendeteksi 97,83% dari semua *instance* positif yang ada dalam data.

1. Untuk menghitung nilai *F1-Score* menggunakan rumus persamaan (9), hasil perhitungan sebagai berikut:

*F1-Score* adalah *harmonic mean* dari *precision* dan *recall*. Nilai 96,77% menunjukkan keseimbangan antara *precision* dan *recall* dalam model ini.

#### *Train Set*

Evaluasi performa model *Multi-Layer Perceptron* (MLP) pada data latih dilakukan untuk mengukur sejauh mana model mampu mempelajari pola dari data yang digunakan saat pelatihan [5].



Gambar 4. 4 *Confusion Matrix Train Set Heatmap*

Berdasarkan Gambar 4.4, dijelaskan bahwa terdapat 162 *True Positif* (TP), 24 *False Negatif* (FN), 21 *False Positif* (FP), dan 165 *True Negatif* (TN).

Tabel 4. 3 *Confusion Matrix Train Size*

|  |  |  |
| --- | --- | --- |
|  | **Prediksi** | |
| **Aktual** | *True* | *False* |
| *True* | TP=162 | FN=24 |
| *False* | FP=21 | TN=165 |

Berdasarkan Tabel 4.3, ditampilkan *Confusion Matrix* hasil evaluasi model MLP pada data latih dengan proporsi 80% dari total dataset. Dari hasil tersebut, diproleh:

1. *True Positive* (TP), 162 data Layak berhasil diklasifikasi dengan benar.
2. *False Negatif* (FN), 24 data Tidak Layak diklasifikasikan salah sebagai Tidak Layak.
3. *False Positive* (FP), 21 data Tidak Layak diklasifikasikan salah sebagai Layak.
4. *True Negatif* (TN), 165 data dengan label Tidak Layak berhasil diklasifikasikan dengan benar.

Dari *Confusion Matrix* tersebut, dapat dihitung metrik evaluasi sebagai berikut:

1. Nilai *Accuracy*:

Model memberikan prediksi yang benar untuk 87,90% dari semua *instance* dalam data.

1. Nilai *Precision*:

Dalam konteks ini, dari semua prediksi positif yang dibuat, 88,52% di antaranya adalah benar-benar positif, sementara sisanya mungkin adalah kesalahan.

1. Nilai *Recall*:

Dalam kasus ini, model berhasil mendeteksi 87,10% dari semua *instance* positif yang ada dalam data.

1. Nilai *F1-Score*:

Nilai 87,80% menunjukkan keseimbangan antara *precision* dan *recall* dalam model ini.

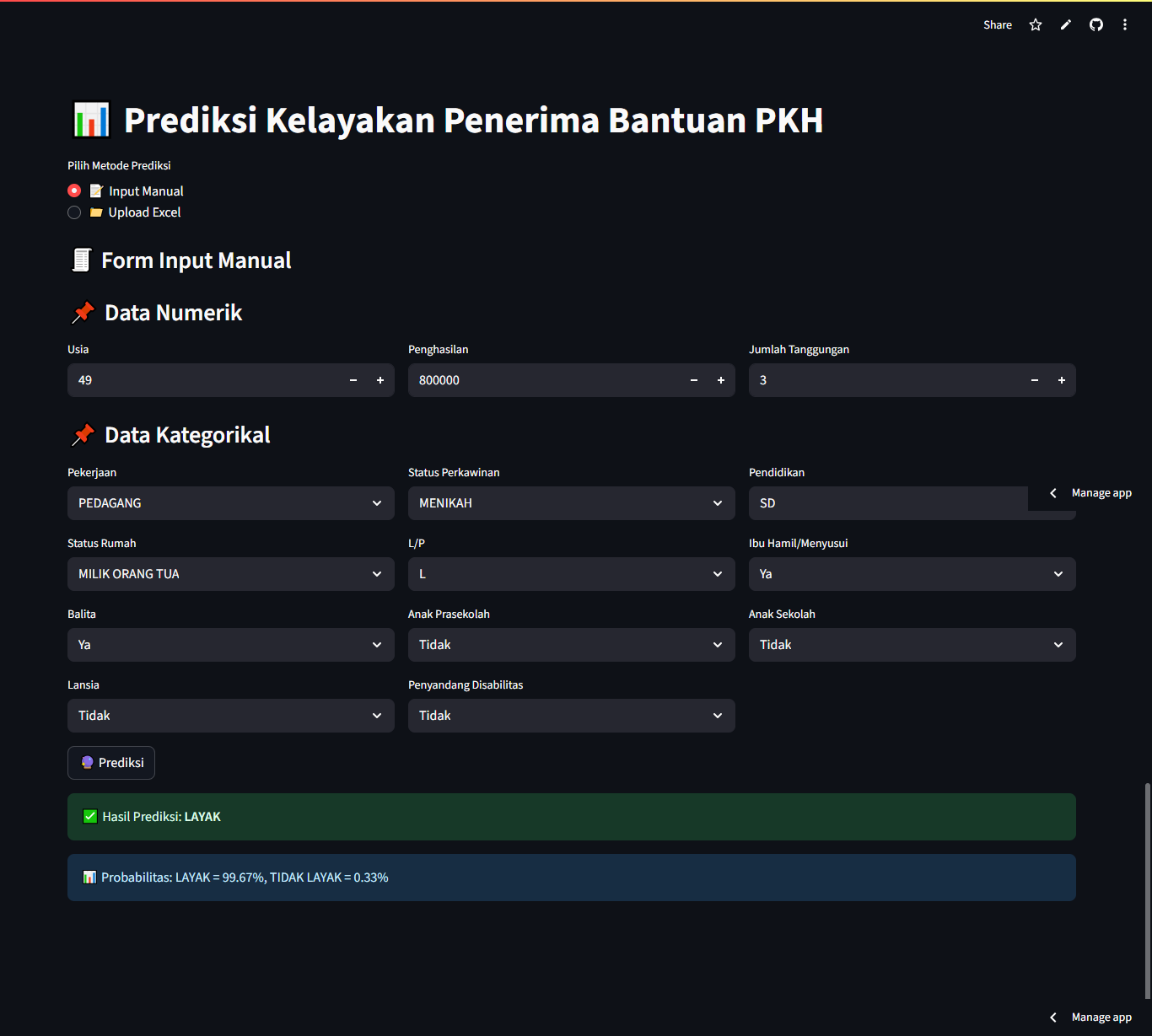
Evaluasi pada data latih menunjukkan bahwa model MLP memiliki kinerja yang baik dalam mempelajari pola dari data, dengan nilai *recall* dan *precision* yang seimbang. Hasil ini akan menjadi dasar untuk dibandingkan dengan evaluasi pada data uji guna menilai kemampuan generaliasi model [5].

Dengan membandingkan hasil evaluasi pada data latih dan data uji, terlihat bahwa performa model relatif stabil pada kedua data, tanpa perbedaan yang signifikan. Hal ini mengindikasikan bahwa model tidak mengalami *overfitting* dan cukup andal untuk digunakan dalam memprediksi penerima bantuan sosial PKH di Kelurahan Cipamokolan.

### *Deployment*

Tahap ini memastikan bahwa model tidak hanya berfungsi dalam hal teknis namun juga dapat memberikan kepraktisan kepada pengguna. Peneliti mengimplementasikan penerimaan bantuan PKH melalui algoritma klasifikasi yang diterapkan dalam sebuah sistem informasi berbasis *web* menggunakan *framework streamlit*.

Aplikasi ini menyediakan dua mode *input* data, yaitu *input* manual untuk data tunggal (satu data), dan *input* data dalam jumlah banyak (multi data) melalui file berformat *Excel*. Untuk prediksi data tunggal, data yang akan diuji coba sama dengan data yang digunakan pada perhitungan manual, yaitu data yang disajikan pada Tabel 4.1.



Gambar 4. 5 Tampilan *Input* Data Tunggal

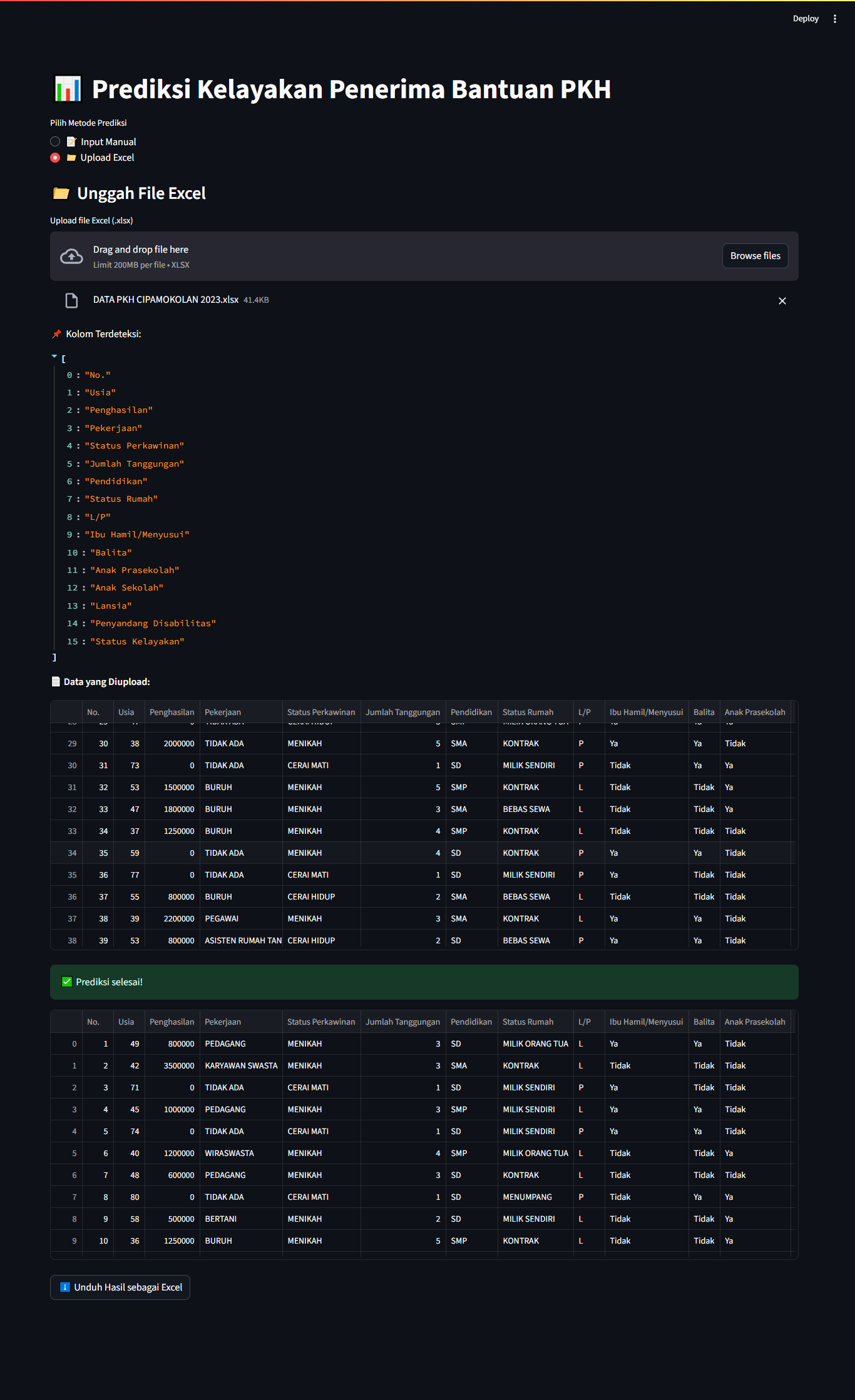
Berdasarkan Gambar 4.5, data yang dimasukkan berupa Usia “49”, Penghasilan “800000”, Jumlah Tanggungan “3”, Pekerjaan “Pedagang”, Status Perkawinan “Menikah”, Pendidikan “SD”, Status Rumah “Milik Orang Tua”, Jenis Kelamin “L/P”, Ibu Hamil/Menyusui “Ya”, Balita “Ya”, Anak Prasekolah “Tidak”, Anak Sekolah “Tidak”, Lansia “Tidak”, Penyandang Disabilitas “Tidak”.



Gambar 4. 6 Hasil Prediksi *Input* Data Tunggal

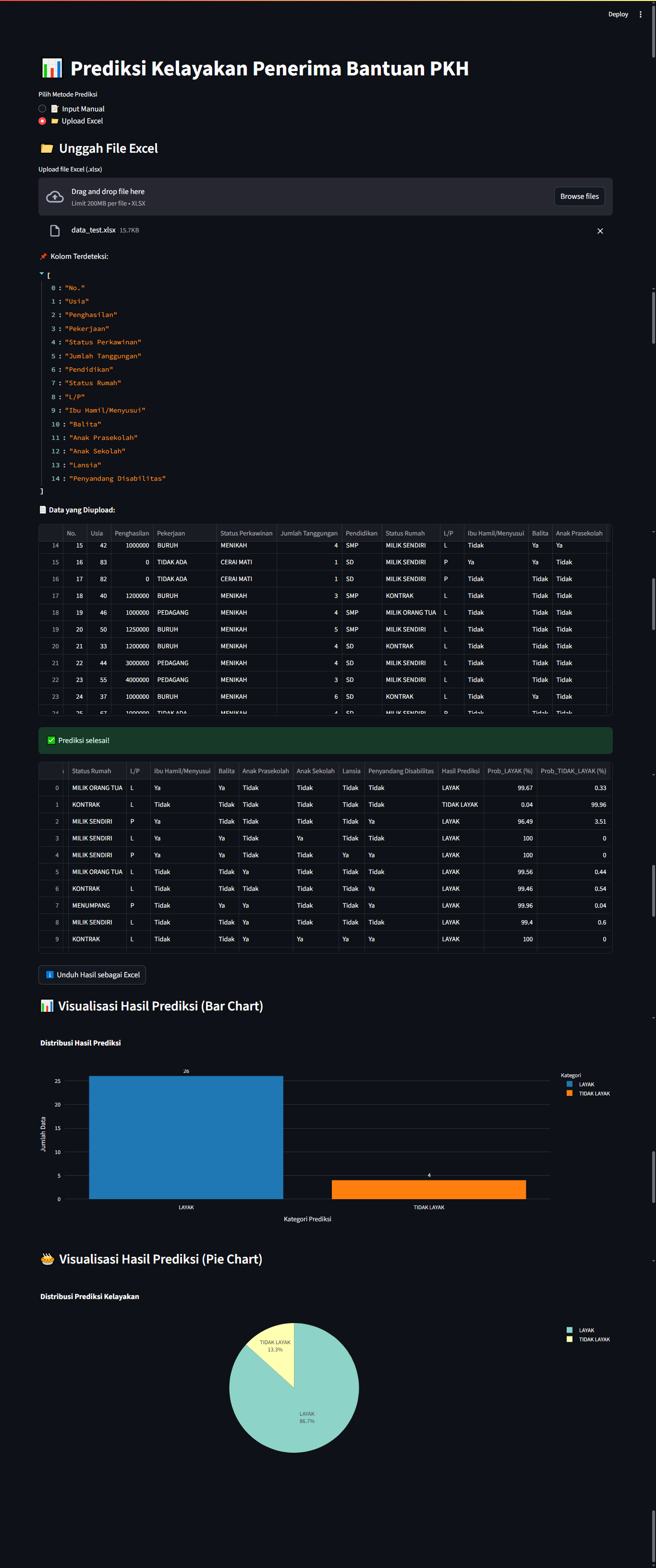
Berdasarkan Gambar 4.6, hasil pengujian aplikasi *web* klasifikasi penerima bantuan sosial Program Keluarga Harapan menggunakan *streamlit* menunjukkan bahwa data yang diujii diprediksi sebagai “Layak” menerima bantuan, dengan probabilitas sebesar 99,67% untuk kelas Layak dan 0,33% untuk kelas Tidak Layak.

Untuk melakukan prediksi multi data, digunakan dataset yang sama dengan yang dipakai dalam penelitian ini, yaitu dataset penerima Bantuan Sosial Program Keluarga Harapan (PKH) Kelurahan Cipamokolan Tahun 2023, sebagaimana ditampilkan pada Gambar 3.7.



Gambar 4. 7 Tampilan *Input* Multi Data

Pada proses input multi data, jumlah fitur dan nama setiap fitur harus sesuai dengan ketentuan yang telah ditetapkan. Apabila struktur fitur tidak sesuai, sistem akan menampilkan pesan peringatan untuk memperbaiki struktur data tersebut, dan proses prediksi tidak akan dapat dilanjutkan. Namun, jika struktur fitur dalam dataset sudah sesuai, maka aplikasi akan menampilkan *output* hasil prediksi sebagai berikut:



Gambar 4. 8 Hasil Prediksi *Input* Multi Data

Berdasarkan pada Gambar 4.8, *output* dari prediksi multi data menghasilkan dataset baru yang telah dilengkapi dengan tiga fitur tambahan, yaitu hasil prediksi, probabilitas layak, dan probabilitas tidak layak. Dataset hasil prediksi ini dapat diunduh oleh pengguna dalam bentuk *file Excel* (*soft file*) untuk keperluan dokumentasi atau analisis lebih lanjut.

## Pembahasan dan Rekomendasi

Pada subbab ini dibahas perbandingan kinerja model klasifikasi yang dibangun dengan model klasifikasi lain. Berdasarkan pembahasan tersebut, diberikan beberapa rekomendasi yang dapat menjadi masukan untuk pengembangan model di masa mendatang.

### Pembahasan

Pada proporsi data 70:30, jumlah data untuk latihan adalah sebanyak 295 dan untuk uji adalah 127 dari total data 422. Selanjutnya, pada proporsi data 80:20. Dimana, terdapat jumlah data untuk latihan adalah sebanyak 338 dan untuk uji adalah sebanyak 84. Dari sisi parameter, Arsitektur jaringanyang digunakan adalah 100 dan 50 dengan nilai epoch 300. Hasil dari klasifikasi dapat dilihat pada Tabel 4.2.

Tabel 4. 4 Hasil Klasifikasi pada Proporsi Data

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Proporsi Data** | **Accuracy** | **Recall** | **Precision** | **F1-Score** |
| 70:30 | 93,57% | 91,43% | 95,52% | 93,43% |
| 80:20 | 95,70% | 97,83% | 95,74% | 96,77% |

Berdasarkan hasil yang ditunjukkan pada Tabel 4.4, terlihat adanya perbedaan nilai antara proporsi data 70:30 dan 80:20. Nilai akurasi pada proporsi data 80:20 adalah sebesar 95,70% sedangkan pada proporsi data 70:30 hanya 93,57%. Hal yang sama juga terhadi pada perbandingan nilai *recall*, *precision*, dan *F1-score*, dimana proporsi data 80:20 lebih unggul dibanding dengan 70:30.

#### Perbandingan Kinerja MLP dengan Algoritma Lain

Berdasarkan keunggulan kinerja model pada proporsi data 80:20, maka proporsi ini digunakan sebagai dasar dalam membandingkan algoritma *Multi-Layer Perceptron* (MLP) dengan klasifikasi lainnya, yaitu *Naïve Bayes* dan *Decision Tree*. Perbandingan ini bertujuan untuk mengetahui sejauh mana keefektifan algoritma MLP dalam melakukan klasifikasi penerima bantuan sosial Program Keluarga Harapan (PKH) dibandingkan dengan algoritma lainnya.

Setiap algoritma diuji menggunakan dataset yang sama, dengan proporsi data 80% untuk pelatihan dan 20% untuk pengujian. Evaluasi dilakukan berdasarkan metrik akurasi, *recall*, *precision*, dan *F1-score*. Hasil dari perbandingan tersebut akan ditampilkan pada Tabel 4.5.

Tabel 4. 5 Hasil Klasifikasi Perbandingan Algoritma

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Algoritma Klasifikasi** | ***Accuracy*** | ***Recall*** | ***Precision*** | ***F1-Score*** |
| *Multi-Layer Perceptron* | 95,70% | 97,83% | 95,74% | 96,77% |
| *Naïve Bayes Classifier* | 65,59% | 100% | 58,97% | 74,19% |
| *Decision Tree* | 93,54% | 95,65% | 91,67% | 93,62% |

Berdasarkan Tabel 4.5, dapat dilihat perbandingan kinerja tiga algoritma klasifikasi yang digunakan dalam penelitian ini, yaitu *Multi-Layer Perceptron* (MLP), *Naïve Bayes Classifier*, dan *Decision Tree*, dengan proporsi data latih dan data uji sebesar 80:20.

Hasil evaluasi menunjukkan bahwa algoritma MLP memberikan performa terbaik dibandingkan dua algoritma lainnya. MLP memperoleh nilai akurasi sebesar 95,70%, *recall* 97,83%, *precision* 95,74%, dan *F1-scroe* 96,77%. Nilai-nilai tersebut menunjukkan bahwa MLP mampu mengklasifikasikan data dengan sangat baik dan seimbang antara kemampuan mendeteksi kelas yang benar (*recall*) dan ketepatan prediksi (*precision*).

Sementara itu, *Naïve Bayes Classifier* memiliki nilai *recall* tertinggi yaitu 100%, yang berarti semua data pada kelas positif berhasil terdeteksi. Namun nilai akurasi dan *precision*-nya jauh lebih rendah, masing-masing sebesar 65,59% dan 58,97%, sehingga menghasilkan *F1-score* yang juga rendah, yaitu 74,19%. Hal ini menunjukkan bahwa meskipun *Naïve Bayes* mampu mengenali semua penerima yang seharusnya layak, namun banyak terjadi kesalahan klasifikasi untuk data yang sebenarnya tidak layak.

Namun, rendahnya akurasi yang dihasilkan oleh algoritma *Naïve Bayes* kemungkinan disebabakan oleh karakteristik algoritma tersebut yang lebih optimal digunakan pada data diskrit atau kategorikal. Dataset yang digunakan dalam penelitian ini telah melalui proses transforamsi numerik untuk keperluan algoritma MLP, yang memang membutuhkan input numerik terstandarisasi [29].

Sementara itu, algoritma *Naïve Bayes* bekerja lebih efektif apabila fitur berupa data kategorikal, karena perhitungan probabilitas kondisionalnya berdasarkan frekuensi kemunculan kelas. Jika ingin tetap menggunakan data diskrit untuk *Naïve Bayes*, maka data numerik seperti penghasilan, usia, atau jumlah tanggungan harus diubah ke dalam bentuk kategori atau rentang nilai [29]. Namun, penerapan sistem rentang tersebut justru dapat mengurangi performa algoritma MLP, yang membutuhkan nilai numerik agar pembelajaran melalui proses propagasi bobot dan aktivasi neuron berjalan optimal [9]. Dengan demikian, transformasi numerik yang digunakan pada dataset lebih menguntungkan algoritma MLP, namun kurang sesuai untuk karakteristik *Naïve Bayes*. Hal ini menjelaskan *Naïve Bayes* tampak memiliki akurasi yang lebih rendah dibandingkan algoritma lainnya dalam penelitian ini.

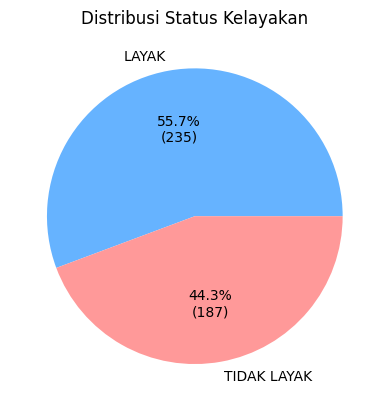
Adapun algoritma *Decision Tree* menunjukkan kinerja yang cukup baik, dengan nilai akurasi 93,54%, *recall* 95,65%, *precision* 91,67%, dan *F1-score* 93,62%. Algoritma *Decision Tree* mampu bekerja dengan baik pada data numerik, karena dapat memecah data berdasarkan nilai ambang tanpa memerlukan transformasi diskrit [34]. Hal ini membuat *Decision Tree* tetap menunjukkan performa yang tinggi meskipun tetap sedikit di bawah MLP dalam hal akurasi dan keseimbangan klasifikasi.

Dari hasil perbandingan ini, dapat disimpulkan bahwa algoritma *Multi-Layer Perceptron* (MLP) merupakan algoritma terbaik yang digunakan dalam penelitian ini untuk melakukan klasifikasi kelayakan penerima bantuan sosial Program Keluarga Harapan (PKH) di Kelurahan Cipamokolan. Rangkuman dari hasil klasifikasinya dapat dilihat pada Tabel 4.6.

Tabel 4. 6 Klasifikasi Data Kelayakan Penerima Bantuan Sosial PKH

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **No.** | **Usia** | **Penghasilan** | **Pekerjaan** | **…** | **Status Kelayakan** | **Hasil Prediksi** |
| 1 | 49 | 800000 | Pedagang | … | Layak | Layak |
| … | … | … | … | … | … | … |
| 106 | 53 | 1200000 | Wiraswasta | … | Layak | Tidak Layak |
| … | … | … | … | … | … | … |
| 422 | 56 | 3500000 | Karyawan Swasta | … | Tidak Layak | Tidak Layak |

Tabel 4.6 menunjukkan adanya beberapa ketidaksamaan antara klasifikasi status kelayakan dengan prediksi yang dilakukan oleh MLP. Dimana, nilai akurasi-nya adalah sebesar 95,70% seperti yang terlihat pada Tabel 4.5. Adapun hasil dari klasifikasi yang dilakukan adalah sebanyak 235 kepala keluarga dikategorikan layak dan 187 kepala keluarga dinyatakan tidak layak. Visualisasi dari hasil klasifikasi data terlihat pada Gambar 4.10.



Gambar 4. 9 Visualisasi Hasil Klasifikasi

Kontribusi penelitian ini bahwa pemodelan metode *Multi-Layer Perceptron* (MLP) menghasilkan model performa terbaik, sehingga kedepannya dapat diimplementasikan pada sebuah sistem pendukung keputusan untuk membantu proses penentuan bantuan sosial Program Keluarga Harapan (PKH) agar lebih cepat dan mudah.

### Rekomendasi

Berdasarkan hasil evaluasi performa model *Multi-Layer Perceptron* (MLP) yang telah dilakukan, model menunjukkan kinerja yang sangat baik dengan nilai akurasi sebesar 95,70%, *precision* sebesar95,74%, *recall* sebesar 97,83%, dan *F1-Score* sebesar 96,77%. Meskipun nilai evaluasi tergolong tinggi, terdapat beberapa hal yang dijadikan rekomendasi untuk meningkatkan performa model maupun pengembangan sistem ke depannya. Berikut adalah beberapa rekomendasi yang disarankan:

1. Penambahan variabel fitur

Model saat ini dilatih dengan fitur-fitur terbatas. Direkomendasikan untuk melakukan eksplorasi dan penambahan variabel baru yang relevan seperti kepemilikan aset (rumah dan kendaraan) dan riwayat penerimaan bantuan tahun-tahun sebelumnya. Penambahan fitur ini diharapkan dapat meningkatkan ketepatan klasifikasi penerima bantuan.

1. Eksperimen arsitektur jaringan

Kinerja model MLP sangat dipengaruhi oleh jumlah neuron dan *hidden layer.* Direkomendasikan untuk melakukan eksperimen jumlah neuron dan *layer*. Dengan melakukan eksplorasi dan optimasi arsitektur jaringan, diharapkan model MLP tidak hanya memiliki akurasi yang tinggi pada data pelatihan, tetapi juga mampu mempertahankan performa yang baik pada data baru atau data aktual di lapangan.

1. Bandingkan dengan algoritma klasifikasi lain

Untuk mendapatkan pemahaman yang lebih komprehensif tentang kinerja *Multi-Layer Perceptron*, direkomendasikan untuk melakukan evaluasi dan perbandingan dengan model klasifikasi lain seperti *Random Forest* dan *Support Vector Machine* untuk melihat model mana yang paling sesuai dengan karakteristik data.

# KESIMPULAN DAN SARAN

## Kesimpulan

Berdasarkan penerapan *data mining* menggunakan metode *Multi-Layer Perceptron* (MLP) untuk klasifikasi penerima bantuan sosial Program Keluarga Harapan (PKH) Kelurahan Cipamokolan, maka diperoleh beberapa kesimpulan sebagai berikut:

1. Metode *Multi-Layer Perceptron* berhasil diterapkan untuk melakukan klasifikasi terhadap kelayakan penerima bantuan sosial PKH. Dataset yang digunakan telah melalui tahapan *preprocessing*, sehingga siap digunakan dalam pelatihan model.Hasil pengujian menunjukkan bahwa model MLP memiliki akurasi sebesar 95,70%, yang mengindikasikan kemampuan model dalam mengklasifikasikan penerima dengan benar sebanyak 95,70% dari total data yang diuji.
2. Nilai *precision* sebesar 95,74% menunjukkan bahwa dari semua prediksi yang dilakukan oleh model sebagai penerima bantuan sosial PKH, sekitar 95,74% di antaranya adalah prediksi benar.
3. Nilai *recall* sebesar 97,83% menggambarkan bahwa model mampu mendeteksi sekitar 97,83% dari total penerima bantuan sosial PKH yang sebenarnya.
4. Nilai *F1-score* sebesar 96,77% adalah harmonisasi antara *precision* dan *recall*. *F1-score* mengukur keseimbangan antara kemampuan model dalam mengklasifikasikan dengan tepat dan kemampuan dalam mendeteksi penerima yang sebenarnya.
5. Perbandingan dengan algoritma lain menunjukkan bahwa MLP memiliki performa lebih unggul dibandingkan algoritma *Naïve Bayes* dan *Decision Tree* dalam aspek akurasi, *precision*, *recall*, dan *F1-score*. Hal ini menunjukkan bahwa MLP lebih efektif untuk digunakan dalam kasus klasifikasi berbasis data numerik kompleks seperti pada penelitian ini.
6. Implementasi *website* sederhana menggunakan *framework streamlit* untuk membantu pengambilan keputusan dan mengurangi resiko terjadinya kekeliruan dalam penyaluran bantuan sosial PKH kepada masyarakat Kelurahan Cipamokolan. Pengguna dapat melakukan prediksi secara langsung baik untuk data individu maupun multi-data dalam format *ExcelI*, serta mendapatkan hasil klasifikasi beserta probabilitasnya prediksi yang dihasilkan oleh model.

## Saran

Berdasarkan hasil penelitian dan implementasi yang telah dilakukan, terdapat beberapa saran yang dapat dipertimbangkan untuk pengembangan lebih lanjut:

1. Penelitian ini hanya menggunakan dataset yang diperoleh dari wilayah Kelurahan Cipamokolan. Untuk meningkatkan generalisasi model, disarankan untuk mengembangkan dataset dari wilayah yang lebih luas atau mencakup beberapa kelurahan/kecamatan guna memperkuat akurasi dan daya prediksi model terhadap data yang lebih beragam.
2. Kinerja model MLP sangat dipengaruhi oleh arsitektur jaringan serta parameter pelatihan. Disarankan untuk melakukan eksperimen lebih lanjut menggunakan teknik seperti *grid search* atau *random search* untuk mendapatkan konfigurasi optimal.
3. Selain *Naïve Bayes* dan *Decision Tree*, algoritma lain seperti *Random Forest*, *Support Vector Machine* (SVM), atau *Gradient Boosting* dapat diuji untuk mengetahui apakah terdapat metode lain yang mampu memberikan performa lebih baik dibandingkan MLP.
4. Aplikasi *web* yang dikembangkan sudah berfungsi dengan baik untuk kebutuhan prediksi dasar. Kedepannya, aplikasi ini dapat diperluas dengan penambahan fitur seperti histori prediksi, sistem login multi-level (admin dan petugas), serta validasi hasil prediksi oleh pihak terkait untuk mendukung proses pengambilan keputusan yang lebih terstruktur.

# DAFTAR PUSTAKA

[1] A. Sutanti, M. Komaruddin, P. Damayanti, and P. U. Studi Sistem Informasi Metro, “RANCANG BANGUN APLIKASI PERPUSTAKAAN KELILING MENGGUNAKAN PENDEKATAN TERSTRUKTUR,” *Jurnal Ilmiah Komputer dan Informatika (KOMPUTA)*, vol. 9, no. 1, 2020.

[2] A. A. Musaddad and A. Kriswibowo, “Optimalizing Big Data in Reducing Miss-Targeting Family Hope Program (PKH) in Sidoarjo Disctrict with Approach Machine Learning,” *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 15, no. 1, p. 99, Jan. 2021, doi: 10.22146/ijccs.62589.

[3] A. K. Fajri, D. Larasati, S. P. Alifkah, D. Augustin, and R. Herawati, “ANALISIS KEBIJAKAN PENANGGULANGAN KEMISKINAN MELALUI PROGRAM KELUARGA HARAPAN,” *Jurnal Manajemen dan Kebijakan Publik*, vol. 7, no. 1, pp. 158–170, Mar. 2022, doi: https://doi.org/10.14710/gp.7.1.2022.158-170.

[4] C. Lestari, T. Ratih, D. Saputri, S. Charles, and P. T. Siahaan, “Analisis Sentimen Pandangan Netizen Indonesia Terhadap Vaksin COVID-19 Menggunakan Multi-Layer Perceptron,” *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 9, no. 4, pp. 2795–2803, Dec. 2022, [Online]. Available: http://jurnal.mdp.ac.id

[5] P. H. Simanjuntak, “IMPLEMENTASI ALGORITMA RANDOM FOREST DALAM KLASIFIKASI KELAYAKAN MASYARAKAT UNTUK MENDAPATKAN BANTUAN SOSIAL,” Medan, Jan. 2025. Accessed: May 01, 2025. [Online]. Available: repository.uma.ac.id

[6] A. Khaleel Faieq and M. M. Mijwil, “Prediction of heart diseases utilising support vector machine and artificial neural network,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 26, no. 1, pp. 374–380, Apr. 2022, doi: 10.11591/ijeecs.v26.i1.pp374-380.

[7] K. Devendran, S. K. Thangarasu, P. Keerthika, R. Manjula Devi, and B. K. Ponnarasee, “Effective prediction on music therapy using hybrid SVM-ANN approach,” *ITM Web of Conferences*, vol. 37, p. 01014, Mar. 2021, doi: 10.1051/itmconf/20213701014.

[8] A. Dutta, T. Batabyal, M. Basu, and S. T. Acton, “An Efficient Convolutional Neural Network for Coronary Heart Disease Prediction,” Apr. 2020, doi: https://doi.org/10.48550/arXiv.1909.00489.

[9] S. H. Gulo and A. H. Lubis, “Penerapan Multi-Layer Perceptron untuk Mengklasifikasi Penduduk Kurang Mampu,” *Journal of Computer Science and Information Technology*, vol. 4, no. 2, pp. 51–59, Jul. 2024, [Online]. Available: https://www.researchgate.net/publication/382800582

[10] A. H. A. Abdel-aziem and T. H. M. Soliman, “A Multi-Layer Perceptron (MLP) Neural Networks for Stellar Classification: A Review of Methods and Results,” *International Journal of Advances in Applied Computational Intelligence*, vol. 3, no. 2, pp. 29–37, 2023, doi: 10.54216/IJAACI.030203.

[11] A. Secilmis, N. Aksu, F. A. Dael, I. Shayea, and A. A. El-Saleh, “Machine Learning Based Fire Detection: A Comprehensive Review and Evaluation of Classification Models,” *International Journal on Informatics Visualization*, vol. 7, pp. 1982–1988, Nov. 2023, doi: http://dx.doi.org/10.30630/joiv.7.3-2.2332.

[12] A. Nursikuwagus, T. Hartono, M. A. Nurwicaksono, M. M. Choir, and M. A. Saputri, “Hyperparameter tuning of multilayer convolutional network and augmentation method for classification motive of batik,” *Jurnal Informatika*, vol. 17, no. 1, pp. 28–38, Jan. 2023, doi: 10.26555/jifo.v17i1.a25823.

[13] I. W. Supriana, M. A. Raharja, and I. M. S. Bimantara, “Pengembangan Sistem Prediksi Bantuan Program Keluarga Harapan (PKH) Berbasis Machine Learning,” *SINTECH Journal*, vol. 6, Apr. 2023, doi: https://doi.org/10.31598.

[14] I. M. Dewi, L. Lisdiana, and I. M. Sakir, “Implementasi Kebijakan Program Keluarga Harapan di Kecamatan Lawang Kidul,” *Jurnal Ilmiah Ilmu Sosial*, vol. 10, no. 2, pp. 118–128, Dec. 2024, doi: 10.23887/jiis.v10i2.79706.

[15] Muhammad Arhami and Muhammad Nasir, *Data Mining - Algoritma dan Impelementasi*. Penerbit Andi, 2020.

[16] Rahmadini, E. E. LorencisLubis, A. Priansyah, Y. R. W. N., and T. Meutia, “PENERAPAN DATA MINING UNTUK MEMPREDIKSI HARGA BAHAN PANGAN DI INDONESIA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR,” *Jurnal Mahasiswa Akuntansi Samudra (JMAS)*, vol. 4, no. 4, pp. 223–235, Aug. 2023.

[17] K. Handoko and L. Sabda Lesmana, “Computer Based Information System Journal PENGELOMPOKKAN DATA MINING PADA JUMLAH PENUMPANG DI BANDARA HANG NADIM INFORMASI ARTIKEL KATA KUNCI,” *CBIS JOURNAL*, 2018, [Online]. Available: http://ejournal.upbatam.ac.id/index.php/cbis

[18] F. Gullo, “From patterns in data to knowledge discovery: What data mining can do,” in *Physics Procedia*, Elsevier B.V., 2015, pp. 18–22. doi: 10.1016/j.phpro.2015.02.005.

[19] E. E. Barito, J. T. Beng, and D. Arisandi, “PENERAPAN ALGORITMA C4.5 UNTUK KLASIFIKASI MAHASISWA PENERIMA BANTUAN SOSIAL COVID-19”.

[20] Y. P. Pasrun, “PENERAPAN METODE NAÏVE BAYES UNTUK KLASIFIKASI PENYAKIT ISPA,” *semanTIK*, vol. 5, no. 2, pp. 295–302, Dec. 2019, doi: 10.5281/zenodo.3515564.

[21] J. Nurhakiki *et al.*, “Studi Kepustakaan: Pengenalan 4 Algoritma Pada Pembelajaran Deep Learning Beserta Implikasinya,” *Jurnal Pendidikan Berkarakter*, no. 1, pp. 270–281, 2024, doi: 10.51903/pendekar.v2i1.598.

[22] S. Anggreani *et al.*, “Pengantar dan Pengenalan Konsep Dasar Algoritma Pemrograman,” *Jurnal Pendidikan Berkarakter*, no. 1, pp. 282–294, 2024, doi: 10.51903/pendekar.v2i1.599.

[23] R. R. Asyrofi and R. Asyrofi, “IMPLEMENTASI APLIKASI JUPYTER NOTEBOOK SEBAGAI ANALISIS KRETERIA PLAGIASI DENGAN TEKNIK SIMANTIK,” *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 8, no. 2, pp. 627–637, May 2023, doi: 10.29100/jipi.v8i2.3699.

[24] K. Surya Ningsih, N. Jamilah Aruan, and A. Taufik Al Afkari Siahaan, “APLIKASI BUKU TAMU MENGGUNAKAN FITUR KAMERA DAN AJAX BERBASIS WEBSITE PADA KANTOR DISPORA KOTA MEDAN,” *SITek*, vol. 1, no. 3, pp. 94–99, Dec. 2022.

[25] T. Susilawati, F. Yuliansyah, M. Romzi, and R. Aryani, “MEMBANGUN WEBSITE TOKO ONLINE PEMPEK NTHREE MENGGUNAKAN PHP DAN MYSQL,” *JTIM*, vol. 3, no. 1, pp. 35–44, Jan. 2020.

[26] A. Tholib, *Implementasi Algoritma Machine Learning Berbasis Web dengan Framework Streamlit*, 1st ed. Probolinggo: Pustaka Nurja, 2023.

[27] I. Kurniawan, D. Cahya Putri Buani, W. Apriliah, R. Amegia Saputra, and P. Korespondensi, “IMPLEMENTASI ALGORITMA RANDOM FOREST UNTUK MENENTUKAN PENERIMA BANTUAN RASKIN,” *Journal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 10, no. 2, pp. 421–428, Apr. 2023, doi: 10.25126/jtiik.202396225.

[28] I. Made, B. Adnyana, S. B. Jln, and R. Puputan, “Penerapan Feature Selection untuk Prediksi Lama Studi Mahasiswa,” *JURNAL SISTEM DAN INFORMATIKA*, vol. 13, no. 2, pp. 72–76, May 2019.

[29] D. Kurniadi, F. Nuraeni, M. Firmansyah, and P. Korespondensi, “KLASIFIKASI MASYARAKAT PENERIMA BANTUAN LANGSUNG TUNAI DANA DESA MENGGUNAKAN NAÏVE BAYES DAN SMOTE,” *JTIIK*, vol. 10, no. 2, pp. 309–320, Apr. 2023, doi: 10.25126/jtiik.2023106453.

[30] M. P. Pulungan, A. Purnomo, and A. Kurniasih, “Penerapan SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Kepribadian MBTI Menggunakan Naive Bayes Classifier,” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 5, pp. 1033–1042, Oct. 2024, doi: 10.25126/jtiik.2024117989.

[31] R. Indransyah, Y. H. Chrisnanto, and P. N. Sabrina, “KLASIFIKASI SENTIMEN PERGELARAN MOTOGP DI INDONESIA MENGGUNAKAN ALGORITMA CORRELATED NAÏVE BAYES CLASIFIER,” *INFOTECH*, vol. 8, no. 2, pp. 60–66, Dec. 2022, doi: 10.31949/infotech.v8I2.3103.

[32] I. Daniel, A. F. Limas Ptr, and A. Ichsan, “Klasifikasi Risiko Penyakit Jantung Dengan Multilayer Perceptron,” *Data Sciences Indonesia (DSI)*, vol. 4, no. 1, pp. 78–82, Sep. 2024, doi: 10.47709/dsi.v4i1.4667.

[33] P. L. Romadhoni, B. A. Kusuma, and W. M. Baihaqi, “KOMPARASI METODE PEMBELAJARAN MESIN UNTUK IMPLEMENTASI PENGAMBILAN KEPUTUSAN DALAM MENENTUKAN PROMOSI JABATAN KARYAWAN,” *JATI*, vol. 6, no. 2, pp. 622–628, Sep. 2022.

[34] Muhammad Akram Fais, M. Revano Ananda Lubis, Annisa Aulia, and Indri Syafitri, “Implementasi Algoritma Decision Tree untuk Klasifikasi Serangan Jantung,” *Jurnal Sistem Informasi dan Ilmu Komputer*, vol. 1, no. 4, pp. 207–212, Nov. 2023, doi: 10.59581/jusiik-widyakarya.v1i4.1895.

# LAMPIRAN

**KONTAK PENULIS DAN KONTRIBUTOR PENELITIAN**

Informasi korespondensi penulis dan kontributor penelitian adalah sebagai berikut :

**PENULIS**

Nama Lengkap Mahasiswa : Dheandy Divallino

NIM : 10521049

e-Mail yang aktif : Dheandy.10521049@mahasiswa.unikom.ac.id

Kode Prodi PDDIKTI : 57201

**KONTRIBUTOR**

**Pembimbing I**

Nama Lengkap Dosen : Agus Nursikuwagus,Dr.Assoc.Prof.ST.,MT.,MM.,MOS.,MTA

NIP : 41277026119

NIDN/NIDK : 0009077509

e-Mail yang aktif : agusnursikuwagus@email.unikom.ac.id



**FOTO FORMAL BERWARNA 3X4**

**DAFTAR RIWAYAT HIDUP**

**DATA PRIBADI**

Nama : Dheandy Divallino  
Jenis Kelamin : Laki – laki  
Tempat, Tanggal Lahir : Bekasi, 05 Januari 2004  
Kewarganegaraan : Indonesia  
Status Perkawinan : Belum Menikah  
Agama : Islam  
Email : [dheandyvalino@gmail.com](mailto:dheandyvalino@gmail.com) (email aktif)  
Alamat Lengkap : Jl. Terusan Saluyu, Perumahan Kawistara, No. C10  
Nomor Telepon : 083168139437

**PENDIDIKAN FORMAL**

2010 – 2015 : SDN 114 Pekanbaru  
2015 – 2018 : SMPN 4 Pekanbaru  
2018 – 2021 : SMAN 12 Bandung  
2018 – Sekarang : Program S1 Sistem Informasi di Universitas Komputer Indonesia

**PRESTASI**

1. Juara I Lomba Lukis Semarak Seni Lukis Pekanbaru (2015)
2. Juara I Melukis Pekan Raya Biologi Pekanbaru (2017)
3. Juara II Melukis FLS2N Tingkat Provinsi Riau (2017)
4. *Certificate of Completion for successfully completing* Bangkit (2024)

Demikian Daftar Riwayat Hidup ini. Saya buat dengan sesungguhnya.

Hormat saya,

**Dheandy Divallino**

