**Metode *Multi-Layer Perceptron* untuk Klasifikasi Penerima Bantuan Sosial Program Keluarga Harapan (PKH) di Kelurahan Cipamokolan**

***Multi-Layer Perceptron Method for Classifying Beneficiaries of The Family Hope Program (PKH) Social Assistance in Cipamokolan Village***

**Dheandy Divallino**1, **Agus Nursikuwagus**2

1,2Program Studi Sistem Informasi, Universitas Komputer Indonesia, Bandung, Indonesia

\*E-mail: dheandyvalino@gmail.com

***Abstrak***

*Bantuan sosial berpotensi tidak tepat sasaran jika proses seleksinya dilakukan secara musyawarah. Penelitian ini bertujuan menerapkan metode Multi-Layer Perceptron (MLP) untuk mengklasifikasikan kelayakan calon penerima Program Keluarga Harapan (PKH) guna meningkatkan efisiensi dan akurasi. Metode kuantitatif digunakan dengan dataset sebanyak 422 data dari Kelurahan Cipamokolan tahun 2023, yang dibagi menjadi 338 data latih dan 84 data uji. Proses penelitian meliputi preprocessing, transformasi fitur, dan pelatihan model MLP. Hasil pengujian model menunjukkan performa sangat baik dengan akurasi 95,70 %, precision 95,74%, recall 97,83%, dan f1-score 96,77%. Sebuah aplikasi website berbasis framework Streamlit juga dikembangkan untuk memudahkan prediksi. Penelitian ini membuktikan bahwa MLP dapat menjadi alat pendukung keputusan yang optimal untuk mengurangi kesalahan dalam seleksi penerima bantuan, sehingga penyaluran bantuan sosial menjadi lebih tepat sasaran. Penelitian ke depan dapat diperluas dengan menambah data dari wilayah lain dan mengeksplorasi parameter model lebih mendalam.*

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

***Abstract***

*Social assistance has the potential to be inaccurately targeted if the selection process is conducted through communal deliberation (musyawarah). This research aims to implement the Multi-Layer Perceptron (MLP) method to classify the eligibility of prospective recipients of the Family Hope Program (Program Keluarga Harapan/PKH) to improve efficiency and accuracy. A quantitative method was employed using a dataset of 422 records from Cipamokolan Village for the year 2023, which was divided into 338 training data points and 84 testing data points. The research process included preprocessing, feature transformation, and MLP model training. The model test results demonstrated excellent performance with an accuracy of 95.70%, precision of 95.74%, recall of 97.83%, and an f1-score of 96.77%. A web application based on the Streamlit framework was also developed to facilitate easy prediction. This research proves that MLP can be an optimal decision-support tool for reducing errors in the selection of aid recipients, thereby making the distribution of social assistance more targeted. Future research can be expanded by adding data from other regions and exploring model parameters more deeply.*

***Keywords:*** *Data mining, Family Hope Program, Classification, Multi-Layer Perceptron, Streamllit.*

1. Pendahuluan

Kemiskinan merupakan salah satu faktor utama yang menghambat pembangunan suatu wilayah karena sifat permasalahannya yang kompleks dan sulit diselesaikan secara mandiri oleh daerah tersebut. Kemiskinan dapat dipahami sebagai kondisi ketika keluarga tidak mampu memenuhi kebutuhan dasar untuk hidup layak sesuai standar di lingkungannya [1]. Untuk mengatasi permasalahan tersebut, pemerintah melaksanakan berbagai program bantuan sosial, salah satunya adalah Program Keluarga Harapan (PKH) yang bertujuan menurunkan angka kemiskinan melalui dukungan di bidang pendidikan, kesehatan, dan kesejahteraan keluarga [2].

Salah satu wilayah pelaksana program ini adalah Kelurahan Cipamokolan, Kecamatan Rancasari, Kota Bandung. Berdasarkan data Badan Pusat Statistik dan Dinas Sosial Kota Bandung tahun 2023, tercatat sebanyak 236 keluarga penerima PKH dari total 6.339 keluarga. Meskipun mekanisme pendataan penerima bantuan telah berjalan sesuai prosedur, proses verifikasi dan validasi masih dilakukan secara manual melalui musyawarah kelurahan, RW, dan RT. Cara konvensional ini sering menyebabkan keterlambatan pembaruan data serta potensi ketidaktepatan sasaran penyaluran bantuan. Kondisi ini sejalan dengan temuan Badan Pemeriksa Keuangan (BPK) yang menunjukkan masih adanya penerima bantuan yang sebenarnya sudah tidak layak, seperti warga yang telah meninggal dunia atau memiliki penghasilan di atas UMP/UMK (Bisnis.com, 2024).

Permasalahan tersebut menegaskan perlunya solusi berbasis teknologi *data mining* untuk membantu proses klasifikasi penerima bantuan secara lebih akurat dan efisien. *Data mining* adalah proses menganalisis data historis untuk menemukan pola, keteraturan, atau hubungan tertentu yang dapat mendukung pengambilan keputusan di masa depan [3]. Salah satu teknik yang relevan adalah klasifikasi, yaitu proses membangun model untuk memprediksi kelas objek berdasarkan variabel tertentu [4]. Dalam konteks penelitian ini, klasifikasi berperan dalam membedakan antara keluarga yang layak dan tidak layak menerima PKH.

Salah satu algoritma yang banyak digunakan dalam klasifikasi adalah Artificial Neural Network (ANN) [5]. ANN dirancang menyerupai cara kerja otak manusia melalui lapisan *input*, lapisan tersembunyi, dan lapisan *output* yang saling terhubung [6]. Keunggulan ANN terletak pada kemampuannya menangani data non-linear, menyesuaikan bobot secara adaptif, serta mengenali pola yang kompleks. Salah satu arsitektur ANN yang populer adalah *Multi-Layer Perceptron* (MLP) [7]. MLP termasuk jaringan feedforward dengan minimal satu lapisan tersembunyi yang mampu melakukan pemodelan fungsi non-linear, ekstraksi fitur otomatis, serta menghasilkan performa klasifikasi yang tinggi pada data tabular [8], [9].

Beberapa penelitian terdahulu telah mengkaji penerapan *data mining* dan neural network untuk kasus serupa. Agus Nursikuwagus dkk. (2023) membangun sistem klasifikasi motif batik menggunakan Convolutional Neural Network (CNN) dan mencapai akurasi 98,5% [10]. Persamaannya terletak pada penggunaan deep learning, namun berbeda dari penelitian ini karena objeknya berupa data citra, sedangkan penelitian ini menggunakan data tabular dengan MLP . Senang Hati Gulo dan Andre H. Lubis (2024) mengembangkan model MLP untuk klasifikasi penduduk kurang mampu dan memperoleh akurasi 96,9% [8]. Persamaannya adalah sama-sama menggunakan MLP untuk data sosial-ekonomi, sementara penelitian ini lebih spesifik pada klasifikasi penerima PKH dengan label biner (layak/tidak layak). I Wayan Supriana dkk. (2023) mengembangkan sistem prediksi PKH berbasis ANN dengan algoritma backpropagation [11]. Fokus penelitian tersebut adalah prediksi jumlah bantuan, sedangkan penelitian ini menekankan klasifikasi status penerima.

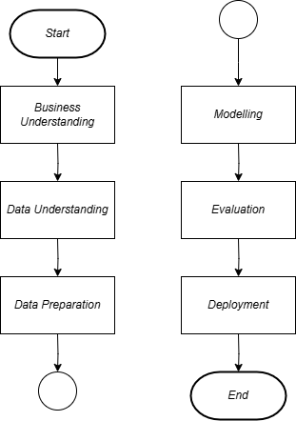
Berdasarkan ulasan tersebut, kebaruan penelitian ini terletak pada penerapan metode *Multi-Layer Perceptron* (MLP) untuk klasifikasi penerima PKH di Kelurahan Cipamokolan, dengan fokus pada kelayakan penerimaan bantuan. Penelitian ini tidak hanya menghasilkan model prediktif yang lebih akurat, tetapi juga menyiapkan langkah deployment berbasis *web*, sehingga dapat langsung digunakan oleh pemerintah daerah dalam proses seleksi calon penerima.

Adapun tujuan penelitian ini adalah untuk menentukan kelayakan keluarga penerima bantuan PKH berdasarkan hasil klasifikasi menggunakan metode *Multi-Layer Perceptron* (MLP). Selain itu, penelitian ini juga bertujuan untuk menjelaskan tahapan preprocessing data yang digunakan, membangun serta mengevaluasi model prediksi klasifikasi PKH berbasis MLP, mengukur tingkat akurasi hasil prediksi, serta menerapkan deployment model berbasis *web* yang dapat dimanfaatkan sebagai alat bantu pengambilan keputusan.

1. Metode Penelitian

Metode penelitian yang digunakan adalah metode kuantitatif karena melibatkan pengolahan data numerik dan algoritma *machine learning* untuk klasifikasi penerima bantuan sosial PKH. Data penelitian diperoleh dari sumber primer dan sekunder. Data primer dikumpulkan melalui observasi langsung di Kantor Kecamatan Rancasari serta wawancara dengan narasumber dari Seksi Kesejahteraan Sosial. Sementara itu, data sekunder diperoleh dari buku, jurnal, dan dokumen relevan melalui kegiatan membaca, mempelajari, dan menganalisis. Penelitian dilaksanakan di Kantor Kecamatan Rancasari selama enam bulan, yaitu Februari–Juli 2025.

Tahapan penelitian digambarkan melalui diagram sebagai berikut:



Gambar . Alir Penelitian  
(Sumber: Jurnal Teknologi Informasi dan Ilmu Komputer)

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

2.1 *Business Understanding*

Tahap awal penelitian berfokus pada pemahaman tujuan dan kebutuhan bisnis yang diterjemahkan menjadi rumusan masalah untuk diatasi dengan teknik *data mining* [12]. Kegiatan dimulai dengan wawancara, baik langsung maupun tidak langsung, guna mengidentifikasi permasalahan dan mencari solusi. Wawancara dilakukan dengan pegawai Seksi Kesejahteraan Sosial Kecamatan Rancasari.

2.2 *Data Understanding*

Tahap ini bertujuan memahami karakteristik data penerima bantuan sosial PKH Kelurahan Cipamokolan tahun 2023 melalui identifikasi, evaluasi kualitas, dan eksplorasi data. Statistik deskriptif digunakan untuk menggambarkan karakteristik umum seperti mean dan standar deviasi, sedangkan visualisasi membantu memahami distribusi, mendeteksi data hilang, serta menghapus fitur bermasalah guna mendukung pembangunan model *machine learning* yang lebih baik [12].

2.3 *Data Preparation*

Tahap ini berfokus pada penyusunan dataset akhir penerima bantuan sosial PKH Kelurahan Cipamokolan tahun 2023 yang siap digunakan dalam model. Data mentah diproses melalui tahapan data preparation agar memenuhi standar yang diperlukan [12]. Adapun tahapan pengolahan data yang akan dilakukan dalam penelitian ini meliputi beberapa proses sebagai berikut:

1. *Data Selection*, dilakukan dengan teknik *Feature Subset Selection* untuk memilih fitur paling relevan yang berkontribusi pada peningkatan akurasi model sekaligus mengurangi dimensi data. Penelitian ini menggunakan *Correlation Feature Selection* (CFS), yang menilai korelasi tiap fitur dengan variabel *output*, mempertahankan fitur dengan korelasi tinggi, dan membuang fitur dengan korelasi rendah [13].
2. *Data Cleansing*, adalah proses pembersihan data dari nilai kosong, tidak valid, fitur kurang relevan, missing value, maupun *outlier*. Data kosong dideteksi dengan fungsi ‘*isna(’* dan diisi menggunakan *‘fillna()*’ dengan *mean*, median, atau modus sesuai karakteristik fitur. *Outlier* dideteksi dengan metode *Z-Score* dan ditangani melalui penyesuaian atau penghapusan agar distribusi data lebih bersih dan representative [14].
3. *Data Transformation*, adalah proses mengubah data agar sesuai untuk diproses dengan teknik *data mining*. Pada penelitian ini, fitur kategorikal diubah menjadi numerik menggunakan *Label Binarizer* untuk label biner (0/1) dan *One-Hot Encoding* untuk fitur dengan lebih dari dua kategori, sehingga dapat dibaca dan diolah oleh algoritma *machine learning* [14].
4. *Data Balancing*, bertujuan menyeimbangkan distribusi kelas dalam dataset agar model tidak bias terhadap kelas mayoritas. Penelitian ini menggunakan metode *Synthetic Minority Oversampling Technique* (SMOTE), yaitu teknik *oversampling* yang menambah data pada kelas minoritas sehingga prediksi lebih seimbang dan akurat [15].
5. *Data Splitting*, membagi dataset menjadi data latih dan data uji dengan perbandingan umum 80:20 atau 70:30, bertujuan memastikan model memiliki performa baik serta mampu melakukan generalisasi pada data baru [14].

2.4 Modelling

Tahap ini mencakup pemilihan model dengan mempertimbangkan karakteristik dataset yang bersifat tabular dan berlabel, sehingga sesuai menggunakan algoritma *supervised learning* [12], [16]. Berdasarkan kajian literatur, metode *Multi-Layer Perceptron (MLP)* dipilih karena terbukti efektif dalam klasifikasi data tabular, mampu menangkap hubungan nonlinear antar fitur, serta fleksibel dalam penggunaan fungsi aktivasi seperti *ReLU* atau *sigmoid*. Dengan keunggulan tersebut, MLP digunakan sebagai algoritma utama dalam penelitian ini. Persamaan fungsi aktivasi ditulis sebagai berikut [17].

(1)

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

(2)

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

(3)

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

(4)

Secara umum, proses per *layer* dituliskan:

(5)

Keterangan:

Z : Nilai *input* total sebelum aktivasi

: Bobot antara *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

Tahap pemodelan bertujuan menghasilkan *Confusion matrix* untuk mengevaluasi kinerja model klasifikasi. *Confusion matrix* menampilkan perbandingan prediksi model dengan kondisi aktual dalam empat kategori: *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Dalam konteks penerima bantuan, FP berarti orang tidak layak dinyatakan layak, sedangkan FN berarti orang layak tidak mendapat bantuan. Matriks ini membantu menilai akurasi, kesalahan, serta menjadi dasar perbaikan model [18].

2.5 Evaluation

Model *Multi-Layer Perceptron (MLP)* untuk klasifikasi penerima bantuan sosial PKH dievaluasi menggunakan metrik akurasi, presisi, *recall*, dan *F1-score*. Nilai tersebut dihitung berdasarkan *Confusion matrix*, sehingga memungkinkan analisis menyeluruh terhadap efektivitas konfigurasi dan fungsi aktivasi dalam memprediksi penerima bantuan [12], [19], [20].

1. *Accuracy*

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

(6)

1. *Precision*

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

(7)

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:

(8)

1. *F1-score*

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

(9)

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

2.6 Deployment

Model *Multi-Layer Perceptron (MLP)* yang telah dibangun akan diimplementasikan ke dalam aplikasi *web* berbasis *Streamlit*. Aplikasi ini memungkinkan pengguna meng*input* data calon penerima bantuan sosial PKH dan memperoleh prediksi kelayakan secara langsung [12].

1. Hasil dan Pembahasan

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.

3.1 Hasil

Adapun hasil dari penelitian ini akan dijelaskan berdasarkan tahapan penelitian. Untuk lebih jelasnya dapat dilihat pada sub poin berikut:

3.1.1 *Business Understanding*

Berdasarkan wawancara dengan salah satu pegawai Seksi Kesejahteraan Sosial Kecamatan Rancasari, diperoleh pemahaman yang lebih mendalam mengenai pelaksanaan Program Bantuan Sosial Keluarga Harapan (PKH) di lapangan. Wawancara ini bertujuan untuk mengidentifikasi permasalahan, memahami alur data, serta mengkaji kriteria yang digunakan dalam menentukan keluarga penerima bantuan. Dari hasil wawancara, diketahui bahwa proses verifikasi data masih dilakukan secara manual sehingga berpotensi menimbulkan keterlambatan dan ketidaktepatan dalam penyaluran bantuan. Oleh karena itu, diperlukan sistem klasifikasi yang dapat membantu menentukan keluarga yang berhak menerima bantuan sosial PKH secara lebih efektif dan akurat.

3.1.2 *Data Understanding*

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

1. Sumber

Sumber data primer penelitian ini adalah data penerima bantuan sosial PKH tahun 2023 di Kelurahan Cipamokolan, yang diperoleh melalui wawancara dan koordinasi dengan koordinator PKH Kecamatan Rancasari serta pegawai Seksi Kesejahteraan Sosial.

1. Validator

Data penelitian ini telah melalui proses verifikasi dan validasi resmi. Data diperoleh dari Koordinator PKH Kecamatan Rancasari berdasarkan pendataan Kelurahan Cipamokolan, kemudian dimusyawarahkan di tingkat RT, RW, dan kelurahan, serta diverifikasi lebih lanjut oleh Kecamatan dan Dinas Sosial. Data masih berupa data mentah dengan 422 record dan 15 fitur, meliputi usia, penghasilan, pekerjaan, status pernikahan, jumlah tanggungan, pendidikan, status rumah, jenis kelamin, kriteria PKH (6 fitur), serta status kelayakan.

1. Keterikatan Antar Fitur

Berdasarkan data dan hasil wawancara, setiap fitur dalam penelitian ini memiliki keterikatan dengan kriteria PKH dalam menentukan kelayakan penerima bantuan. Usia, penghasilan, pekerjaan, status pernikahan, jumlah tanggungan, pendidikan, dan status rumah menjadi faktor yang mencerminkan kondisi sosial ekonomi keluarga. Fitur-fitur tersebut saling berkaitan dan berkontribusi dalam proses klasifikasi untuk menentukan apakah sebuah keluarga layak atau tidak layak menerima bantuan sosial PKH.

1. 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 1. 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 |

Tabel 1 menunjukkan rata-rata usia kepala keluarga 55 tahun dengan penghasilan Rp1.117.606 dan tanggungan 2 orang, sedangkan standar deviasi 15 tahun menggambarkan sebaran usia 40–70 tahun.

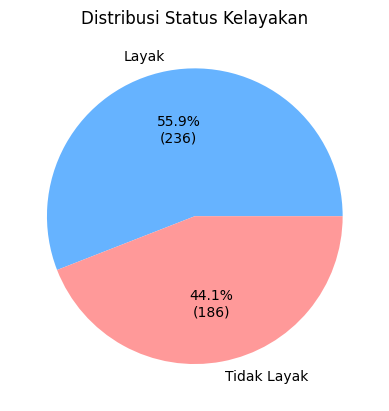
Tabel 2. Statistik Deskriptif Fitur Kategorikal

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 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. 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 2 dan 3, statistik deskriptif fitur kategorikal menunjukkan distribusi tiap kategori. Pada fitur pekerjaan terdapat 9 kategori, dengan mayoritas kepala keluarga tidak memiliki pekerjaan (142 orang). Status pernikahan didominasi kategori menikah (254 orang), tingkat pendidikan terbanyak adalah SMA (138 orang), dan status rumah mayoritas milik sendiri (105 orang). Fitur kriteria PKH dan status kelayakan masing-masing memiliki 2 kategori, di mana dari 422 data terdapat 236 keluarga layak menerima bantuan. Distribusi ini menggambarkan proporsi dan dominasi tiap kelompok dalam dataset.



Gambar 2. Chart Status Kelayakan

Gambar 2 menampilkan diagram batang status kelayakan, di mana 55,9% kepala keluarga tercatat layak. Visualisasi ini menunjukkan distribusi kelas sekaligus mengidentifikasi potensi ketidakseimbangan data. Analisis deskriptif ini menjadi dasar awal untuk memahami pola sebelum tahap klasifikasi.

3.1.3 *Data Preparation*

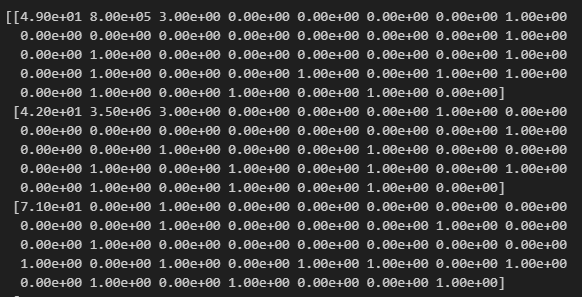
Data mentah yang diperoleh pada tahapan sebelumnya perlu dilakukan *Data Preparation*. Proses yang dilakukan dalam pengolahan data pada penelitian ini adalah sebagai berikut

1. *Data cleansing*

Tahap ini meliputi persiapan data dengan menghapus fitur tidak relevan, memisahkan fitur kategori dan target, serta menangani missing value.

1. *Data transformation*

Pada tahap transformasi data, digunakan *Label Binarizer* untuk fitur target dan *One-Hot Encoding* untuk fitur kategori.

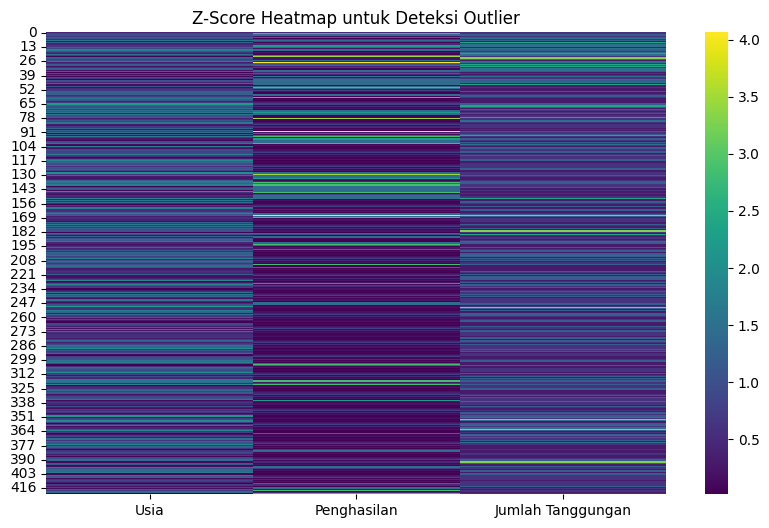


Gambar 3. Hasil Transformasi Fitur Kategori

Gambar 3 menunjukkan bahwa transformasi *One-Hot Encoding* menambah jumlah fitur dari 14 menjadi 39. Fitur tambahan berasal dari kategori pekerjaan, status pernikahan, pendidikan, status rumah, jenis kelamin, serta kriteria PKH (ibu hamil/menyusui, balita, anak prasekolah, anak sekolah, lansia, dan penyandang disabilitas). Fitur target diubah ke format numerik (0 dan 1) agar dapat diproses optimal oleh model dalam pelatihan dan prediksi.

1. *Outlier*

*Outlier* adalah nilai yang menyimpang jauh dari data lainnya. Pengecekan dilakukan setelah tahap transformasi dan seleksi fitur.

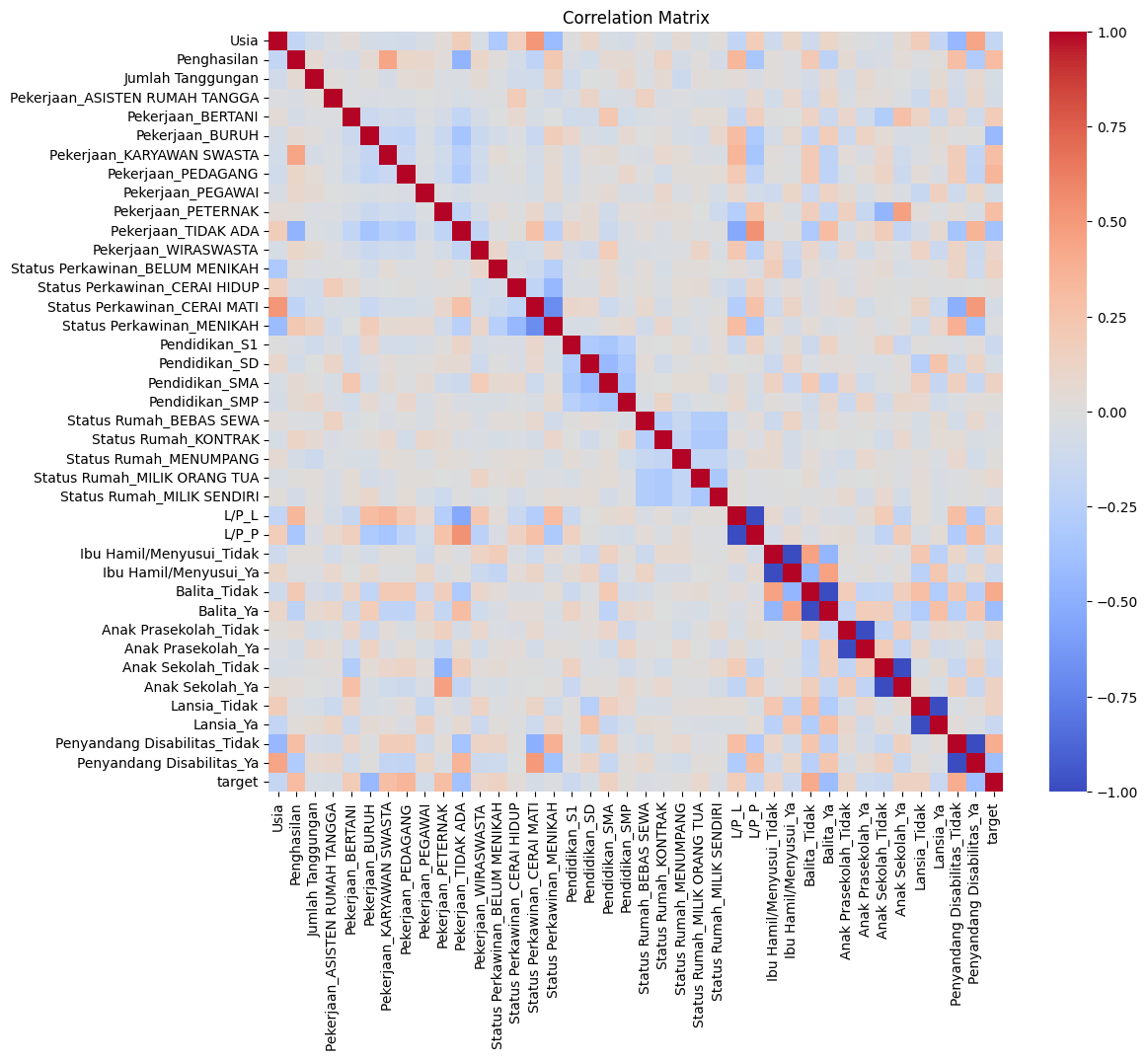


Gambar 4. Visualisasi *Outlier*

Berdasarkan Gambar 4, warna ungu menunjukkan nilai *Z-Score* mendekati nol, sedangkan kuning menunjukkan nilai tinggi. Data dengan *Z-Score* > 3 dihapus agar kualitas data tetap terjaga dan model tidak dipengaruhi *outlier* [14].

1. *Data selection*

Tahap *Data Selection* dilakukan setelah transformasi data dan penanganan *outlier*, karena *Correlation Feature Selection* (CFS) memerlukan data numerik yang bersih [13]. Setelah semua fitur kategorikal ditransformasikan, CFS digunakan untuk memilih fitur yang paling signifikan terhadap target klasifikasi.

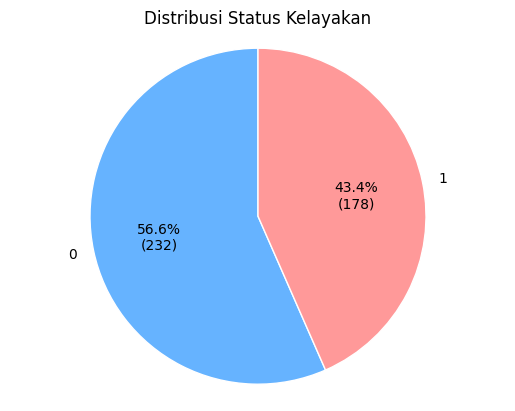


Gambar 5. *Correalation Matrix*

Gambar 5 menampilkan *Correlation Matrix* yang menunjukkan hubungan antar fitur dengan target. Nilai korelasi berkisar -1 hingga 1, di mana nilai mendekati 1 berarti korelasi positif kuat, mendekati -1 korelasi negatif kuat, dan mendekati 0 menunjukkan hubungan lemah [13]. Dari 39 fitur, terpilih 30 fitur yang memiliki hubungan signifikan dengan target klasifikasi.

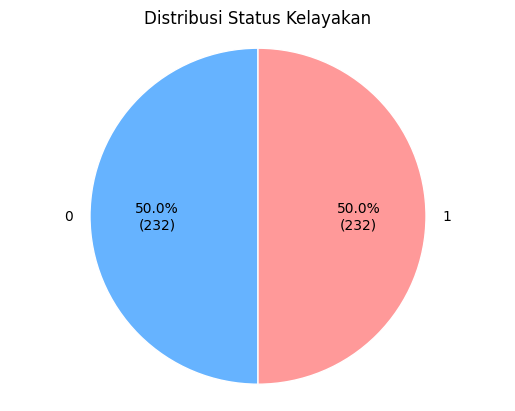
1. Data balancing

Tahap data balancing menggunakan metode SMOTE untuk mengatasi ketidakseimbangan kelas Layak dan Tidak Layak, sebagaimana ditunjukkan pada Gambar 7.



Gambar 6. Diagram *Imbalanced Class*

Berdasarkan diagram, terdapat ketidakseimbangan kelas dengan 56,6% layak (232 data) dan 43,4% tidak layak (178 data). Untuk mengatasinya, digunakan teknik SMOTE.



Gambar 7. Diagram Status Kelayakan Setelah Proese SMOTE

Berdasarkan Gambar 7, jumlah kelas Layak dan Tidak Layak sudah seimbang setelah proses balancing, sehingga model tidak bias dan dapat dilanjutkan ke tahap berikutnya.

1. Data splitting

Pada tahap ini, data dibagi menjadi dua bagian: data latih untuk melatih model dan data uji untuk mengukur kinerjanya. Pembagian data ditunjukkan pada tabel.

Tabel 5. Pembagian Proporsi Data

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

Berdasarkan Tabel 5, data splitting dilakukan dengan dua skenario: rasio 70:30 dan 80:20. Pada rasio 70:30, 70% data digunakan untuk pelatihan dan 30% untuk pengujian, sedangkan pada rasio 80:20, 80% untuk pelatihan dan 20% untuk pengujian. Tujuannya adalah membandingkan performa model pada proporsi data yang berbeda guna melihat pengaruhnya terhadap akurasi dan generalisasi.

3.1.4 *Modelling*

Pada tahap ini, pemodelan dilakukan dengan algoritma *Multi-Layer Perceptron* (MLP) menggunakan *library scikit-learn*. Model dibangun dengan dua *hidden layer* berisi 100 dan 50 *neuron*, dilatih menggunakan parameter *max\_iter=300* dan *random\_state=42* untuk konsistensi. nilai aktivasi masing-masing *neuron* pada *hidden layer* 1 dihitung dengan persamaan (1):

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

𝑧1 = ((−0,49 . − 1,24)+. . . +(−0,56 . −2,19)) + 0,11 = 0,21

Hasil ini kemudian dihitung menggunakan fungsi aktivasi *ReLu* pada persamaan (3):

𝑓(𝑧1) = 𝑚𝑎𝑥(0, 0,21) = 0,21

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:

𝑧1 = ((0,21 . − 1,75) + (0 . 9,5)+. . . ) + 0,11 = −0,05

Kemudian diaplikasikan fungsi aktivasi *ReLu*:

𝑓(𝑧1) = 𝑚𝑎𝑥(0, −0,05) = 0

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

𝑧1 = ((0. 0,009) + (0 . 0,33)+. . . ) + (−0,19) = 4,95

Nilai ini kemudian diolah dengan fungsi aktivasi *sigmoid* pada persamaan (4):

Jika probabilitas *sigmoid* > 0.5, data diklasifikasikan sebagai “Layak”. Ambang 0.5 dipilih karena merupakan titik tengah umum pada klasifikasi biner.

3.1.5 Metrik

Evaluasi dilakukan pada data latih untuk menilai kemampuan model mempelajari pola, dan pada data uji untuk mengukur generalisasi terhadap data baru. Perbandingan keduanya membantu mendeteksi potensi *overfitting* atau *underfitting* [4]. Kinerja model dievaluasi pada data uji menggunakan *Confusion matrix* untuk menghitung *Accuracy*, *Precision*, *Recall*, dan *F1-Score*. Berdasarkan metrik evaluasi, dijelaskan bahwa terdapat 45 *True Positif* (TP), 2 *False Negatif* (FN), 2 *False Positif* (FP), dan 44 *True Negatif* (TN).

Tabel 6. *Confusion matrix*

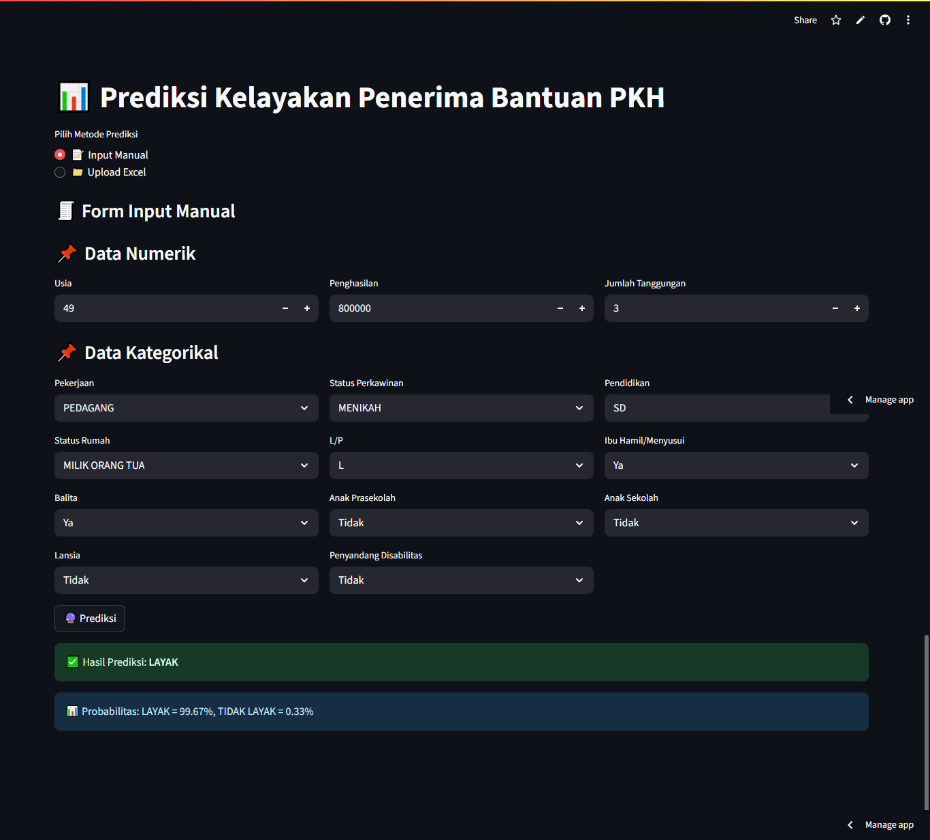
|  |  |  |
| --- | --- | --- |
|  | Prediksi | |
| Aktual | *True* | *False* |
| *True* | TP=45 | FN=2 |
| *False* | FP=2 | TN=44 |

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:
2. Untuk menghitung nilai *Precision* menggunakan rumus persamaan (7), hasil perhitungan sebagai berikut:
3. Untuk menghitung nilai *Recall* menggunakan rumus persamaan (8), hasil perhitungan sebagai berikut:
4. Untuk menghitung nilai F1-Score menggunakan rumus persamaan (9), hasil perhitungan sebagai berikut:

3.1.6 *Deployment*

Tahap ini mengimplementasikan model MLP ke dalam aplikasi *web* berbasis Streamlit, menyediakan *input* manual untuk satu data dan *input* file Excel untuk multi data.



Gambar 8. Tampilan *Input* Data Tunggal

3.2 Pembahasan

Pada proporsi 70:30, data latih berjumlah 295 dan data uji 127; sedangkan pada proporsi 80:20, data latih 338 dan data uji 84. Model menggunakan arsitektur 100-50 *neuron* dengan 300 *epoch*, dan hasil klasifikasi ditunjukkan pada Tabel 7.

Tabel 7. 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 Tabel 7, proporsi 80:20 menghasilkan kinerja lebih baik dengan akurasi 95,70%, dibanding 70:30 yang hanya 93,57%, demikian juga pada *recall*, *precision*, dan *F1-score*. Berdasarkan keunggulan proporsi 80:20, proporsi ini digunakan untuk membandingkan MLP dengan *Naïve Bayes* dan *Decision Tree*. Semua algoritma diuji pada dataset yang sama, dengan evaluasi menggunakan akurasi, *recall*, *precision*, dan *F1-score*.

Tabel 8. 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 8, perbandingan kinerja MLP, *Naïve Bayes*, dan *Decision Tree* dengan proporsi data 80:20 menunjukkan bahwa MLP memberikan performa terbaik dengan akurasi 95,70%, *recall* 97,83%, *precision* 95,74%, dan *F1-score* 96,77%, menunjukkan keseimbangan tinggi antara deteksi kelas benar dan ketepatan prediksi. *Naïve Bayes* memiliki *recall* 100% tetapi akurasi dan *precision* rendah (65,59% dan 58,97%), sehingga *F1-score* hanya 74,19%. Hal ini disebabkan *Naïve Bayes* lebih optimal untuk data kategorikal, sedangkan dataset ini telah ditransformasikan menjadi numerik untuk MLP. *Decision Tree* menunjukkan performa cukup baik dengan akurasi 93,54%, *recall* 95,65%, *precision* 91,67%, dan *F1-score* 93,62%, karena mampu bekerja langsung dengan data numerik [21]. Kesimpulannya, MLP merupakan algoritma terbaik untuk klasifikasi kelayakan penerima bantuan sosial PKH di Kelurahan Cipamokolan.

1. Kesimpulan

Berdasarkan hasil penelitian yang diperoleh, maka dapat ditarik kesimpulan yaitu: 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%, nilai *precision* sebesar 95,74%, nilai *recall* sebesar 97,83%, nilai F1-score sebesar 96,77%. 2) 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. 3) Implementasi *web*site 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 untuk penelitian selanjutnya yaitu: 1) 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) Melakukan eksperimen lebih lanjut menggunakan teknik seperti *grid search* atau *random search* untuk mendapatkan konfigurasi optimal. 3) Menggunakan algoritma lain seperti *Random Forest, Support Vector Machine* (SVM), atau *Gradient Boosting* untuk mengetahui apakah terdapat metode lain yang mampu memberikan performa lebih baik dibandingkan MLP. 4) Menambahkan 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. 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.

[2] 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.

[3] 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

[4] 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

[5] 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.

[6] 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.

[7] 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.

[8] 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

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

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

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.