Jurnal Cermatika
LPPM Universitas Graha Nusantara Volume 4, Nomor 1, Oktober 2024
https://jurnalugn.id/index.php/cermatika **p-ISSN : 2502-5139**
DOI https://doi.org/10.64168/cermatika.v4i1.1324 **e-ISSN : 2776-6039**

STUDI KASUS SMA N 1 SINUNUKAN IMPLEMENTASI ALGORITMA K-NEAREST NEIGHBOR UNTUK KLASIFIKASI PENERIMA BEASISWA PROGRAM INDONESIA PINTAR (PIP)

Torang Siregar [1*], Riski Ardian[2], Ahmad Arisman[3], Iskandarsyah[4]

Mahasiswa Pascasarjana Tadris Matematika UIN Syahada Padangsidimpuan, Sumatera Utara, Indonesia

[1*] Penulis Korespodensi :torangsir@uinsyahada.ac.id

Abstrak

Program Indonesia Pintar (PIP) merupakan salah satu kebijakan pemerintah yang diharapkan dapat meningkatkan aksesibilitas dan pemerataan pendidikan di Indonesia, namun dalam pelaksanaannya pemberian beasiswa dari program ini masih dijumpai banyak kasus yang kurang tepat sasaran.Salah satu permasalahannya adalah masih ditemukan siswa penerima bantuan pendidikan yang berasal dari keluarga yang mampu, sedangkan siswa yang kurang mampu justru tidak mendapatkan bantuan. Sehingga diperlukan suatu sistem klasifikasi berbasis web yang dapat mengklasifikasikan siswa layak atau tidak untuk mendapatkan PIP. Penelitian ini mengimplementasikan algoritma *k*-nearest neighbor untuk mengklasifikasikan siswa penerima beasiswa PIP. Penelitian ini menggunakan data siswa/i SMAN 1 Sinunukan, Mandailing Natal yang didapat melalui penyebaran angket sesuai dengan kriteria yang telah ditentukan. Data yang telah diperoleh kemudian dilakukan preprocessing data dengan menggunakan Label Encoder dan Normalisasi Min-Max.Data dibagi menjadi dua jenis yaitu data training dan data testing.*K*-fold cross validation digunakan untuk menentukan nilai *k* yang optimal. Hasil penelitian ini memperlihatkan tingkat akurasi yang dihasilkan berdasarkan hasil pengujian yang dilakukan untuk implentasi algoritma *k*-nearest neighbor dalam klasifikasi kelayakan penerima beasiswa PIP yaitu sebesar 70% dengan nilai *k*=3.

Kata kunci: Program Indonesia Pintar (PIP), *K*-Nearest Neighbor, Klasifikasi

Abstract

The Smart Indonesia Program (PIP) is one of the government policies which is expected to increase the accessibility and equality of education in Indonesia, but in its implementation there are still many cases where scholarships from this program are not well targeted. One of the problems is that there are still students who receive educational assistance who come from well-off families, while less well-off students don't get help. So we need a web-based classification system that can classify students as eligible or not to get PIP. This research implements the k-nearest neighbor algorithm to classify students who receive PIP scholarships. This research uses student data from SMAN 1 Sinunukan, Mandailing Natal which was obtained through distributing questionnaires according to predetermined criteria. The data that has been obtained is then preprocessed using the Label Encoder and Min-Max Normalization. The data is divided into two types, namely training data and testing data. K-fold cross validation is used to determine the optimal k value. The results of this research

show that the level of accuracy produced based on the results of tests carried out for the implementation of the k-nearest neighbor algorithm in the classification of eligibility for PIP scholarship recipients is 70% with a value of k=3.

Keywords : *Smart Indonesia Program (PIP), K-Nearest Neighbor, Classification*

Jurnal Cermatika
LPPM Universitas Graha Nusantara
https://jurnalugn.id/index.php/cermatika
DOI httpsdoi.org10.64168cermatika.v4i1
Volume 4, Nomor 1, Oktober 2024
p-ISSN : 2502-5139
e-ISSN : 2776-6039

1. PENDAHULUAN

Pendidikan ialah aspek yang sangat penting dalam membantu suatu bangsa untuk mencapai kemajuan.Kualitas pendidikan menjadi sangat penting dalam menentukan kemajuan suatu bangsa, khususnya Indonesia yang saat ini masih mengalami berbagai masalah dalam sistem pendidikannya.Selain perlu ditingkatkan, masalah utama adalah kesenjangan sosial-ekonomi yang menyebabkan banyak warga tidak mampu membayar biaya pendidikan. Pemerintah berusaha menyelesaikan masalah ini dengan berbagai cara, salah satunya yaitu melalui program beasiswa (Ridho *et al*., 2021).

Salah satu kebijakan pemerintah dalam upaya pemerataan pendidikan yaitu dengan diterbitkannya Program Indonesia Pintar (PIP) oleh presiden Joko Widodo pada tahun 2014 melalui Instruksi Presiden Nomor 7 tahun 2014. Instruksi tersebut menugaskan Kementerian Pendidikan dan Kebudayaan untuk menyiapkan PIP melalui pemberian Kartu Indonesia Pintar (KIP) untuk membantu siswa miskin agar mereka bisa mendapatkan pendidikan yang layak (Rakista, 2021).

Sesuasi dengan hasil wawancara yang sudah dilaksanakan dengan salah satu guru di SMA Negeri 1 Sinunukan, Mandailing Natal, diperoleh informasi bahwa dalam menjalankan pendidikan di SMA Negeri 1 Sinunukan, Mandailing Natal, masih banyak siswa yang mengalami kendala dalam hal finansial.Saat ini, implementasi PIP di SMA Negeri 1 Sinunukan, Mandailing Natal belum dapat memberikan manfaat secara merata bagi siswa yang berhak menerimanya.Salah satu faktor penyebabnya adalah ketidakadanya sistem yang efektif, yang menyulitkan petugas dalam mengklasifikasikan calon penerima PIP. Proses pengolahan data masih dilaksanakan secara manual, di mana pendekatan ini membutuhkan waktu yang lama sehingga berpotensi menghasilkan bantuan yang tidak tepat sasaran, dan juga membawa risiko seperti kehilangan atau kerusakan dokumen.

Maka, dibutuhkan suatu sistem pengklasifikasian yang dapat menghasilkan data yang akurat dan sesuai sasaran, sehingga bantuan beasiswa PIP dapat dimanfaatkan secara optimal oleh siswa yang membutuhkannya.

Machine learning ialah sebuah subbidang pada ilmu kecerdasan buatan (AI), yang bertujuan untuk memprogram komputer agar dapat memperlihatkan perilaku yang cerdas layaknya manusia serta bisa meningkatkan pemahamannya melalui *trial* dan *error*.Salah satu tugas yang dapat dilakukan oleh *machine learning* yaitu melakukan klasifikasi (Retnoningsih & Pramudita, 2020).

Klasifikasi merupakan suatu teknik dalam machine learning yang mengelompokkan data atau objek ke dalam kategori atau label tertentu berdasarkan ciriciri yang dimilikinya (Nasution *et al*., 2019).

Penelitian ini memakai algoritma *KNearest Neighbor* (KNN) karena memiliki tingkat akurasi yang lebih tinggi dan kemudahan penerapannya dibandingkan dengan algoritma-algoritma lainnya.KNN ialah algoritma yang tidak memerlukan asumsi tentang distribusi data yang digunakan dan tidak memerlukan informasi tentang

Volume 4, Nomor 1, Oktober 2024
p-ISSN : 2502-5139
e-ISSN : 2776-6039

parameter populasi (non-paramaetric) (Suryaman *et al*., 2021).KNN mempunyai keunggulan dalam pelatihan yang cepat dan sederhana sehingga mudah dipelajari dan diimplementasikan.Selain itu, algoritma KNN juga sangat efektif untuk melatih data yang berjumlah banyak (Iriantoro *et al*., 2018).

Akurasi *K-Nearest Neighbor* sangat terpengaruh dengan memilih jumlah *k* tetangga terdekat. Jika nilai *k* yang tinggi bisa mengurangi efek gangguan atau ketidakakuratan yang terdapat dalam data (*noise*) pada klasifikasi tetapi membuat batasan setiap kelas menjadi kabur sedangkan nilai *k* yang terlalu kecil bisa mengakibatkan algoritma terlalu sensitif terhadap *noise* (Cholil *et al*., 2021). Nilai *k* terbaik bisa dipilih dengan optimasi parameter, salah satunya yaitu dengan memakai *K-fold cross validation*. *K-fold cross validation* atau dapat disebut estimasi rotasi ialah sebuah teknik validasi yang dilakukan dengan mengulang proses perhitungan sampai beberapa kali (Tempola *et al*., 2018).

Dari uraian pada latar belakang di atas, telah dilakukan penelitian dengan judul "Implementasi Algoritma *K-Nearest Neighbor* untuk Klasifikasi Penerima Beasiswa Program Indonesia Pintar (PIP)".

2. BAHAN DAN METODE

Penelitian ini ialah penelitian terapan yang bertujuan untuk menerapkan algoritma KNN pada sebuah sistem klasifikasi penentuan penerima beasiswa PIP di SMA Negeri 1 Kota Sinunukan, Mandailing Natal.

Proses pengumpulan data pada penelitian ini dilaksanakan dengan metode wawancara dan survei. Data yang dijadikan sebagai dasar penelitian ini merupakan jenis data primer. Data tersebut diperoleh dari populasi siswa SMA Negeri 1 Kota Sinunukan, Mandailing Natal dan digunakan untuk melakukan klasifikasi terhadap penentuan penerima beasiswa Program Indonesia Pintar (PIP), serta dilakukan pemilihan sampel menggunakan teknik *sampling* sistematis, dengan jumlah anggota sampel sebanyak 100 siswa.

Sampling sistematis ialah teknik pengambilan sampel yang dilakukan dengan memanfaatkan urutan nomor dari anggota populasi yang telah diberi nomor urut.Sebagai contoh, dalam populasi yang terdiri dari 100 orang, setiap anggota diberi nomor urut dari 1 hingga 100. Pengambilan sampel dilakukan dengan cara memilih anggota populasi berdasarkan kriteria tertentu, seperti hanya mengambil anggota dengan nomor ganjil, nomor genap, atau nomor yang merupakan kelipatan dari bilangan tertentu, seperti kelipatan dari bilangan lima (Sugiyono, 2013).

Preprocessing data, yang juga dikenal sebagai pra-pemrosesan data, mengacu pada rangkaian langkah yang dilakukan untuk membersihkan, mengubah, atau menyesuaikan data mentah sebelum digunakan dalam analisis atau pemodelan. Tujuannya adalah untuk meningkatkan kualitas data dengan menghilangkan gangguan (*noise*) atau pencilan (*outlier*) pada data, serta mempersiapkan data agar sesuai dengan persyaratan algoritma atau teknik yang akan digunakan dalam tahap

Volume 4, Nomor 1, Oktober 2024
p-ISSN : 2502-5139
e-ISSN : 2776-6039

selanjutnya (Alghifari & Juardi, 2021). *Preprocessing* data pada penelitian ini dilakukan dalam 2 tahap yatu tahap transformasi data menggunakan *Label Encoder* dan tahap normalisasi data menggunakan normalisasi *min-max*.

Label Encoder adalah teknik yang digunakan pada tahap *prepocessing* data untuk mengubah data kategorikal menjadi data numerik.*Label Encoder* mengubah setiap kategori menjadi angka (label numerik) secara berurutan sesuai dengan indeks atau urutan uniknya (Santoso *et al*., 2023).

Normalisasi *min-max* merupakan teknik normalisasi yang melibatkan transformasi linier pada data asli untuk mencapai keseimbangan dalam perbandingan nilai antara data sebelum dan setelah proses normalisasi. Rumus yang digunakan untuk metode ini dapat dinyatakan sebagai berikut (Nasution *et al*., 2019):

$$x_{new} = \frac{x \quad x_{min}}{x_{max} - x_{min}}$$

Keterangan :

x_{new} = nilai setelah normalisasi

x = nilai sebelum normalisasi

x_{min} = nilai minimal pada data sebelum normalisasi

x_{max} = nilai maksimal pada data sebelum normalisasi

Data yang telah dinormalisasi dibagi menjadi dua bagian, yaitu data *training* serta data *testing*. Data *training* digunakan pada proses pelatihan model klasifikasi, sedangkan data testing digunakan untuk menguji performa model yang telah dilatih. Hal ini bertujuan untuk memastikan bahwa model mempunyai kemampuan yang baik dalam mengklasifikasikan data yang belum pernah dilihat sebelumnya.

Tahap selanjutnya adalah menentukan nilai k yang optimal melalui proses *k-fold cross validation*, yang akan digunakan dalam proses klasifikasi dengan menggunakan algoritma *K-Nearest Neighbor*.

K -fold cross validation atau dapat disebut estimasi rotasi ialah sebuah teknik validasi model untuk menilai bagaimana hasil statistik analisis akanmenggeneralisasi kumpulan data independen (Tempola *et al*., 2018). *K-fold cross validation* adalah salah satu teknik yang sering digunakan untuk menentukan nilai k terbaik pada algoritma KNN dalam klasifikasi. Dalam metode *k-fold cross validation*, data dibagi menjadi k bagian yang sama besar, dengan nilai k yang disesuaikan oleh peneliti. Namun, nilai k harus dipilih secara hati-hati agar tidak terlalu besar sehingga dapat menyebabkan *overfitting*, atau terlalu kecil sehingga menyebabkan bias (Widyaningsih *et al*., 2021).

Gambar 1. Model 3-*fold cross validation*

Gambar 1 merupakan penggunaan 3 *fold cross validation*. Dimana setiap data akan di eksekusi sebanyak 3 kali dan setiap

Volume 4, Nomor 1, Oktober 2024
p-ISSN : 2502-5139
e-ISSN : 2776-6039

subset data akan memiliki kesempatan sebagai data *testing* atau data *training*. Model pengujian seperti berikut dengan diasumsikan nama setiap pembagian data yaitu D1, D2, dan D3 (Tempola *et al*., 2018):

1. Percobaan pertama data D1 sebagai data *testing* sedangkan D2 dan D3 sebagai data *training*.
2. Percobaan kedua data D2 sebagai data *testing* sedangkan data D1 dan D3 sebagai data *training*.
3. Pada percobaan terakhir atau percobaan ketiga data D3 sebagai data *testing* sedangkan D1 dan D2 sebagai data *training.*

Selanjutnya, menerapkan model klasifikasi dengan algoritma KNN berdasarkan variabel yang sudah ditentukan.KNN adalah salah satu algoritma klasifikasi yang bekerja dengan mengambil sejumlah k data terdekat sebagai acuan untuk menentukan kelas dari data baru.Algoritma ini mengklasifikasikan data berdasarkan kemiripan atau kedekatannya terhadap data lainnya.KNN ialah salah satu algoritma nonparametrik yang digunakan pada pengklasifikasian dan tidak memerlukan asumsi distribusi sehingga sebaran data bebas.*K-Nearest Neighbor* ialah jenis algoritma *machine learning* yang paling dasar dalam kasus klasifikasi (Cholil *et al*.,
2021). Elemen yang penting dalam keberhasilan algoritma k *-nearest neighbor* yaitu perhitungan jarak, di mana jarak digunakan untuk menentukan tingkat kemiripan antara data uji dan data latih.Semakin jauh jarak antara kedua data, semakin rendah kemiripannya, dan sebaliknya (Cholil *et al*., 2021).

Penelitian ini menggunakan perhitungan jarak *euclidean distance* dalam algoritma k *-nearest neighbor*, karena sifatnya mudah dipahami atau diinterpretasikan tanpa memerlukan penjelasan yang rumit atau kompleks (Id, 2021).Selain itu, *euclidean distance* juga memberikan hasil yang optimal dibandingkan dengan metode perhitungan jarak lainnya (Pribadi *et al*.,
2022).Keunggulan *euclidean distance* juga terletak pada kemampuannya untuk mengukur jarak dalam dua, tiga, atau lebih dimensi, sehingga dapat diterapkan dalam berbagai bidang, termasuk pengenalan pola, analisis data, dan pengolahan citra ., 2019):

$$d_{(x,y)} = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

(Miftahuddin *et al*., 2020). Rumus *euclidean distance*, yaitu sebagai berikut (Muslim *et al*

Keterangan :

$d_{(x,y)}$ = jarak antara dua titik atau vektor, yaitu titik zx (titik yang tidak diketahui atau titik baru) dan titik y (titik pada dataset yang sudah diketahui)

x_i = data *testing* ke-i

y_i = data *training* ke-i

Metode KNN memiliki beberapa langkah dalam penghitungan, yaitu (Cahyanti *et al*., 2020):

Volume 4, Nomor 1, Oktober 2024
p-ISSN : 2502-5139
e-ISSN : 2776-6039

1. Menentukan parameter *k* yang akan digunakan.
   Nilai *k* adalah jumlah tetangga terdekat yang letaknya paling dekat dengan objek data testing yang diuji. Berdasarkan nilai *k* inilah klasifikasi objek baru pada data testing dapat ditentukan dan akan masuk ke kelas yang telah ada nantinya.
2. Menghitung jarak antara data training dengan data testing.
   Menghitung jarak *euclidean* antara data *testing* dan data *training* untuk mencari peringkat dalam *k* yang telah ditentukan dan berfungsi untuk menguji ukuran yang bisa dipakai sebagai interpretasi kedekatan antara dua objek.
3. Mengurutkan jarak yang dihasilkan dalam urutan terkecil hingga terbesar. Hasil perhitungan yang diurutkan adalah hasil dari perhitungan jarak *euclidean*.Pengurutan ini dilakukan dari yang terkecil hingga yang paling besar nilainya.
4. Menentukan kelompok data mayoritas. Menentukan kelompok data mayoritas pada *k* untuk menentukan kelas objek berdasakan *k* yang ditentukan.

Gambar 2. Flowchart Algoritma *K-Nearest Neighbor*

Confusion matrix ialah sebuah matriks yang digunakan untuk mengevaluasi tingkat akurasi dari suatu algoritma klasifikasi terhadap kelas yang dihasilkan (Pratiwi *et al*., 2020).*Confusion matrix* biasanya disajikan dalam bentuk tabel matriks yang memuat informasi mengenai kelas asli dan hasil prediksi kelas. Contoh bentuk dari *confusion matrix* adalah sebagai berikut:

Tabel 1.*Confussion Matrix*

Confussion Matrix		**Nilai Prediksi**	
		TRUE	***FALSE***
Nilai Asli	***TRUE***	TP (*True Positive*)	FN (*False Negative*)
	FALSE	FP (*False Positive*)	TN (*True Negative*)

Tabel tersebut menjelaskan empat istilah penting dalam *confusion matrix* yang dipakai dalam mengukur kinerja suatu model klasifikasi dengan keterangan sebagai berikut:

1. *True Positive* (TP) ialah banyak data dengan kelas positif yang secara benar diklasifikasikan sebagai positif oleh model.
2. *False Positive* (FP) ialah banyak data dengan kelas negatif yang salah diklasifikasikan sebagai positif oleh model.
3. *False Negative* (FN) ialah banyak data dengan kelas positif yang salah diklasifikasikan sebagai negatif oleh model.

Volume 4, Nomor 1, Oktober 2024
p-ISSN : 2502-5139
e-ISSN : 2776-6039

4. *True Negative* (TN) ialah banyak data dengan kelas negatif yang secara benar diklasifikasikan sebagai negatif oleh model.

Dalam proses evaluasi akurasi suatu algoritma, salah satu metode yang bisa dipakai ialah menggunakan *confusion matrix* dan menghitung akurasi menggunakan rumus sebagai berikut (Han *et al.,* 2011).

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

Akurasi klasifikasi adalah ukuran seberapa tepat model klasifikasi yang dibentuk secara keseluruhan.Semakin tinggi akurasi, semakin baik model yang dihasilkan.Oleh karena itu, klasifikasi berupaya membentuk model dengan akurasi yang tinggi.

$$Precission = \frac{TP}{TP + FP} \times 100\%$$

Precission menggambarkan jumlah data kategori positif yang diklasifikasi dengan benar dibagi dengan total data yang diklasifikasi positif.

$$Recall = \frac{TP}{TP + FN} \times 100\%$$

Recall memperlihatkan beberapa persen data kategori positif yang terklasifikasi dengan benar oleh sistem.

PHP atau *Hypertext Preprocessor* ialah bahasa pemrograman yang dipakai untuk membuat *website* dengan pendekatan *server-side scripting*. Kelebihan utama dari PHP adalah sifatnya yang dinamis, memungkinkan pengembang untuk membuat aplikasi *web* yang responsif dan interaktif. ebagai bahasa pemrograman skrip, PHP dirancang khusus untuk membangun aplikasi *web* yang efisien dan handal (Novendri *et al*., 2019).

MySQL ialah sebuah sistem manajemen basis data relasional (RDBMS) yang memiliki sifat *open-source* serta mendapatkan popularitas yang luas.MySQL menyediakan *platform* yang andal dan efisien untuk penyimpanan, pengelolaan, dan akses data.Kehadirannya yang fleksibel dalam berbagai bahasa pemrograman membuat MySQL menjadi pilihan yang sangat populer sebagai basis data *backend* untuk pengembangan aplikasi *web* dan sistem informasi (Dhika *et al*., 2019).

XAMPP merupakan sebuah paket perangkat lunak web yang komprehensif yang bisa dipakai untuk keperluan belajar pemrograman *web*, terutama dalam bahasa PHP dan MySQL. XAMPP berfungsi sebagai *server* mandiri yang beroperasi di *localhost*, yang terdiri dari program *Apache* HTTP *Server*, database MySQL, serta interpreter bahasa pemrograman PHP (Anggraini *et al*., 2020).

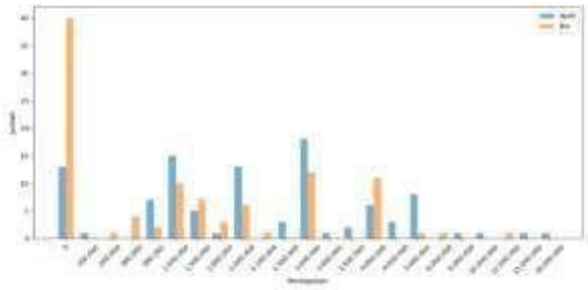
3. HASIL DAN PEMBAHASAN

Pengumpulan data pada penelitian ini dilaksanakan melalui survei (penyebaran angket). Berdasarkan hasil survei yang telah dilakukan, diperoleh dataset siswa SMA Negeri 1 Sinunukan, Mandailing Natal sebanyak 231 data dengan 12 variabel yaitu nama, jenis kelamin, pendapatan ayah, pendapatan ibu, pekerjaan ayah, pekerjaan ibu,

Volume 4, Nomor 1, Oktober 2024
p-ISSN : 2502-5139
e-ISSN : 2776-6039

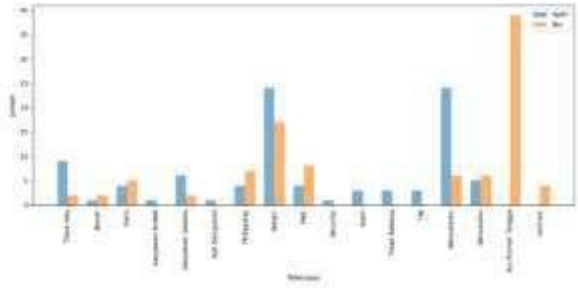
tanggungan orangtua, status keluarga, status tempat tinggal, transportasi, uang saku, dan penerima PIP.

Data yang telah diperoleh, kemudian dilakukan pemilihan sampel sebanyak 100 data yang digunakan sebagai sampel pada penelitian ini.Pemilihan sampel pada penelitian ini dilakukan menggunakan teknik sampling sistematis, dimana data yang diambil sebagai sampel pada dataset dilakukan dengan mengambil data yang memiliki no urut ganjil. Sehingga didapat sampel pada penelitian ini yaitu 100 data siswa yang memiliki 11 variabel yang akan digunakan dalam klasifikasi, yaitu pendapatan ayah (x_1), pendapatan ibu (x_2), pekerjaan ayah (x_3), pekerjaan ibu (x_4), tanggungan orangtua (x_5), status keluarga (x_6), status tempat tinggal (x_7), transportasi (x_8), uang saku (x_9), dan penerima PIP (y), sebagai target klasifikasi. Berikut ini adalah deskripsi data dari 11 variabel yang akan digunakan pada penelitian ini.



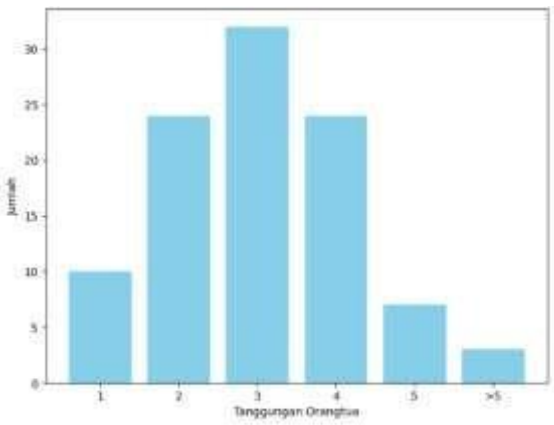
Gambar 3. Diagram Variabel Pendapatan Ayah dan Pendapatan Ibu

Berdasarkan gambar 3.diatas, dapat dilihat sebaran data pada variabel pendapatan ayah (x_1) dan pendapatan ibu (x_2). Pada variabel pendapatan ayah (x_1), nominal pendapatan yang memiliki jumlah terbanyak, yaitu pada nominal 3.000.000 dengan jumlah sebanyak 18 orang.Sedangkan pada variabel pendapatan ibu (x_2), nominal pendapatan yang memiliki jumlah terbanyak, yaitu pada nominal 0 dengan jumlah 40 orang.



Gambar 4. Diagram Variabel Pekerjaan Ayah dan Pekerjaan Ibu

Berdasarkan gambar 4.diatas, dapat dilihat sebaran data pada variabel pekerjaan ayah (x_3) dan pekerjaan ibu (x_4). Pada variabel pekerjaan ayah (x_3), pekerjaan yang memiliki jumlah terbanyak adalah pekerjaan wiraswasta dengan jumlah sebanyak 24 orang.Sedangkan pada variabel pekerjaan ibu (x_4), pekerjaaan yang memiliki jumlah terbanyak, yaitu pada pekerjaan Ibu Rumah Tangga dengan jumlah sebanyak 39 orang.

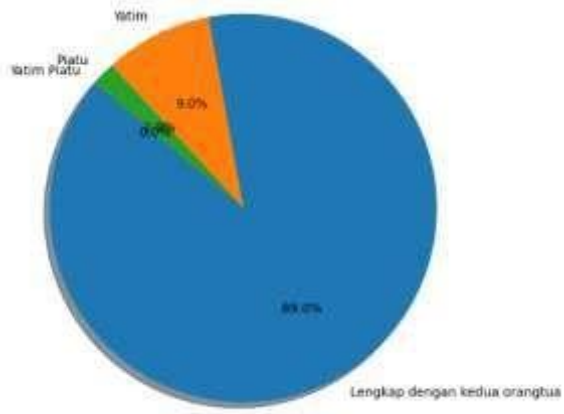


Gambar 5. Diagram Variabel Tanggungan Orangtua

Berdasarkan gambar 5.diatas, bisa diamati sebaran data pada variabel

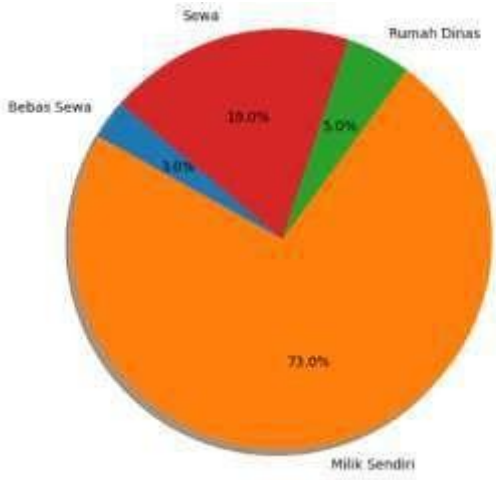
Volume 4, Nomor 1, Oktober 2024
p-ISSN : 2502-5139
e-ISSN : 2776-6039

tanggungan orangtua (x_5), dimana jumlah terbanyak terdapat pada kategori 3 sebanyak 32 orang dan jumlah terendah terdapat pada kategori >5 sebanyak 3 orang.



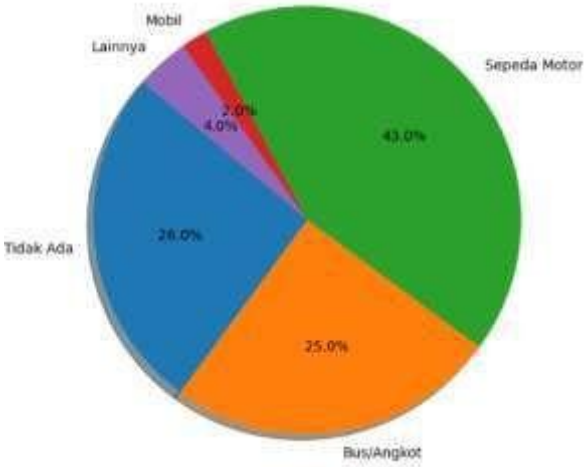
Gambar 6. Diagram Variabel Status Keluarga

Berdasarkan gambar 6.diatas, dapat dilihat sebaran data pada variabel status keluarga (x_6) dimana persentase terbanyak terdapat pada kategori Lengkap dengan kedua orangtua dengan persentase 89% dan persentase terendah terdapat pada kategori Yatim Piatu dengan persentase 0%.



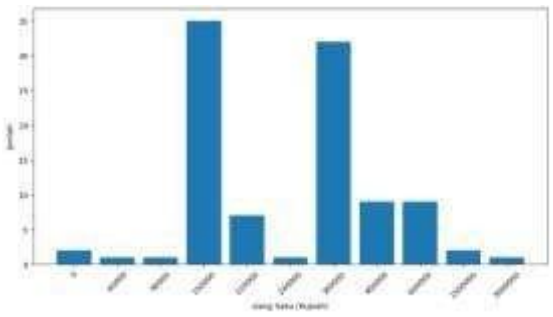
Gambar 7. Diagram Variabel Status Tempat Tinggal

Berdasarkan gambar 7. diatas, dapat dilihat sebaran data pada variabel status tempat tinggal (x_7) dimana persentase terbanyak terdapat pada kategori Milik Sendiri dengan persentase 73% dan persentase terendah terdapat pada kategori Bebas Sewa dengan persentase 3%.



Gambar 8. Diagram Variabel Transportasi

Berdasarkan gambar 8. diatas, bisa diamati sebaran data pada variabel status transportasi (x_8) dimana persentase terbanyak ada pada kategori Sepeda Motor dengan persentase 43% serta persentase terendah ada pada kategori Mobil dengan persentase 2%.

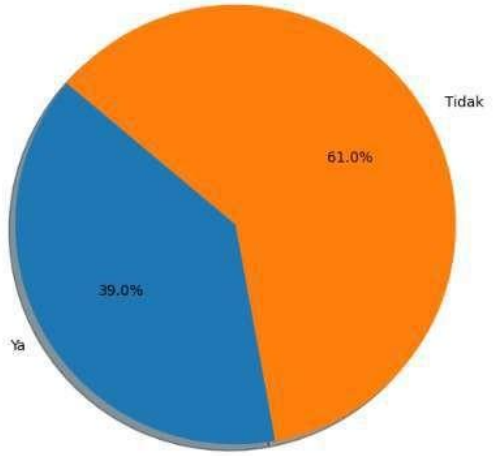


Gambar 9. Diagram Variabel Uang Saku

Berdasarkan gambar 9.diatas, dapat dilihat sebaran data pada variabel uang saku (x_9), dimana nominal uang saku yang memiliki jumlah terbanyak,

Volume 4, Nomor 1, Oktober 2024
p-ISSN : 2502-5139
e-ISSN : 2776-6039

yaitu pada nominal 150.000 dengan jumlah sebanyak 35 orang.



Gambar 10. Diagram Variabel Penerima PIP

Berdasarkan gambar 10.diatas, dapat dilihat sebaran data pada variabel penerima PIP (y) dimana persentase terbanyak ada pada kategori Tidak dengan persentase 61% serta persentase terendah ada pada kategori Ya dengan persentase 39%.

Setelah sampel ditentukan, selanjutnya yang dilakukan yaitu melakukan *prepocessing* data.Penelitian ini *preprocessing* data dilakukan dalam 2 tahap, yaitu transformasi data menggunakan *label encoder* dan normalisasi data menggunakan normalisasi *min-max*.

Label encoder digunakan untuk mengubah data atau variabel kategorikal pada dataset menjadi format numerik sesuai dengan indeks atau urutan uniknya, agar dapat digunakan dalam algoritma *k-nearest neighbor*. Terdapat beberapa variabel pada dataset yang ditranformasikan dari data kategorikal ke dalam format numerik, yaitu variabel pekerjaan ayah (x_3), pekerjaan ibu (x_4), status keluarga (x_6), status tempat tinggal (x_7), transportasi (x_8), dan penerima PIP (y).

Tabel 2.*Label Encoder* pada Variabel Pekerjaan Ayah

Pekerjaan Ayah (x_3)	**Label**
Tidak ada	0
Buruh	1
DPRD	2
Guru	3
Karyawan BUMN	4
Karyawan Swasta	5
Koperasi	6
Kuli Bangunan	7
PNS	8
POLRI	9
Pedagang	10
Pendeta	11
Pengacara	12
Pensiunan	13
Pensiunan BUMN	14
Petani	15
Security	16
Supir	17
TNI	18
Tidak Bekerja	19
Wiraswasta	20
Wirausaha	21

Berdasarkan tabel 2.diatas, jika pekerjaan ayah adalah Buruh, maka data pada dataset ditransformasi menjadi angka 1, dan seterusnya.

Tabel 3. *Label Encoder* pada Variabel Pekerjaan Ibu

Volume 4, Nomor 1, Oktober 2024
p-ISSN : 2502-5139
e-ISSN : 2776-6039

(x_4)	**Label**
Tidak ada	0
Buruh	1
Guru	2
Ibu Rumah Tangga	3
Karyawan Swasta	4
PNS	5
Pedagang	6
Pegawai Sekolah	7
Pengacara	8
Perawat	9
Petani	10
Pendeta	11
Petenun	12
Wiraswasta	13
Wirausaha	14

Pekerjaan Ibu

Berdasarkan tabel 3.diatas, jika pekerjaan ibu adalah Guru, maka data pada dataset ditransformasi menjadi angka 2, dan seterusnya.

Tabel 4. *Label Encoder* pada Variabel Status Keluarga

Status Keluarga (x_6)	**Label**
Lengkap dengn kedua orangtua	0
Piatu	1
Yatim	2
Yatim Piatu	3

Berdasarkan tabel 4.diatas, jika status keluarga adalah Piatu, maka data pada dataset ditransformasi menjadi angka 1, dan seterusnya.

Tabel 5.*Label Encoder* pada Variabel Status Tempat Tinggal

(x_7)	**Label**
Bebas Sewa	0
Milik Sendiri	1
Rumah Dinas	2
Sewa	3

Status Tempat Tinggal

Berdasarkan tabel 5.diatas, jika status tempat tinggal adalah Bebas Sewa, maka data pada dataset ditransformasi menjadi angka 0, dan seterusnya.

Tabel 6.*Label Encoder* pada Variabel Transportasi

Transportasi (x_8)	**Label**
Bus/Angkot	0
Lainnya	1
Mobil	2
Sepeda Motor	3
Tidak ada	4

Berdasarkan tabel 6.diatas, jika transportasi adalah Sepeda Motor, maka data pada dataset ditransformasi menjadi angka 3, dan seterusnya.

Tabel 7. *Label Encoder* pada Variabel Penerima PIP

(y)	**Label**
Ya	0
Tidak	1

Penerima PIP

Berdasarkan tabel 7.diatas, jika data pada variabel penerima pip termasuk dalam kategori Ya, maka data pada dataset ditransformasi menjadi angka 0, kemudian

Volume 4, Nomor 1, Oktober 2024
p-ISSN : 2502-5139
e-ISSN : 2776-6039

jika data pada variabel penerima pip termasuk dalam kategori Tidak, maka data pada dataset ditransformasi menjadi angka 1.

Setelah transformasi data menggunakan *label encoder* dilakukan, langkah selanjutnya yaitu melakukan normalisasi data pada dataset yang sudah ditransformasikan menggunakan normalisasi *min-max*. Variabel-variabel pada dataset yang dinormalisasikan, yaitu variabel pendapatan ayah (x_1), pendapatan ibu (x_2), pekerjaan ayah (x_3), pekerjaan ibu (x_4), tanggungan orangtua (x_5), status keluarga (x_6), status tempat tinggal (x_7), transportasi
(x_8), dan uang saku (x_9).

Tabel 8. Hasil Normalisasi *Min-Max* (x_{new}) pada Variabel Pendapatan Ayah

Pendapatan Ayah (x_1)	$x_{new} = \frac{x - x_{min}}{x_{max} - x_{min}}$
0	0
100.000	0.0056
500.000	0.0278
1.000.000	0.0556
1.500.000	0.0833
1.800.000	0.1000
2.000.000	0.1111
2.500.000	0.1389
3.000.000	0.1667
3.400.000	0.1889
3.500.000	0.1944
4.000.000	0.2222
4.500.000	0.2500
5.000.000	0.2778
8.000.000	0.4444
10.000.000	0.5556
15.000.000	0.8333
18.000.000	1

Berdasarkan tabel 8.diatas, jika nominal pendapatan ayah pada data adalah 3.000.000, maka nominal data pada dataset setelah normalisasi berubah menjadi 0.1667, dan seterusnya.

Tabel 9. Hasil Normalisasi *Min-Max* (x_{new}) pada Variabel Pendapatan Ibu

Pendapatan Ibu (x_2)	$x_{new} = \frac{x - x_{min}}{x_{max} - x_{min}}$
0	0
200.000	0.0167
300.000	0.0250
500.000	0.0417
1.000.000	0.0833
1.500.000	0.1250
1.800.000	0.1500
2.000.000	0.1667
2.100.000	0.1750
3.000.000	0.2500
4.000.000	0.3333
5.000.000	0.4167
6.000.000	0.5000
12.000.000	1

Berdasarkan tabel 9.diatas, jika nominal pendapatan ibu pada data adalah 1.500.000, maka nominal data pada dataset setelah normalisasi berubah menjadi 0.1250, dan seterusnya.

Tabel 10. Hasil Normalisasi *Min-Max*

Volume 4, Nomor 1, Oktober 2024
p-ISSN : 2502-5139
e-ISSN : 2776-6039

(x_{new}) pada Variabel Pekerjaan Ayah

Pekerjaan Ayah (x_3)	$x_{new} = \frac{x - x_{min}}{x_{max} - x_{min}}$
0	0
1	0.0476
2	0.0952
3	0.1428
4	0.1905
5	0.2380
6	0.2857
7	0.3333
8	0.3810
9	0.4286
10	0.4762
11	0.5238
12	0.5714
13	0.6190
14	0.6667
15	0.7143
16	0.7619
17	0.8095
18	0.8571
19	0.9048
20	0.9524
21	1

Berdasarkan tabel 10.diatas, jika nilai pekerjaan ayah pada data adalah 15, maka nilai data pada dataset setelah normalisasi berubah menjadi 0.7143, dan seterusnya.

Tabel 11.Hasil Normalisasi *Min-Max* (x_{new}) pada Variabel Pekerjaan Ibu

Pekerjaan Ibu (x_4)	$x_{new} = \frac{x - x_{min}}{x_{max} - x_{min}}$
0	0
1	0.0714
2	0.1429
3	0.2143
4	0.2857
5	0.3571
6	0.4286
7	0.5000
8	0.5714
9	0.6429
10	0.7143
11	0.7857
12	0.8571
13	0.9286
14	1

Berdasarkan tabel 11.diatas, jika nilai pekerjaan ibu pada data adalah 14, maka nilai data pada dataset setelah normalisasi berubah menjadi 1, dan seterusnya.

Tabel 12.Hasil Normalisasi *Min-Max* (x_{new}) pada Variabel Tanggungan Orangtua

Tanggungan Orangtua ($x5$)	$x_{new} = \frac{x - x_{min}}{xmax - xmin}$
1	0
2	0.1667
3	0.3333
4	0.5000
5	0.6667
7	1

Berdasarkan tabel 12.diatas, jika nilai tanggungan orangtua pada data adalah 3, maka nilai data pada dataset setelah normalisasi berubah menjadi 0.3333, dan seterusnya.

Tabel 13. Hasil Normalisasi *Min-Max*

Volume 4, Nomor 1, Oktober 2024
p-ISSN : 2502-5139
e-ISSN : 2776-6039

(x_{new}) pada Variabel Status Keluarga

Status Keluarga (x_6)	$x_{new} = \frac{x - x_{min}}{x_{max} - x_{min}}$
0	0
1	0.5000
2	1

Berdasarkan tabel 13.diatas, jika nilai status keluarga pada data adalah 1, maka nilai data pada dataset setelah normalisasi berubah menjadi 0.5000, dan seterusnya.

Tabel 14. Hasil Normalisasi *Min-Max* (x_{new}) pada Variabel Status Tempat Tinggal

Status Tempat Tinggal (x_7)	$x_{new} = \frac{x - x_{min}}{x_{max} - x_{min}}$
0	0
1	0.3333
2	0.6667
3	1

Berdasarkan tabel 14.diatas, jika nilai status tempat tinggal pada data adalah 2, maka nilai data pada dataset setelah normalisasi berubah menjadi 0.6667, kemudian jika nilai status tempat tinggal pada data adalah 1, maka nilai data pada dataset setelah normalisasi berubah menjadi 0.3333, dan seterusnya.

Tabel 15.Hasil Normalisasi *Min-Max* (x_{new}) pada Variabel Transportasi

Transportasi (x_8)	$x_{new} = \frac{x - x_{min}}{x_{max} - x_{min}}$
0	0
1	0.2500
2	0.5000
3	0.7500
4	1

Berdasarkan tabel 15.diatas, jika nilai variabel transportasi pada data adalah 3, maka nilai data pada dataset setelah normalisasi berubah menjadi 0.7500, kemudian jika nilai variabel transportasi pada data adalah 4, maka nilai data pada dataset setelah normalisasi berubah menjadi 1, dan seterusnya.

Tabel 16.Hasil Normalisasi *Min-Max* (x_{new}) pada Variabel Uang Saku

Uang Saku (x_9)	$x_{new} = \frac{x - x_{min}}{x_{max} - x_{min}}$
0	0
60.000	0.0200
90.000	0.0300
150.000	0.0500
210.000	0.0700
240.000	0.0800
300.000	0.1000
450.000	0.1500
600.000	0.2000
1.500.000	0.5000
3.000.000	1

Jurnal Cermatika
LPPM Universitas Graha Nusantara
https://jurnalugn.id/index.php/cermatika
DOI httpsdoi.org10.64168cermatika.v4i1

Volume 4, Nomor 1, Oktober 2024
p-ISSN : 2502-5139
e-ISSN : 2776-6039

Berdasarkan tabel 16.diatas, jika nominal uang saku pada data adalah 150.000, maka nominal data pada dataset setelah normalisasi berubah menjadi 0.0500, kemudian jika nominal uang saku pada data adalah 450.000, maka nominal data pada dataset setelah normalisasi berubah menjadi 0.1500, dan seterusnya.

Setelah data di normalisasi, langkah berikutnya yang dilakukan adalah melakukan *split* data menjadi data *training* serta data *testing* dengan perbandingan 90:10. Setelah data dibagi, langkah selanjutnya adalah menentukan nilai *k* yang optimal untuk KNN menggunakan *k-fold cross validation*. Nilai yang dipakai pada penentuan nilai *k* untuk algoritma KNN dalam penelitian ini adalah 3, 4, 5, 7, dan 9.Sehingga diperoleh hasil perhitungan penentuan nilai *k* untuk algoritma KNN dengan menggunakan 5-*fold cross validation* ialah sebagai berikut.

Tabel 17. Hasil perhitumgan penentuan nilai *k* untuk algoritma *k-nearest neighbor* menggunakan *5-fold cross validation*

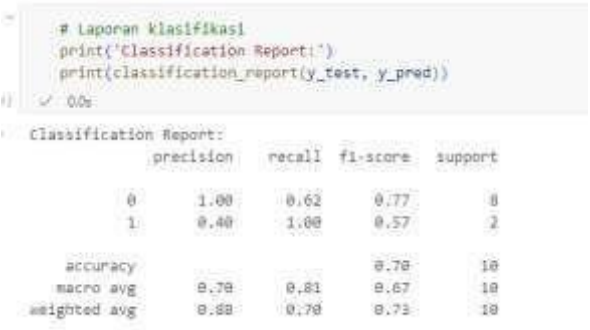
Nilai *k*	***Cross Validation Accuracies***					***Mean Cross Validation***
	1fold	***2-fold***	***3-fold***	***4fold***	***5fol d***	***Accuracy***
k = 3	0.65	0.95	0.7	0.6	0.8	0.74
k = 4	0.75	0.85	0.7	0.55	0.75	0.72
k = 5	0.75	0.8	0.65	0.5	0.75	0.69
k = 7	0.8	0.8	0.65	0.6	0.75	0.72
k = 9	0.75	0.8	0.7	0.65	0.75	073

Berdasarkan hasil perhitungan penentuan nilai *k* untuk algoritma KNN menggunakan 5-*fold cross validation* pada tabel 17. diatas, dapat disimpulkan bahwa nilai *k* yang optimal untuk digunakan pada algoritma KNN yaitu *k = 3*dengan nilai *mean cross validation accuracy* sebesar 0.74, yang merupakan nilai *mean cross validation accuracy* yang paling tinggi daripada nilai *k* lainnya.

Setelah menentukan nilai *k* untuk algoritma KNN, langkah selanjutnya yang dilakukan adalah penerapan model klasifikasi dengan menggunakan algoritma KNN.Penerapan model klasifikasi ini dilakukan dengan menggunakan bahasa pemograman *python* melalui bantuan *software Visual Studio Code*. Berikut ini ialah hasil penerapan model klasifikasi algoritma KNN dengan menggunakan nilai *k = 3*.

Jurnal Cermatika
LPPM Universitas Graha Nusantara
https://jurnalugn.id/index.php/cermatika
DOI httpsdoi.org10.64168cermatika.v4i1

Volume 4, Nomor 1, Oktober 2024
p-ISSN : 2502-5139
e-ISSN : 2776-6039



```
# Laporan klasifikasi
print('Classification Report:')
print(classification_report(y_test, y_pred))
```

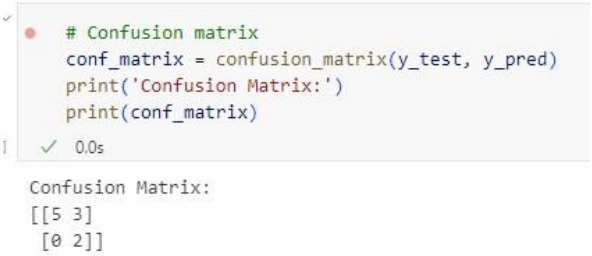
0.0s

```
Classification Report:
              precision    recall  f1-score   support

           0       1.00      0.62      0.77         8
           1       0.40      1.00      0.57         2

    accuracy                           0.70        10
   macro avg       0.70      0.81      0.67        10
weighted avg       0.88      0.70      0.73        10
```

Gambar 11. Hasil penerapan model klasifikasi algoritma KNNdengan menggunakan nilai *k = 3*

Berdasarkan gambar 11.diatas, diperoleh hasil laporan penerapan model klasifikasi algoritma KNN dengan akurasi sebesar 70% dan nilai *precission*sebesar 70% serta nilai *recall* sebesar 81%. Setelah nilai akurasi, *precission*, dan *recall* diperoleh, langkah selanjutnya adalah melakukan evaluasi dengan menggunakan *confussion matrix*.

Berikut ini adalah hasil *confussion matrix* yang diperoleh dari penerapan model klasifikasi dengan menggunakan bahasa pemograman *python* melalui bantuan *software Visual Studio Code*.



```
# Confusion matrix
conf_matrix = confusion_matrix(y_test, y_pred)
print('Confusion Matrix:')
print(conf_matrix)
```

0.0s

```
Confusion Matrix:
[[5 3]
 [0 2]]
```

Gambar 12. Hasil *Confussion Matrix*

Berdasarkan hasil *confussion matrix* pada gambar diatas, diperoleh hasil sebagai berikut.

1. *True Positif* (TP) = 5, artinya ada 5 data penerima PIP yang secara benar diklasifikasikan sebagai penerima PIP.
2. *False Positive* (FP) = 0, artinya tidak ada data yang bukan penerima PIP yang diklasifikasikan sebagai penerima PIP.
3. *False Negative* (FN) = 3, artinya ada 3 data penerima PIP yang diklasifikasikan sebagai bukan penerima PIP.
4. *True Negative* (TN) = 2, artinya ada 2 data bukan penerima PIP yang diklasifikasikan secara benar sebagai bukan penerima PIP.

Dengan menggunakan hasil evaluasi *confussion matrix* tersebut, maka dapat dihitung hasil evalusi akurasi, *precission* dan *recall* pada model. Berdasarkan evaluasi *confussion matrix,* maka diperoleh nilai akurasi sebesar 70%, nilai *precission* sebesar 100%, dan nilai *recall* sebesar 62,5%. Sehingga diperoleh kesimpulan bahwa setelah dilakukan evaluasi mode menggunakan *confussion matrix*, tidak ada peningkata dan penurunan pada nilai akurasi. Sementara itu, pada nilai *precission* terjadi peningkatan sebesar 30%, dan pada nilai
recall terjadi penurunan sebesar 18,5%

Penerapan Sistem

1. Halaman *Dashboard* Halaman *dashboard* pada sistem menampilkan penjelasan mengenai klasifikasi metode *k-nearest neighbor*.Halaman *dashboard* ini bisa diakses oleh *user* dan *admin*.

Jurnal Cermatika
LPPM Universitas Graha Nusantara
https://jurnalugn.id/index.php/cermatika
DOI httpsdoi.org10.64168cermatika.v4i1

Volume 4, Nomor 1, Oktober 2024
p-ISSN : 2502-5139
e-ISSN : 2776-6039

Gambar 13. Halaman *Dashboard*

2. Halaman Klasifikasi Halaman klasifikasi digunakan untuk mengklasifikasikan siswa layak untuk menerima PIP atau tidak dengan menginput data siswa sesuai dengan atribut/kriteria yang telah ditentukan. Setelah data diinput, *web* akan menampilkan perhitungan algoritna *k nearest neighbor* mulai dari dataset yang ada, normalisasi, dan perhitungan *euclidean distance* serta hasil klasifikasi Halaman klasifikasi ini bisa diakses oleh *user* dan *admin*.

Gambar 14. Halaman Klasifikasi 1

Gambar 15. Halaman Klasifikasi 2

3. Halaman *Login*

Halaman *login* hanya digunakan oleh *admin* untuk mengakses beberapa halaman lainnya.

Gambar 16. Halaman *Login*

4. Halaman Atribut

Halaman atribut menampilkan seluruh atribut yang digunakan pada sistem.Halaman atribut juga digunakan untuk menambah dan menghapus atribut.Halaman atribut ini hanya digunakan oleh *admin*.

Gambar 17. Halaman Atribut

5. Halaman Nilai Atribut Halaman nilai atribut menampilkan nilai untuk atribut yang kategorikal (tidak numerik).Halaman nilai atribut juga digunakan untuk menambah dan menghapus nilai atribut berdasarkan atribut yang sudah ada.Halaman nilai atribut ini hanya digunakan oleh *admin.*

Volume 4, Nomor 1, Oktober 2024
p-ISSN : 2502-5139
e-ISSN : 2776-6039

Gambar 18. Halaman Nilai Atribut

6. Halaman Dataset
   Halaman dataset menampilkan seluruh data yang digunakan pada sistem untuk klasifikasi.Halaman dataset ini hanya digunakan oleh *admin*.Halaman ini juga digunakan untuk menghapus dan menambah data.

Gambar 19. Halaman Dataset

7. Halaman Akurasi Halaman akurasi menampilkan hasil akurasi yang dilakukan terhadap data *testing*.Halaman ini hanya digunakan oleh *admin*. Jumlah data *testing* dan nilai *k* dapat diinput sesuai keinginan *admin* serta data *testing* yang digunakan hanya terdapat 3 pilihan saja yaitu data dari awal, acak, dan data dari akhir. Jika data yang dipilih adalah data acak maka akurasi yang dihasilkan akan berbedabeda.

Gambar 20. Halaman Akurasi 1

Gambar 21. Halaman Akurasi 2

8. Halaman Hasil
   Halaman hasil menampilkan hasil klasifikasi yang telah dilakukan pada halaman klasifikasi.Halaman ini hanya digunakan oleh *admin*.

Gambar 22. Halaman Hasil

9. Halaman *Password* Halaman *password* digunakan untuk mengubah *password admin*.Halaman ini hanya digunakan oleh *admin.*

Volume 4, Nomor 1, Oktober 2024
p-ISSN : 2502-5139
e-ISSN : 2776-6039

Gambar 23. Halaman *Password*

4. KESIMPULAN

Sesuai dengan hasil penelitian yang sudah dilakukan,maka diperoleh kesimpulan sebagai berikut :

1. Sistem klasifikasi penentuan calon penerima beasiswa PIP menggunakan algoritma *k-nearest neighbor* berbasis *web* yang telah di rancang dapat diakses melalui *localhost* setelah mendownload dan mengesktrak file pada link *googledrive* berikut : https://drive.google.com/drive/folders/1 T8-r15QIa3xd4M3bhUzGyviiKvYS8h_S?u sp=drive_link
2. Hasil evaluasi dari model klasifikasi penerima beasiswa Program Indonesia Pintar (PIP) menggunakan algoritma *knearest neighbor* dengan pembagian data 90:10 dan nilai *k = 3*, diperoleh hasil akurasi sebesar 70%, nilai precission sebesar 100%, serta nilai recall sebesar 62,5%.
3. Hasil penelitian yang telah dilaksanakan ialah sistem klasifikasi calon penerima beasiswa PIP menggunakan algoritma *k-nearest neighbor*, dengan hasil pengujian dari 10 data uji, terdapat 7 (70%) data yang bernilai benar, dan 3 (30%) data yang bernilai salah.

UCAPAN TERIMA KASIH

Puji dan syukur kepada Tuhan Yang Maha Esa yang telah memberikan rahmat dan karuniaNya, sehingga penulis dapat menyelesaikan skripsi ini dengan baik. Penulis mengucapkan terima kasih kepada kedua orang tua dan kedua adik penulis atas segala doa dan dukungan yang penuh dengan cinta kasih kepada penulis. Terima kasih kepada dosen pembimbing skripsi Ibu Anita Adinda, S.Si., M.Si. atas atas bimbingan, motivasi, dan arahannya yang luar biasa sehingga semua tahap dalam penyusunan ini dapat dilalui dengan baik. Terima kasih juga penulis ucapkan untuk semua pihak yang membantu dalam penyelesaian penelitian ini.

DAFTAR PUSTAKA

Alghifari, F., & Juardi, D. (2021).Penerapan Data Mining Pada Penjualan Makanan Dan Minuman Menggunakan Metode Algoritma Naïve Bayes.*Jurnal Ilmiah Informatika (JIF)*, *9*(2), 75–80.

Anggraini, Y., Pasha, D., Damayanti, & Setiawan, A. (2020).Sistem Informasi Penjualan Sepeda Berbasis Web Menggunakan Framework Codeigniter.*Jurnal Teknologi Dan Sistem Informasi*, *1*(2), 64–70. https://doi.org/10.33365/jtsi.v1i2.236

Cahyanti, D., Rahmayani, A., & Husniar, S. A. (2020).Analisis performa metode Knn pada Dataset pasien pengidap Kanker

Volume 4, Nomor 1, Oktober 2024
p-ISSN : 2502-5139
e-ISSN : 2776-6039

Payudara.*Indonesian Journal of Data and Science*, *1*(2), 39–43.

Cholil, S. R., Handayani, T., Prathivi, R., & Ardianita, T. (2021). Implementasi Algoritma Klasifikasi K-Nearest Neighbor (KNN) Untuk Klasifikasi Seleksi Penerima Beasiswa. *IJCIT (Indonesian Journal on Computer and Information Technology)*, *6*(2), 118– 127.

Dhika, H., Isnain, N., & Tofan, M. (2019).Manajemen Villa Menggunakan Java Netbeans Dan Mysql.*IKRA-ITH INFORMATIKA : Jurnal Komputer Dan Informatika*, *3*(2), 104–110. https://journals.upiyai.ac.id/index.php/ikraithinformatika/article/view/324

Han, J., Kamber, M., & Pei, J. (2011). Data Mining: Concepts and Techniques. In *Data Mining: Concepts and Techniques* (Third Edit). Elsvier. https://doi.org/10.1016/C2009-0-61819-5

Id, I. D. (2021). MACHINE LEARNING: Teori, Studi Kasus dan Implementasi Menggunakan Python. In *UR PRESS* (1st ed.). UR PRESS.

Iriantoro, D. N. D., Dewi, C., & Fitriani, D. (2018). Klasifikasi pada Penyakit Dental Caries Menggunakan Gabungan KNearest Neighbor dan Algoritme Genetika.*Jurnal Pengembangan Teknologi Dan Ilmu Komputer*, *2*(8), 2926–2933.

Miftahuddin, Y., Umaroh, S., & Karim, F. R. (2020). Perbandingan Metode Perhitungan Jarak Euclidean, Haversine, Dan Manhattan Dalam Penentuan Posisi Karyawan. *Jurnal Tekno Insentif*, *14*(2), 69–77. https://doi.org/10.36787/jti.v14i2.270

Muslim, M. A., Prasetiyo, B., Mawarni, E. L. H., Herowati, A. J., Mirqotussa'adah, Rukmana, S. H., & Nurzahputra, A. (2019). Data Mining Algoritma C4.5 Disertai contoh kasus dan penerapannya dengan program computer.In *UNNES Repository*.

Nasution, D. A., Khotimah, H. H., & Chamidah, N. (2019). Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma K-NN.*Computer Engineering, Science and System Journal*, *4*(1), 78–82.

Novendri, M. S., Saputra, A., & Firman, C. E. (2019).Aplikasi Inventaris Barang Pada MTS Nurul Islam Dumai Menggunakan PHP dan MySQL.*Lentera Dumai*, *10*(2), 46–57.

Pratiwi, B. P., Handayani, A. S., & Sarjana.(2020). Pengukuran Kinerja Sistem Kualitas Udara Dengan Teknologi WSN Menggunakan Confusion Matrix.*Jurnal Informatika UPGRIS*, *6*(2), 66–75.

Pribadi, W. W., Yunus, A., & Wiguna, A. S. (2022). Perbandingan Metode K-Means Euclidean Distance Dan Manhattan Distance Pada Penentuan Zonasi Covid19 Di Kabupaten Malang. *JATI*

Volume 4, Nomor 1, Oktober 2024
p-ISSN : 2502-5139
e-ISSN : 2776-6039

(Jurnal Mahasiswa Teknik Informatika), *6*(2), 493–500. https://doi.org/10.36040/jati.v6i2.4808

Rakista, P. M. (2021). Implementasi Kebijakan Program Indonesia Pintar (PIP).*Sawala : Jurnal Administrasi Negara*, *8*(2), 224–232.

Retnoningsih, E., & Pramudita, R. (2020). Mengenal Machine Learning Dengan Teknik Supervised Dan Unsupervised Learning Menggunakan Python. *Bina Insani Ict Journal*, *7*(2), 156–165.

Ridho, M. R., Hairani, H., Latif, K. A., & Hammad, R. (2021).Kombinasi Metode AHP dan TOPSIS untuk Rekomendasi Penerima Beasiswa SMK Berbasis Sistem Pendukung Keputusan.*Jurnal Tekno Kompak*, *15*(1), 26–39.

Santoso, H., Putri, R. A., & Sahbandi.(2023). Deteksi Komentar Cyberbullying pada Media Sosial Instagram Menggunakan Algoritma Random Forest.*Jurnal Manajemen Informatika (JAMIKA)*, *13*(1), 62–72.

Sugiyono.(2013). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Alfabeta.

Suryaman, S. A., Magdalena, R., & Sa'idah, S. (2021). Klasifikasi Cuaca Menggunakan Metode VGG-16, Principal Component Analysis Dan KNearest Neighbor.*Jurnal Ilmu Komputer Dan Informatika*, *1*(1), 1–8.

Tempola, F., Muhammad, M., & Khairan, A. (2018).Perbandingan Klasifikasi Antara KNN dan Naive Bayes pada Penentuan Status Gunung Berapi dengan K-Fold Cross Validation.*Jurnal Teknologi Informasi Dan Ilmu Komputer (JTIIK)*, *5*(5), 577–584.

Widyaningsih, Y., Arum, G. P., & Prawira, K. (2021).Aplikasi K-Fold Cross Validation Dalam Penentuan Model Regresi Binomial Negatif Terbaik.*BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, *15*(2), 315– 322. https://doi.org/10.30598/barekengvol15i ss2pp315-322