

## IMPLEMENTASI ALGORITMA K-NEAREST NEIGHBOR DALAM SISTEM KLASIFIKASI CALON PENERIMA BEASISWA PROGRAM INDONESIA PINTAR (PIP) (STUDI KASUS : SMP NEGERI 2 SIBOLANGIT)

Ekin Adhi Guna, Debi Yandra Niska

Ilmu Komputer, Universitas Negeri Medan

Jalan Willem Iskandar, Percut Sei Tuan, Deli Serdang, Sumatera Utara, Indonesia

ekinadhiguna@gmail.com

### ABSTRAK

Pendidikan berperan penting dalam kemajuan bangsa, namun di Indonesia akses pendidikan masih terkendala biaya. Untuk mengatasinya, pemerintah meluncurkan Program Indonesia Pintar (PIP) guna pemerataan akses pendidikan. SMP Negeri 2 Sibolangit menjadi salah satu sekolah pelaksana PIP dan aktif mengusulkan bantuan bagi siswa yang memenuhi kriteria. Permasalahan muncul karena proses klasifikasi penerima masih dilakukan secara manual, tidak hanya memerlukan waktu yang cukup lama tetapi juga berisiko mengalami kesalahan serta kerusakan data. Penelitian ini bertujuan mengembangkan sistem klasifikasi penerima PIP yang lebih efisien dan akurat. Metode yang digunakan adalah algoritma *K-Nearest Neighbor* (KNN) yang membandingkan karakteristik calon penerima dengan data sebelumnya berdasarkan kemiripan. Data penelitian diperoleh melalui angket kepada 250 siswa, mencakup 11 variabel seperti status PKH, status KKS, pendapatan ayah, pendapatan ibu, pekerjaan ayah, pekerjaan ibu, jumlah tanggungan, status keluarga, status tempat tinggal, kondisi rumah, dan status penerimaan PIP. Hasil evaluasi menunjukkan model K-NN memperoleh nilai *accuracy*, *precision*, dan *recall* sebesar 94%, yang membuktikan sistem mampu mengklasifikasikan calon penerima PIP secara efektif, cepat, dan akurat, sehingga mendukung pemerataan bantuan pendidikan di SMP Negeri 2 Sibolangit.

**Kata kunci :** *K-Nearest Neighbor*, beasiswa, machine learning, klasifikasi, website

### 1. PENDAHULUAN

Pendidikan memegang peran penting dalam kemajuan bangsa. Di Indonesia, sebagai negara berkembang, pendidikan masih menghadapi berbagai tantangan, selain sistem yang perlu terus diperbaiki, masih banyak masyarakat yang kesulitan mengakses pendidikan karena kendala biaya. Pemerintah merespons masalah ini melalui berbagai program bantuan pendidikan, salah satunya Program Indonesia Pintar (PIP) yang diatur dalam Instruksi Presiden Nomor 7 Tahun 2014. Program ini dijalankan Kementerian Pendidikan dan Kebudayaan melalui distribusi Kartu Indonesia Pintar (KIP) untuk membantu peserta didik dari keluarga kurang mampu agar dapat mengakses pendidikan yang layak [1]. Untuk mencapai tujuan tersebut, diperlukan peran aktif setiap sekolah dalam mengimplementasikan kebijakan PIP melalui KIP secara efektif, agar program ini dapat tepat sasaran dan sesuai dengan harapan yang telah ditetapkan [2].

Berdasarkan wawancara dengan salah satu guru di SMP Negeri 2 Sibolangit, 201 dari total 339 siswa menerima PIP pada tahun ajaran 2023/2024, sementara lainnya masih mengajukan. Namun demikian, pelaksanaan PIP di SMP Negeri 2 Sibolangit belum sepenuhnya mampu menjangkau seluruh siswa yang berhak menerima manfaatnya. Salah satu kendala utama yang dihadapi adalah belum tersedianya sistem klasifikasi yang memadai, sehingga menyulitkan petugas dalam mengidentifikasi dan menetapkan calon penerima secara tepat. Proses

verifikasi masih manual, memakan waktu yang relatif lama, rawan ketidaktepatan, dan berisiko kehilangan dokumen.

Meski PIP bertujuan meningkatkan pemerataan pendidikan, di lapangan masih menghadapi sejumlah permasalahan, khususnya terkait ketepatan sasaran penerima bantuan [2].

Salah satu isu yang sering muncul adalah adanya penerima beasiswa yang berasal dari keluarga yang secara ekonomi tergolong mampu. Sebaliknya, masih terdapat siswa dari keluarga kurang mampu yang justru tidak memperoleh bantuan tersebut [3].

Penelitian ini bertujuan membangun sistem klasifikasi yang efisien dan akurat guna mengoptimalkan distribusi bantuan PIP kepada siswa yang benar-benar membutuhkan. Peneliti mengusulkan pemanfaatan algoritma *K-Nearest Neighbor* (KNN) untuk membantu sekolah mengklasifikasikan calon penerima secara cepat dan efektif. Sistem yang dirancang akan mengelola data secara terpusat berbasis *website*, sehingga memudahkan akses, pengelolaan, dan monitoring. Variabel yang digunakan mencakup status penerima program keluarga harapan (PKH), kepemilikan kartu keluarga sejahtera (KKS), pendapatan ayah, pendapatan ibu, pekerjaan ayah, pekerjaan ibu, jumlah tanggungan, status keluarga, serta kondisi tempat tinggal. Seluruh variabel ini dipertimbangkan untuk mendukung penentuan penerima PIP secara tepat di SMP Negeri 2 Sibolangit.

## 2. TINJAUAN PUSTAKA

### 2.1. Penelitian Terdahulu

Berdasarkan penelitian yang sudah dilakukan oleh Cholil dengan judul “Implementasi Algoritma Klasifikasi *K-Nearest Neighbor* (KNN) Untuk Klasifikasi Seleksi Penerima Beasiswa”, diperoleh akurasi sebesar 90,5% dan presisi sebesar 89% [4].

Penelitian berikutnya oleh Firdaus & Tukiya dengan judul “Analisis Algoritma C 4.5, *Naive Bayes* dan *K-Nearest Neighbor* Untuk Menentukan Penerima Beasiswa”, dimana algoritma KNN mendapat nilai akurasi yang paling tinggi yaitu 88,24% sedangkan algoritma C4.5 dan *Naive Bayes* masing-masing mendapatkan akurasi sebesar 82,35% [5].

Penelitian selanjutnya oleh Kurniadi dengan judul “Aplikasi Sistem Prediksi Mahasiswa Penerima Beasiswa Berbasis Web dengan Menerapkan Model Klasifikasi *K-Nearest Neighbors*”, dimana pada penelitian ini mendapatkan tingkat akurasi sebesar 92,86% [6]. Berdasarkan temuan dari sejumlah penelitian terdahulu, terlihat bahwa algoritma *K-Nearest Neighbor* (KNN) terbukti efektif dalam menghasilkan klasifikasi calon penerima beasiswa dengan tingkat akurasi yang tinggi.

### 2.2. Program Indonesia Pintar (PIP)

Program Indonesia Pintar (PIP) merupakan salah satu kebijakan pemerintah yang bertujuan untuk mewujudkan pemerataan akses pendidikan di seluruh wilayah Indonesia. Pelaksanaan PIP dilakukan melalui distribusi Kartu Indonesia Pintar (KIP) di bawah pengelolaan Kementerian Pendidikan dan Kebudayaan. Program ini menyediakan bantuan berupa uang tunai, serta memperluas akses dan kesempatan belajar bagi peserta didik yang berasal dari keluarga miskin atau rentan miskin, guna mendukung pembiayaan pendidikan mereka [7].

### 2.3. Klasifikasi

Klasifikasi merupakan salah satu tugas utama dalam bidang *Machine Learning*, yang merujuk pada proses identifikasi kesamaan karakteristik dalam sekumpulan objek pada basis data, untuk kemudian mengelompokkannya ke dalam kategori atau kelas yang berbeda sesuai dengan model klasifikasi yang telah ditetapkan. Tujuan dari proses ini adalah membangun sebuah model dari data *training* yang mampu membedakan atribut-atribut ke dalam kelas-kelas tertentu. Model tersebut selanjutnya digunakan untuk mengklasifikasikan data baru yang belum diketahui kelasnya [8].

### 2.4. K-Nearest Neighbor (KNN)

Algoritma KNN merupakan salah satu metode klasifikasi dalam *supervised learning* yang digunakan untuk mengklasifikasikan suatu objek berdasarkan kedekatannya terhadap data latih (*training data*) serta kelas yang telah ditentukan. Prinsip dasar dari algoritma ini adalah mengidentifikasi sejumlah  $k$  tetangga terdekat dari data yang ingin diklasifikasikan,

dengan menggunakan perhitungan jarak sebagai acuan kedekatan. Secara terminologis, algoritma KNN dapat dijelaskan sesuai dengan namanya, di mana huruf  $K$  merujuk pada jumlah tetangga terdekat (*Nearest Neighbors*) yang dijadikan dasar untuk mengukur kemiripan antara data baru dengan data latih yang ada [9].

### 2.5. K-Fold Cross-Validation

*K-fold cross-validation* adalah metode yang membagi dataset ke dalam beberapa lipatan (*fold*) untuk dijadikan sebagai data pelatihan (*training*) dan data pengujian (*testing*) secara bergantian. Teknik ini populer di kalangan peneliti karena mampu mengurangi bias yang mungkin muncul dalam proses pengambilan sampel. Dalam *k-fold cross-validation*, seluruh data akan memiliki kesempatan yang sama untuk menjadi data pengujian, karena proses pembagian dan pelatihan dilakukan secara berulang [10].

Penerapan *k-fold cross validation* memberikan manfaat signifikan dalam meminimalkan risiko *overfitting* maupun *underfitting*, serta meningkatkan kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya.

### 2.6. Website

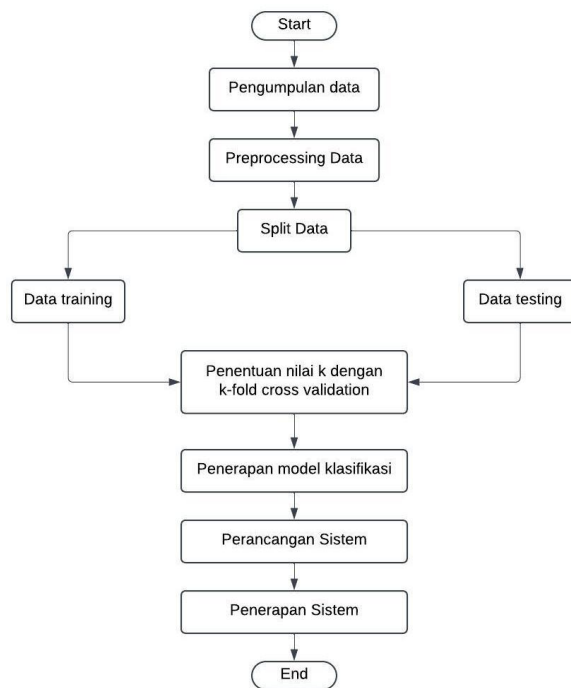
*Website* dapat diartikan sebagai sekumpulan halaman digital yang menyajikan berbagai informasi dalam format teks, gambar, animasi, audio, video, maupun kombinasi dari semuanya. Informasi tersebut disebarluaskan melalui jaringan internet, sehingga dapat diakses oleh pengguna dari berbagai belahan dunia. Secara teknis, *website* terdiri atas barisan kode instruksional yang diterjemahkan oleh *browser* menjadi tampilan visual yang dapat digunakan oleh pengguna. Berdasarkan karakteristiknya, *website* dapat dibagi menjadi dua, yaitu: *website* statis yang merupakan jenis situs *website* yang kontennya bersifat tetap dan tidak dapat diubah secara langsung oleh pengguna melalui *browser*, dan *website* dinamis yang merupakan jenis *website* yang memungkinkan interaksi yang lebih kompleks antara pengguna dan server.

## 3. METODE PENELITIAN

Berikut ini adalah alur penelitian yang digunakan dalam penelitian ini.

### 3.1. Pengumpulan Data

Tahap awal dari penelitian ini melibatkan proses pengumpulan data yang dilaksanakan melalui metode wawancara dan penyebaran kuesioner (*survei*). Jenis data yang dikumpulkan merupakan data primer yang diperoleh langsung dari populasi siswa SMP Negeri 2 Sibolangit. Data tersebut digunakan sebagai dasar dalam proses klasifikasi untuk menentukan calon penerima Program Indonesia Pintar (PIP).



Gambar 1. Flowchart alur penelitian

### 3.2. Preprocessing Data

Proses *preprocessing* pada data dimulai dengan pengidentifikasian atribut, kemudian mengubah jenis atribut kategorial menjadi numerik menggunakan teknik pengkodean *label encoding*, penyeimbangan data melalui teknik *undersampling*, serta normalisasi data dengan pendekatan *min-max normalization*.

### 3.3. Penentuan nilai k

Langkah selanjutnya adalah menentukan nilai parameter  $k$  yang paling optimal untuk algoritma *K-Nearest Neighbor* (KNN) melalui proses *k-fold cross validation*. Teknik ini bertujuan untuk mengidentifikasi parameter yang menghasilkan performa klasifikasi terbaik secara konsisten.

### 3.4. Penerapan Model Klasifikasi

Setelah proses pengumpulan data selesai, variabel-variabel yang digunakan dalam proses klasifikasi penerima bantuan PIP telah ditentukan. Selanjutnya, diterapkan model klasifikasi berdasarkan variabel-variabel yang telah ditentukan tersebut. Proses ini bertujuan untuk melakukan pengklasifikasian terhadap data yang belum diketahui statusnya, sehingga dapat mengidentifikasi siswa/i yang berpotensi menjadi penerima bantuan PIP.

### 3.5. Perancangan Sistem

Pada proses ini, dilakukan perancangan dan desain sistem yang menggambarkan bagian tampilan dan fungsi dari situs *website* pengklasifikasian calon penerima beasiswa PIP yang diperlukan berdasarkan hasil analisis kebutuhan. Proses ini melibatkan gambaran *interface* sistem yang mudah dipahami dan

efisien, serta mengidentifikasi fitur-fitur yang relevan untuk memenuhi kebutuhan pengguna dalam melakukan pengklasifikasian terhadap data siswa.

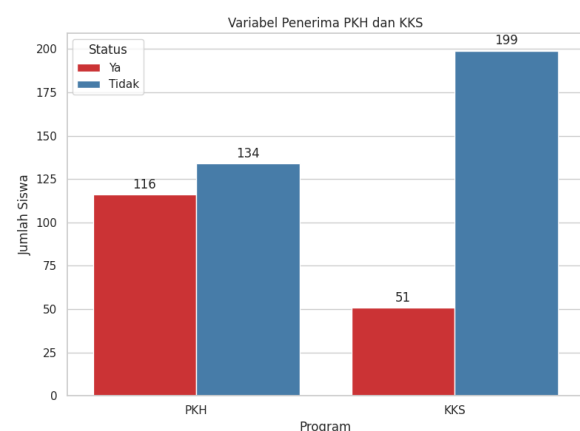
### 3.6. Penerapan Sistem

Langkah terakhir adalah implementasi sistem klasifikasi berbasis *website* yang telah dirancang sebelumnya. Sistem ini digunakan untuk mengelola dan memproses data siswa secara otomatis, serta menentukan siswa yang layak menerima beasiswa Program Indonesia Pintar (PIP) berdasarkan model klasifikasi yang telah dibangun.

## 4. HASIL DAN PEMBAHASAN

### 4.1. Deskripsi Data

Pengumpulan data pada penelitian ini dilakukan dengan menggunakan metode survei, yakni melalui penyebaran angket kepada peserta didik. Berdasarkan hasil survei yang diperoleh, terkumpul sebanyak 250 data siswa dari SMP Negeri 2 Sibolangit. *Dataset* yang diperoleh mencakup 11 variabel, yaitu: Status Penerima PKH, Status Penerima KKS, Pendapatan Ayah, Pendapatan Ibu, Pekerjaan Ayah, Pekerjaan Ibu, Jumlah Tanggungan, Status Keluarga, Status Tempat Tinggal, Kondisi Rumah, serta Status Penerimaan PIP. Berikut adalah uraian mengenai masing-masing variabel yang digunakan dalam proses klasifikasi pada penelitian ini.

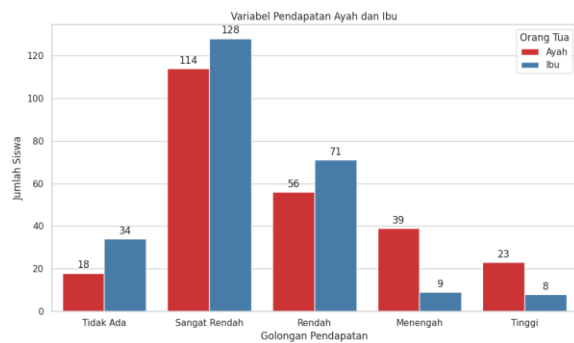


Gambar 2. Diagram variabel Penerima PKH dan KKS

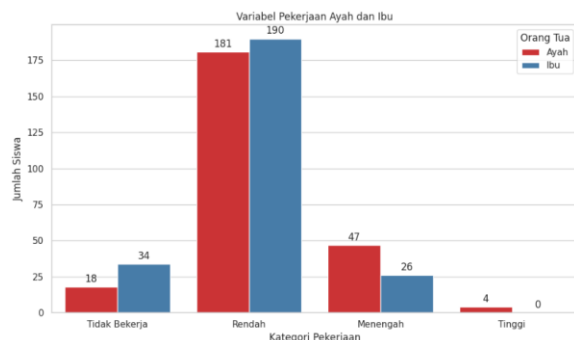
Berdasarkan Gambar 2, terlihat sebaran data pada variabel Penerima PKH dan Penerima KKS. Untuk variabel Penerima PKH, distribusi antara penerima dan non-penerima tergolong seimbang, dengan 116 siswa tercatat sebagai penerima dan 134 siswa sebagai non-penerima. Sementara itu, pada variabel Penerima KKS, terdapat ketimpangan yang cukup signifikan. Hanya 51 siswa yang menerima bantuan KKS, sedangkan 199 siswa lainnya tidak menerima bantuan tersebut.

Berdasarkan Gambar 3, dapat diamati bahwa distribusi data pada variabel Pendapatan Ayah dan Pendapatan Ibu menunjukkan kecenderungan mayoritas siswa berasal dari keluarga dengan tingkat

penghasilan yang sangat rendah, yaitu dengan pendapatan kurang dari Rp. 1.500.000.

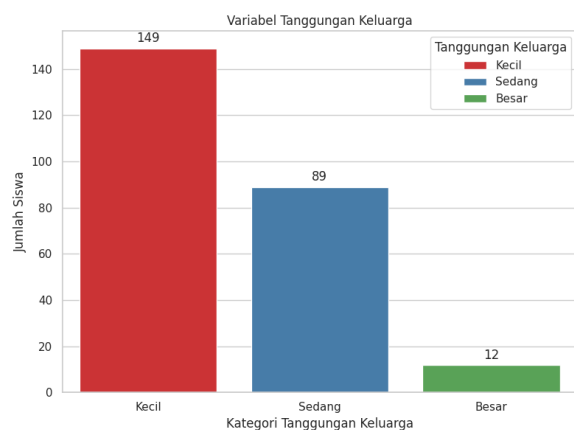


Gambar 3. Diagram variabel pendapatan ayah dan ibu



Gambar 4. Diagram variabel ayah dan ibu

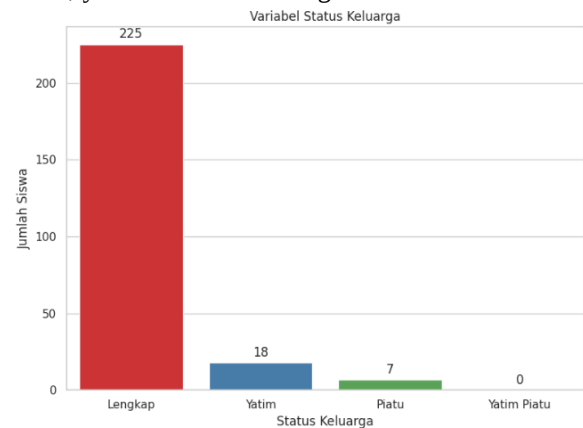
Berdasarkan Gambar 4, dapat terlihat bahwa distribusi data pada variabel Pekerjaan Ayah dan Pekerjaan Ibu menunjukkan dominasi pada kategori pekerjaan dengan tingkat rendah. Hal ini terlihat dari data, di mana terdapat 181 siswa yang ayahnya bekerja di sektor pekerjaan kategori rendah, dan 190 siswa yang ibunya juga bekerja di sektor serupa.



Gambar 5. Diagram variabel tanggungan keluarga

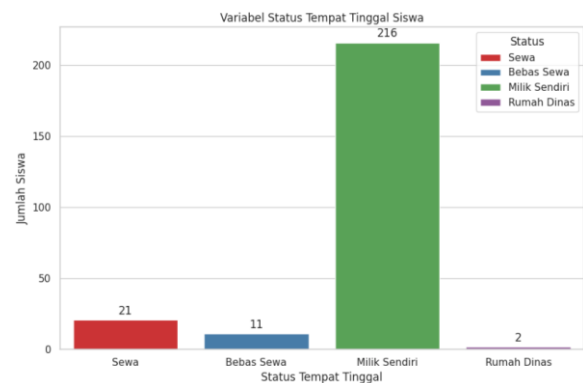
Berdasarkan Gambar 5, distribusi data pada variabel Tanggungan Keluarga menunjukkan bahwa sebagian besar siswa berasal dari keluarga dengan jumlah tanggungan yang tergolong kecil, yakni antara 1 hingga 3 orang, sebanyak 149 siswa. Selanjutnya, terdapat 89 siswa yang berasal dari keluarga dengan tanggungan sedang (4–6 orang), dan sisanya, sebanyak

12 siswa, berasal dari keluarga dengan tanggungan besar, yaitu lebih dari 6 orang.



Gambar 6. Diagram variabel status keluarga

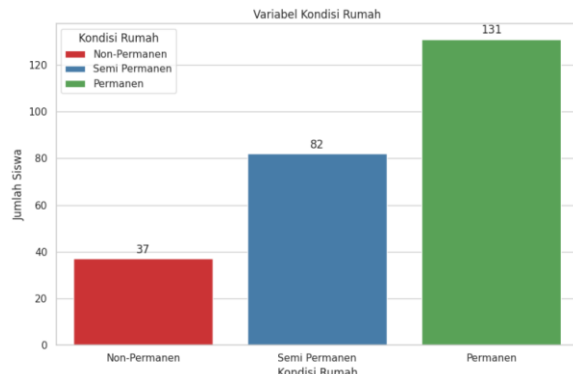
Berdasarkan Gambar 6, terlihat sebaran data pada variabel Status Keluarga. Sebagian besar siswa, yaitu sebanyak 225 orang, berasal dari keluarga yang lengkap, di mana kedua orang tua (ayah dan ibu) masih hidup. Sementara itu, terdapat 25 siswa yang berasal dari keluarga yang telah kehilangan salah satu orang tua (yatim atau piatu). Dan Tidak terdapat siswa yang berasal dari keluarga yatim piatu atau yang telah kehilangan kedua orang tuanya.



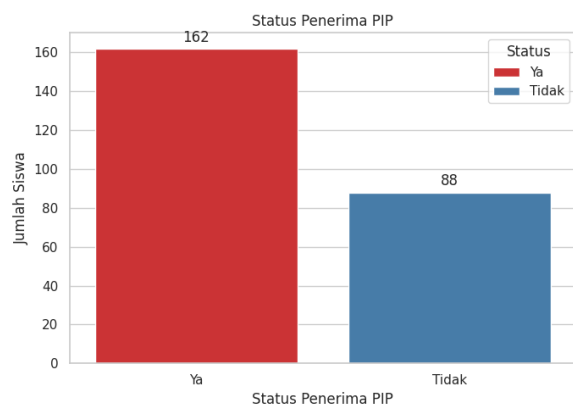
Gambar 7. Diagram variabel status tempat tinggal

Berdasarkan Gambar 7, dapat dilihat sebaran data pada variabel Status Tempat Tinggal. Mayoritas siswa, yakni sebanyak 216 orang, tinggal di rumah milik sendiri. Selanjutnya, terdapat 21 siswa yang tinggal di rumah sewa, dan 11 siswa tinggal secara bebas sewa, yaitu menumpang di rumah keluarga atau kerabat. Sementara itu, hanya 2 siswa yang menempati rumah dinas.

Berdasarkan Gambar 8, terlihat sebaran data pada variabel Kondisi Rumah. Sebagian besar siswa tinggal di rumah dengan kondisi Permanen dan Semi Permanen. Tercatat sebanyak 131 siswa tinggal di rumah Permanen, disusul oleh 82 siswa yang menempati rumah Semi Permanen. Sementara itu, hanya 37 siswa yang tinggal di rumah dengan kondisi Non-Permanen.



Gambar 8. Diagram variabel kondisi rumah



Gambar 9. Diagram status penerima PIP

Berdasarkan Gambar 9, terlihat sebaran data pada variabel Penerima PIP. Dari data tersebut, sebanyak 162 siswa tercatat sebagai penerima bantuan Program Indonesia Pintar (PIP), sementara 88 siswa lainnya tidak menerima bantuan tersebut.

#### 4.2. Preprocessing Data

Setelah data diperoleh, tahap selanjutnya adalah melakukan *preprocessing data*. Dalam penelitian ini, *preprocessing data* mencakup tiga langkah utama, yaitu: transformasi data menggunakan metode *label encoding*, penyeimbangan data melalui teknik *undersampling*, serta normalisasi data dengan pendekatan *min-max normalization*.

Dalam penelitian ini, *Label Encoding* diterapkan untuk mengubah data kategorial dalam dataset ke dalam nilai numerik berdasarkan indeks atau urutan unik dari masing-masing kategori. Berikut adalah variabel pada *dataset* yang akan ditransformasikan dari data kategorial ke dalam format numerik.

Tabel 1. *Label encoding* pada variabel status penerima PKH

| Status Penerima PKH | Label |
|---------------------|-------|
| Ya                  | 0     |
| Tidak               | 1     |

Pada Tabel 1, menampilkan hasil *label encoding* untuk variabel penerima PKH yang memiliki 2 kategori, yaitu Ya dan Tidak.

Tabel 2. *Label encoding* pada variabel penerima KKS

| Status Penerima KKS | Label |
|---------------------|-------|
| Ya                  | 0     |
| Tidak               | 1     |

Pada Tabel 2, menampilkan hasil *label encoding* untuk variabel penerima KKS yang memiliki 2 kategori, yaitu Ya dan Tidak.

Tabel 3. *Label encoding* pada variabel pendapatan orangtua

| Pendapatan Orangtua               | Label |
|-----------------------------------|-------|
| Tidak Ada                         | 0     |
| Golongan Pendapatan Sangat Rendah | 1     |
| Golongan Pendapatan Rendah        | 2     |
| Golongan Pendapatan Menengah      | 3     |
| Golongan Pendapatan Tinggi        | 4     |

Pada Tabel 3, menampilkan hasil *label encoding* untuk variabel pendapatan orang tua, yang terdiri dari dua variabel, yaitu Pendapatan Ayah dan Pendapatan Ibu. Kedua variabel ini memiliki 5 kategori yang sama, yaitu Tidak Ada, Golongan Pendapatan Sangat Rendah, Golongan Pendapatan Rendah, Golongan Pendapatan Menengah, dan Golongan Pendapatan Tinggi.

Tabel 4. *Label encoding* pada variabel pekerjaan orangtua

| Pekerjaan Orangtua          | Label |
|-----------------------------|-------|
| Tidak Bekerja               | 0     |
| Golongan Pekerjaan Rendah   | 1     |
| Golongan Pekerjaan Menengah | 2     |
| Golongan Pekerjaan Tinggi   | 3     |

Pada Tabel 4, menampilkan hasil *label encoding* untuk variabel pekerjaan orang tua, yang terdiri dari dua variabel, yaitu pekerjaan ayah dan pekerjaan ibu. Kedua variabel ini memiliki 4 kategori yang sama, yaitu Tidak Bekerja, Golongan Pekerjaan Rendah, Golongan Pekerjaan Menengah, dan Golongan Pekerjaan Tinggi.

Tabel 5. *Label encoding* pada variabel tanggungan keluarga

| Tanggungan Keluarga | Label |
|---------------------|-------|
| Tanggungan Besar    | 0     |
| Tanggungan Sedang   | 1     |
| Tanggungan Kecil    | 2     |

Pada Tabel 5, menampilkan hasil *label encoding* untuk variabel tanggungan keluarga yang memiliki 3 kategori, yaitu Tanggungan Kecil, Tanggungan Sedang dan Tanggungan Besar.

Tabel 6. *Label encoding* pada variabel status keluarga

| Status Keluarga | Label |
|-----------------|-------|
| Yatim Piatu     | 0     |
| Yatim           | 1     |
| Piatu           | 2     |
| Lengkap         | 3     |

Pada Tabel 6, menampilkan hasil *label encoding* untuk variabel status keluarga yang memiliki 3 kategori, yaitu Yatim Piatu, Yatim, Piatu dan Lengkap.

Tabel 7. *Label encoding* pada variabel status tempat tinggal

| Status Tempat Tinggal | Label |
|-----------------------|-------|
| Sewa                  | 0     |
| Bebas Sewa            | 1     |
| Milik Sendiri         | 2     |
| Rumah Dinas           | 3     |

Pada Tabel 7, menampilkan hasil *label encoding* untuk variabel tempat tinggal yang memiliki 4 kategori, yaitu Sewa, Bebas Sewa, Milik Sendiri dan Rumah Dinas.

Tabel 8. *Label encoding* pada variabel kondisi rumah

| Kondisi Rumah | Label |
|---------------|-------|
| Non-Permanen  | 0     |
| Semi-Permanen | 1     |
| Permanen      | 2     |

Pada Tabel 8, menampilkan hasil *label encoding* untuk variabel kondisi rumah yang memiliki 3 kategori, yaitu Non-Permanen, Semi-Permanen dan Permanen.

Tabel 9. *Label encoding* pada variabel penerima PIP

| Penerima PIP | Label |
|--------------|-------|
| Ya           | 0     |
| Tidak        | 1     |

Pada Tabel 9, menampilkan hasil *label encoding* untuk variabel penerima PIP yang memiliki 2 kategori, yaitu Ya dan Tidak. Setelah proses transformasi data, langkah selanjutnya adalah melakukan penyeimbangan data (*data balancing*) pada *dataset* yang telah ditransformasi. Tercatat terdapat 162 siswa yang berlabel sebagai penerima PIP, sementara hanya terdapat 88 siswa yang berlabel tidak menerima PIP. Untuk mengatasi permasalahan tersebut, digunakan teknik *undersampling*, yaitu menghapus secara acak sebagian data dari kelas penerima PIP hingga jumlahnya seimbang dengan kelas tidak penerima PIP. Hasil dari proses ini adalah terbentuknya *dataset* yang seimbang, dengan masing-masing kelas terdiri atas 88 siswa, sehingga total data yang digunakan untuk pelatihan model adalah sebanyak 176 data.

Setelah proses penyeimbangan data selesai dilakukan, tahap selanjutnya adalah melakukan normalisasi terhadap *dataset*. Dalam penelitian ini, teknik normalisasi yang digunakan adalah metode *min-max normalization*.

Berikut adalah variabel-variabel yang dilakukan normalisasi dalam penelitian. Variabel pendapatan orangtua dapat dibagi menjadi 2 yaitu pendapatan ayah dan pendapatan ibu dimana diketahui kedua variabel memiliki nilai  $x_{min}$  adalah 0 dan nilai  $x_{max}$  adalah 4, sehingga diperoleh hasil normalisasi sebagai berikut.

Tabel 10. Hasil normalisasi variabel pendapatan orangtua

| Pendapatan Orangtua | Hasil Normalisasi |
|---------------------|-------------------|
| 0                   | 0                 |
| 1                   | 0.25              |
| 2                   | 0.5               |
| 3                   | 0.75              |
| 4                   | 1                 |

Variabel pekerjaan orangtua dapat dibagi menjadi 2 yaitu pekerjaan ayah dan pekerjaan ibu dimana diketahui kedua variabel memiliki nilai  $x_{min}$  adalah 0 dan nilai  $x_{max}$  adalah 3, sehingga diperoleh hasil normalisasi sebagai berikut.

Tabel 11. Hasil normalisasi variabel pekerjaan orangtua

| Pekerjaan Orangtua | Hasil Normalisasi |
|--------------------|-------------------|
| 0                  | 0                 |
| 1                  | 0.3333            |
| 2                  | 0.6667            |
| 3                  | 1                 |

Pada variabel tanggungan keluarga diketahui nilai  $x_{min}$  adalah 0 dan nilai  $x_{max}$  adalah 2, sehingga diperoleh hasil normalisasi sebagai berikut.

Tabel 12. Hasil normalisasi tanggungan keluarga

| Tanggungan Keluarga | Hasil Normalisasi |
|---------------------|-------------------|
| 0                   | 0                 |
| 1                   | 0.5               |
| 2                   | 1                 |

Pada variabel status keluarga diketahui nilai  $x_{min}$  adalah 0 dan nilai  $x_{max}$  adalah 3, sehingga diperoleh hasil normalisasi sebagai berikut.

Tabel 13. Hasil normalisasi variabel status keluarga

| Status Keluarga | Hasil Normalisasi |
|-----------------|-------------------|
| 0               | 0                 |
| 1               | 0.3333            |
| 2               | 0.6667            |
| 3               | 1                 |

Pada variabel status tempat tinggal diketahui nilai  $x_{min}$  adalah 0 dan nilai  $x_{max}$  adalah 3, sehingga diperoleh hasil normalisasi sebagai berikut.

Tabel 14. Hasil normalisasi variabel status tempat tinggal

| Status Tempat Tinggal | Hasil Normalisasi |
|-----------------------|-------------------|
| 0                     | 0                 |
| 1                     | 0.3333            |
| 2                     | 0.6667            |
| 3                     | 1                 |

Pada variabel kondisi rumah diketahui nilai  $x_{min}$  adalah 0 dan nilai  $x_{max}$  adalah 2, sehingga diperoleh hasil normalisasi sebagai berikut.

Tabel 15. Hasil normalisasi variabel kondisi rumah

| Kondisi Rumah | Hasil Normalisasi |
|---------------|-------------------|
| 0             | 0                 |
| 1             | 0.5               |
| 2             | 1                 |

#### 4.3. Penentuan Nilai k

Tahapan selanjutnya adalah menentukan nilai  $k$  yang paling optimal pada algoritma *K-Nearest Neighbor* (KNN), untuk mendukung proses ini, digunakan teknik *k-fold cross-validation*. Dalam penelitian ini, digunakan metode *5-fold cross-validation* dimana dengan *5-fold*, proses pelatihan dan validasi dilakukan sebanyak lima kali, di mana pada setiap iterasi, empat *fold* digunakan untuk melatih model dan satu *fold* digunakan sebagai data validasi.

Dalam penelitian ini, nilai  $k$  yang diuji untuk algoritma *K-Nearest Neighbors* (KNN) adalah 2, 3, 4, 5, 6, 7, dan 8. Dengan menerapkan *5-fold cross-validation* pada masing-masing nilai tersebut, diperoleh hasil evaluasi yang digunakan untuk menentukan nilai  $k$  optimal yang menghasilkan performa klasifikasi terbaik. Hasil perhitungan penentuan nilai  $k$  untuk algoritma KNN dengan metode *5-fold cross-validation* disajikan sebagai berikut.

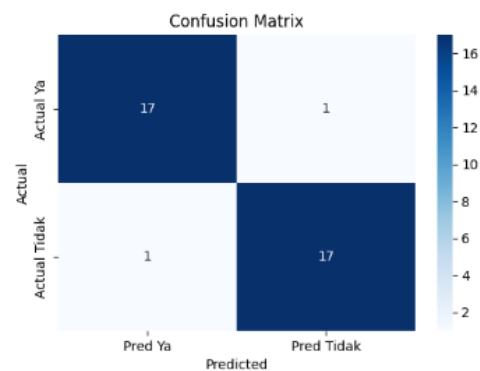
Tabel 16. Hasil perhitungan penentuan nilai k terbaik

| Nilai k | Cross Validation Accuracies |        |        |        |        | Mean Accuracies |
|---------|-----------------------------|--------|--------|--------|--------|-----------------|
|         | Fold 1                      | Fold 2 | Fold 3 | Fold 4 | Fold 5 |                 |
| 2       | 0.9444                      | 0.8857 | 0.9429 | 0.8000 | 1.000  | 0.9146          |
| 3       | 0.9167                      | 0.8857 | 0.9714 | 0.8286 | 0.9429 | 0.9090          |
| 4       | 0.8889                      | 0.8857 | 0.9429 | 0.8000 | 0.9429 | 0.8921          |
| 5       | 0.9167                      | 0.8857 | 0.9714 | 0.8571 | 0.9429 | 0.9148          |
| 6       | 0.9722                      | 0.8286 | 0.9714 | 0.8571 | 0.9143 | 0.9087          |
| 7       | 0.9167                      | 0.8857 | 0.9143 | 0.8857 | 0.9143 | 0.8976          |
| 8       | 0.9167                      | 0.8857 | 0.9714 | 0.8857 | 0.9143 | 0.9090          |

Hasil evaluasi menunjukkan bahwa nilai akurasi tertinggi dicapai ketika  $k$  bernilai 5, dengan rata-rata akurasi hasil *cross-validation* (*mean cross-validation accuracy*) sebesar 0,9148. Oleh karena itu, dapat disimpulkan bahwa nilai  $k$  yang paling optimal untuk digunakan dalam algoritma KNN pada penelitian ini adalah  $k = 5$ , karena memberikan performa klasifikasi terbaik dibandingkan nilai  $k$  lainnya yang diuji.

#### 4.4. Penerapan Model Klasifikasi

Setelah nilai parameter  $k$  untuk algoritma *K-Nearest Neighbor* (KNN) berhasil ditentukan, tahap selanjutnya adalah implementasi model klasifikasi menggunakan algoritma KNN. Proses implementasi dilakukan dengan memanfaatkan bahasa pemrograman Python melalui perangkat lunak Visual Studio Code sebagai lingkungan pengembangan. Untuk mengevaluasi kinerja model klasifikasi yang dibangun, digunakan *confusion matrix* sebagai alat ukur. Berikut disajikan *confusion matrix* yang dihasilkan dari penerapan model klasifikasi menggunakan Python pada platform Visual Studio Code.

Gambar 10. Hasil *confusion matrix*

Berdasarkan hasil *confusion matrix* pada Gambar 10, diperoleh hasil sebagai berikut:

- True Positive* (TP) = 17, yang berarti terdapat 17 siswa yang merupakan penerima PIP berhasil diklasifikasikan dengan benar sebagai penerima PIP.
- False Positive* (FP) = 1, yang berarti terdapat 1 data siswa yang sebenarnya bukan penerima PIP namun salah diklasifikasikan sebagai penerima PIP.
- False Negative* (FN) = 1, yang berarti terdapat 1 data siswa penerima PIP salah diklasifikasikan sebagai bukan penerima.
- True Negative* (TN) = 17, yang berarti terdapat 17 siswa yang bukan penerima PIP berhasil diklasifikasikan dengan benar.

Berdasarkan hasil evaluasi *confusion matrix* tersebut, maka dapat dihitung hasil evaluasi akurasi, *precision* dan *recall* pada model. Berdasarkan evaluasi *confusion matrix*, maka diperoleh nilai akurasi sebesar 94,4%, nilai *precision* sebesar 94,4%, dan nilai *recall* sebesar 94,4%. Setelah memperoleh ketiga metrik evaluasi dari *confusion matrix*, tahap selanjutnya adalah menghasilkan *classification report* untuk memperoleh gambaran yang lebih komprehensif mengenai kinerja model. Berikut merupakan hasil *classification report* dari penerapan model klasifikasi menggunakan algoritma *K-Nearest Neighbors* (KNN) dengan nilai  $k=5$ .

| Classification Report |           |        |          |         |
|-----------------------|-----------|--------|----------|---------|
|                       | precision | recall | f1-score | support |
| Penerima              | 0.94      | 0.94   | 0.94     | 18      |
| Bukan Penerima        | 0.94      | 0.94   | 0.94     | 18      |
| accuracy              |           |        | 0.94     | 36      |
| macro avg             | 0.94      | 0.94   | 0.94     | 36      |
| weighted avg          | 0.94      | 0.94   | 0.94     | 36      |

Gambar 11. Hasil *classification report*

Berdasarkan Gambar 11, diperoleh hasil *classification report* penerapan model klasifikasi algoritma KNN dengan akurasi sebesar 94% dan nilai *precision* sebesar 94% serta nilai *recall* sebesar 94%.

a. *Landing Page*, merupakan halaman awal ketika pengguna mengakses sistem klasifikasi.



- 

c. Halaman *Dashboard*, merupakan tampilan utama yang muncul setelah pengguna berhasil melakukan proses login ke dalam sistem.



- | Dataset Siswa |                        |           |               |              |              |               |               |              |              |                   |               |               |              |
|---------------|------------------------|-----------|---------------|--------------|--------------|---------------|---------------|--------------|--------------|-------------------|---------------|---------------|--------------|
| No            | Nama                   | NISN      | Jenis Kelamin | Pemerintahan | Pemerintahan | Pemerintahan  | Pemerintahan  | Pemerintahan | Pemerintahan | Pemerintahan      | Pemerintahan  | Pemerintahan  | Pemerintahan |
|               |                        |           |               |              |              |               |               |              |              |                   |               |               |              |
| 1             | Yusuf Adhika Nur Tugan | 190601028 | Pemempuan     | Tidak        | Ya           | Menengah      | Sangat Rendah | Rendah       | Rendah       | Tinggangan Rendah | Sangat Rendah | Sangat Rendah | Ya           |
| 2             | Yusuf Adhika Nur Tugan | 190601028 | Pemempuan     | Ya           | Ya           | Sangat Rendah | Sangat Rendah | Rendah       | Rendah       | Tinggangan Rendah | Sangat Rendah | Rendah        | Ya           |
| 3             | Yusuf Adhika Nur Tugan | 190601028 | Pemempuan     | Tidak        | Ya           | Rendah        | Sangat Rendah | Rendah       | Rendah       | Tinggangan Rendah | Sangat Rendah | Sangat Rendah | Ya           |
| 4             | Siti Azzah             | 190601028 | Pemempuan     | Tidak        | Ya           | Menengah      | Tidak Ada     | Rendah       | Tidak Rendah | Tinggangan Rendah | Pada          | Sangat Rendah | Ya           |
| 5             | Alwinda Nur Salsabila  | 191407108 | Pemempuan     | Tidak        | Tidak        | Rendah        | Rendah        | Rendah       | Rendah       | Tinggangan Rendah | Sangat Rendah | Rendah        | Ya           |
| 6             | Alvinus Salsabila      | 190601028 | Pemempuan     | Tidak        | Tidak        | Menengah      | Rendah        | Menengah     | Menengah     | Tinggangan Rendah | Sangat Rendah | Sangat Rendah | Tidak        |

[illegible]

f. Halaman *Form* Klasifikasi, merupakan halaman yang *digunakan* untuk proses penginputan data siswa yang ini dilakukan klasifikasi



Berikut adalah hasil pengujian sistem berbasis website pada penelitian ini.

| No | Aktivitas Pengujian             | Hasil yang Diharapkan                                                                                             | Keterangan                                                                    |
|----|---------------------------------|-------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------|
| 1  | <i>Landing Page</i>             | Menampilkan halaman awal sistem dan memberikan penjelasan singkat mengenai sistem                                 | <input checked="" type="checkbox"/> Valid<br><input type="checkbox"/> Invalid |
| 2  | Halaman <i>Login</i>            | Menampilkan form <i>login</i> dan mengarahkan ke halaman <i>dashboard</i> jika <i>email</i> dan kata sandi valid. | <input checked="" type="checkbox"/> Valid<br><input type="checkbox"/> Invalid |
| 5  | Halaman <i>Dashboard</i>        | Menampilkan jumlah siswa yang telah diklasifikasikan, beserta jumlah yang Layak dan Tidak Layak                   | <input checked="" type="checkbox"/> Valid<br><input type="checkbox"/> Invalid |
| 6  | Halaman <i>Dataset</i>          | Menampilkan <i>dataset</i> yang digunakan untuk melatih model klasifikasi                                         | <input checked="" type="checkbox"/> Valid<br><input type="checkbox"/> Invalid |
| 7  | Halaman Tabel Klasifikasi       | Menampilkan tabel yang berisi data siswa yang telah berhasil diklasifikasikan                                     | <input checked="" type="checkbox"/> Valid<br><input type="checkbox"/> Invalid |
| 8  | Halaman <i>Form</i> Klasifikasi | Menampilkan form yang digunakan untuk menginput data siswa yang digunakan dalam klasifikasi                       | <input checked="" type="checkbox"/> Valid<br><input type="checkbox"/> Invalid |

## 5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian, sistem klasifikasi untuk menentukan calon penerima beasiswa Program Indonesia Pintar (PIP) berhasil dikembangkan menggunakan algoritma *K-Nearest Neighbor* (KNN) dalam bentuk aplikasi berbasis *website* yang memudahkan pihak sekolah dalam mengelola data siswa serta mempercepat, menstrukturkan, dan mengefisienkan proses penilaian. Evaluasi model dengan menggunakan pembagian data training dan testing sebesar 80:20 serta menggunakan nilai  $k=5$ , menunjukkan bahwa model menghasilkan akurasi, precision dan recall sebesar 94%. Hasil ini menunjukkan bahwa model yang dikembangkan memiliki performa klasifikasi yang baik dan efektif dalam mengklasifikasikan calon penerima beasiswa PIP. Sistem telah diuji menggunakan 176 data, dengan hasil klasifikasi yang menunjukkan bahwa 169 data (96%) diklasifikasikan dengan benar, sementara 7 data (4%) mengalami kesalahan klasifikasi. Hasil ini menunjukkan bahwa sistem memiliki tingkat ketepatan klasifikasi yang cukup tinggi. Untuk pengembangan selanjutnya, disarankan penerapan algoritma lain seperti *Decision Tree* atau *Naive Bayes* serta perbandingan performa antar algoritma untuk memperoleh hasil yang lebih optimal, serta memperluas penerapan sistem ini pada program beasiswa lainnya agar memberikan kontribusi yang lebih luas dan menjangkau lebih banyak penerima manfaat.

## DAFTAR PUSTAKA

- [1] P. M. Rakista, "Implementasi Kebijakan Program Indonesia Pintar (PIP) (Studi Kasus Pada Sekolah Dasar Di Kabupaten Banyumas)," *SAWALA*, vol. 8, no. 2, pp. 224–232, 2020.
- [2] B. G. Dimmera and P. D. Purnasari, "Permasalahan Dan Solusi Program Indonesia Pintar Dalam Mewujudkan Pemerataan Pendidikan Di Kabupaten Bengkayang," *SEBATIK*, pp. 307–314, 2020.
- [3] A. Nata and Suparmadi, "Analisis Sistem Pendukung Keputusan Dengan Model Klasifikasi Berbasis Machine Learning Dalam Penentuan Penerima Program Indonesia Pintar," *Journal of Science and Social Research*, vol. 5, no. 3, pp. 697–702. Available: <http://jurnal.goretanpena.com/index.php/JSSR>
- [4] S. R. Cholil, T. Handayani, R. Prathivi, and T. Ardianita, "Implementasi Algoritma Klasifikasi K-Nearest Neighbor (KNN) Untuk Klasifikasi Seleksi Penerima Beasiswa," *IJCIT*, vol. 6, no. 2, pp. 118–127, Jun. 2021.
- [5] M. F. Firdaus and Tukiyyat, "Analisis Algoritma C 4.5, Naive Bayes dan K-Nearest Neighbor Untuk Menentukan Penerima Beasiswa," *Jurnal Penelitian Ilmu Komputer*, vol. 1, no. 3, pp. 156–164. Available: <https://mypublikasi.com/index.php/JUPIK/156>
- [6] D. Kurniadi, F. Nuraeni, and A. F. Hazar, "Aplikasi Sistem Prediksi Mahasiswa Penerima Beasiswa Berbasis Web dengan Menerapkan Model Klasifikasi K-Nearest Neighbors," *Jurnal Algoritma*, vol. 21, no. 1, pp. 68–79, May 2024, doi: 10.33364/algoritma/v.21-1.1424.
- [7] "Data Penyaluran PIP Perwilayah," [pip.dikdasmen.go.id](http://pip.dikdasmen.go.id). Available: [https://pip.dikdasmen.go.id/penyaluran/jenjang\\_sekolah/070103/SMP?tahun=2024](https://pip.dikdasmen.go.id/penyaluran/jenjang_sekolah/070103/SMP?tahun=2024)
- [8] P. P. P. A. N. W. F. I. R. H. Zer, B. H. Hayadi, and A. R. Damanik, "Pendekatan Machine Learning Menggunakan Algoritma C4.5 Berbasis PSO Dalam Analisis Pemahaman Pemrograman Website," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 10, no. 3, pp. 150–156, Aug. 2022, doi: 10.23960/jitet.v10i3.2700.
- [9] I. D. Id, *Machine Learning: Teori, Studi Kasus dan Implementasi Menggunakan Python*, 1st ed. UR PRESS, 2021.
- [10] T. Ridwansyah, "Implementasi Text Mining Terhadap Analisis Sentimen Masyarakat Dunia Di Twitter Terhadap Kota Medan Menggunakan K-Fold Cross Validation Dan Naïve Bayes Classifier," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 2, no. 5, pp. 178–185. Available: <https://djournals.com/klik>