

Implementasi Algoritma K-Nearest Neighbor Untuk Klasifikasi Penerimaan Beasiswa Program Indonesia Pintar

Chusi Yanasari

Universitas Adhirajasa Reswara Sanjaya

Toni Arifin

Universitas Adhirajasa Reswara Sanjaya

Alamat : Jl. Sekolah Internasional No.1-2, Antapani, Bandung 40282, Indonesia

*Korespondensi penulis : chusysdr@gmail.com

Abstract. Scholarships are a form of assistance in the form of educational expenses provided by the government or foundations to students or students who are categorized as from underprivileged families. However, in determining scholarship recipients, there are still many scholarship recipients who come from wealthy families, while those from less fortunate families do not receive this assistance. This may be due to calculations and data processing that still use manual methods, causing scholarship recipients to not be on target. The purpose of this research is to simplify and minimize calculation errors in determining scholarship recipients for the Smart Indonesia Program (PIP) at SMK Karya Medika. Therefore, for calculating and processing PIP scholarship recipients data, data mining techniques can use the classification method using the K-NN algorithm. K-Nearest Neighbor is a data classification method that will be used for data objects based on learning data that is closer to the object. In this study using the Confusion Matrix test so as to obtain an accuracy value of 80.00%.

Keywords: Implementation, Classification, K-Nearest Neighbor, Scholarship Recipients Smart Indonesia Pintar (PIP)

Abstrak. Beasiswa merupakan bentuk bantuan berupa biaya pendidikan yang diberikan oleh pemerintah atau yayasan kepada siswa atau pelajar yang dikategorikan dari keluarga kurang mampu. Namun dalam menentukan penerima beasiswa masih banyak penerima beasiswa yang berasal dari keluarga kaya, sedangkan dari keluarga kurang mampu tidak menerima bantuan ini. Hal ini mungkin dikarenakan perhitungan dan pengolahan data yang masih menggunakan cara manual, sehingga menyebabkan penerima beasiswa tidak tepat sasaran. Tujuan dari penelitian ini adalah mempermudah dan meminimalisir kesalahan perhitungan dalam menentukan penerima beasiswa Program Indonesia Pintar (PIP) di SMK Karya Medika. Oleh karena itu, untuk perhitungan dan mengolah data penerima beasiswa PIP dapat menggunakan teknik *data mining* metode klasifikasi menggunakan algoritma K-NN. K-Nearest Neighbor merupakan suatu metode klasifikasi data yang akan digunakan terhadap objek data berdasarkan data pembelajaran yang jaraknya lebih dekat dengan objek tersebut. Pada penelitian ini menggunakan pengujian *Confusion Matrix* sehingga memperoleh nilai akurasi sebesar 80.00%

Kata Kunci: Implementasi, Klasifikasi, K-Nearest Neighbor, Penerima Beasiswa Program Indonesia Pintar (PIP)

LATAR BELAKANG

Pendidikan mempunyai peran penting dalam kemajuan bangsa dan negara untuk mencetak generasi muda yang lebih maju, akan tetapi untuk mencapai pendidikan tidak selalu berjalan dengan lancar karena masih banyak ditemukan siswa yang putus sekolah penyebabnya faktor perekonomian yang tidak mencukupi untuk melanjutkan pendidikan. Oleh karena itu untuk membantu siswa agar tetap terus melanjutkan pendidikan pemerintah membuat Program Indonesia Pintar (PIP).

Program Indonesia Pintar (PIP) yang merupakan salah satu program bantuan yang diberikan oleh pemerintah kepada siswa atau pelajar dari keluarga yang kurang mampu, agar siswa bisa terus melanjutkan pendidikan (Nata & Suparmadi, 2022). Kategori keluarga kurang mampu dapat dilihat dari penghasilan orang tua, jumlah orang yang berkerja dalam keluarga dan ada berapa tanggungan dalam keluarga, hal tersebut dapat membantu dalam penentuan calon penerima beasiswa. Akan tetapi, permasalahannya dalam menentukan penerimaan beasiswa yaitu masih banyak ditemukan penerima bantuan beasiswa yang berasal dari keluarga mampu sedangkan dari keluarga yang tidak mampu justru tidak mendapatkan bantuan beasiswa tersebut (Nata & Suparmadi, 2022). Hal ini mungkin dikarenakan cara perhitungan dan pengolahan data yang masih menggunakan cara manual, hal tersebut menyebabkan penerima beasiswa yang tidak tepat pada sasaran (Homepage et al., 2021). Agar beasiswa dapat diberikan tepat pada sasarannya, dibutuhkan suatu teknik ataupun metode untuk mengolah data calon penerima beasiswa Program Indonesia Pintar (PIP), untuk mengolah data calon penerima beasiswa Program Indonesia Pintar dapat menggunakan pengolahan *data mining* yang akan diklasifikasikan berdasarkan data *training* dan data *testing* yang diambil dari data calon penerima beasiswa di SMK Karya Medika (Homepage et al., 2021).

Data yang telah diolah menggunakan *data mining* akan menghasilkan sebuah data baru yang berisi suatu informasi yang berguna untuk menentukan keputusan di masa yang akan mendatang. Salah satu teknik analisa *Data Mining* yang akan digunakan dalam penelitian ini yaitu klasifikasi.

Klasifikasi adalah sebuah proses pengkelompokan objek berdasarkan kelompok sebelumnya yang sudah diberi label kelas, data yang diperlukan yaitu data pelatihan atau *training* (Mulyani, 2020). Algoritma yang akan digunakan dalam penelitian ini yaitu algoritma *K-Nearest Neighbor* yang berdasarkan pada penelitian sebelumnya yang dilakukan oleh (Zaman et al., 2021) mengenai Komparasi Metode Klasifikasi Batik Menggunakan *Neural Network* Dan *K-Nearest Neighbor* Berbasis Ekstraksi Fitur Tekstur. Berdasarkan dari hasil studi komparasi yang menggunakan metode *Backpropagation Neural Network* dengan hasil 90,11% sedangkan dengan metode *K-Nearest Neighbor* memperoleh 96,00% dari hasil yang diperoleh dapat disimpulkan bahwa penggunaan metode *K-Nearest Neighbor* merupakan pilihan yang tepat untuk klasifikasi. Tujuan dari penelitian ini adalah mempermudah dan meminimalisir kesalahan perhitungan dalam menentukan penerima beasiswa Program Indonesia Pintar (PIP) di SMK Karya Medika.

KAJIAN TEORITIS

2.1. Data Mining

Menurut (Koda et al., 2020) memberi batasan bahwa “*Data mining* adalah proses yang mempekerjakan satu atau lebih teknik pembelajaran komputer (*mechine learning*) untuk menganalisis dan mengekstraksi pengetahuan (*knowledge*) secara otomatis”. *Data mining* yaitu suatu langkah basis data yang digunakan untuk analisis pengetahuan atau sering disebut dengan *Knowledge Discovery in Database* (Iriane & Nurfaizah, 2023). Tahapan-tahapan dalam *data mining* seperti *data selection*, *data preprocessing*, *transformation data* dan *Knowledge Interpretation* atau *Evaluation*.

2.2. Klasifikasi

Klasifikasi berdasarkan sistem informasi yaitu suatu teknik mengklafikasikan data ke dalam label kelas yang sudah ada sebelumnya (Homepage et al., 2021). Model dari klasifikasi yaitu suatu nilai objek data yang akan dimasukkan kedalam kelas dengan jumlah kelas yang sudah ada (Nata & Suparmadi, 2022). Tujuan utama dari klasifikasi yaitu mengorganisasikan dan mengelompokkan sebuah data ke dalam kelas-kelas yang berbeda (Pauji et al., 2020). Proses klasifikasi didasarkan pada empat komponen yaitu: kelas, Prediktor, *training dataset*, dan *testing dataset*.

2.3. Algoritma *K-Nearest Neighbor*

K-Nearest Neighbor merupakan sebuah algoritma yang digunakan untuk menghitung kedekatan kasus baru dengan kasus lama (Noviana et al., 2019). Salah satu kelebihan menggunakan metode *K-Nearest Neighbor* yaitu teknik yang digunakan sederhana, efisien dan efektif dalam pengolahan objek data dan lain-lain (Karepesina & Zahrotun, 2023). Tujuan dari algoritma klasifikasi adalah mengklasifikasikan objek baru berdasarkan atribut dan data *training* yang digunakan (Koda et al., 2020).

Berikut tahapan proses algoritma *K-Nearest Neighbor* menurut (Soleha et al., 2023) sebagai berikut:

1. Menentukan nilai k.
2. Menghitung kuadrat jarak *Euclid* antara data *training* dengan data *testing*.
3. Mengurutkan jarak *Euclid* dari yang terkecil.
4. Menentukan kelas mayoritas dari nilai K tetangga terdekat.

(2.1)

$$d = \sqrt{\sum_{i=1}^n (a_i - b_i)^2}$$

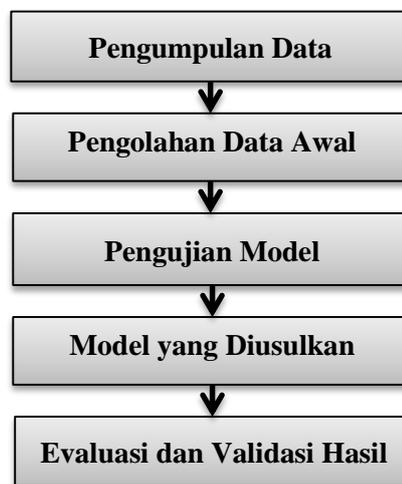
Keterangan :

- d : jarak *Euclidean*
- a_i : data *training*
- b_i : data *testing*
- i : baris ke - i dari tabel
- n : jumlah data *testing*.

METODE PENELITIAN

3.1. Perancangan Penelitian

Menurut Dawson dalam (Andini et al., 2020) pada umumnya terdapat empat metode penelitian yang digunakan antara lain penelitian langsung, eksperimen, studi kasus dan *survey*. Maka dalam penelitian ini menggunakan metode penelitian langsung. Berikut ini adalah tahapan penelitian pada gambar III.1.



Gambar III.I. Tahapan Penelitian

3.2. Pengumpulan Data

Pada penelitian ini penulis menggunakan dataset yang diambil langsung dari SMK Karya Medika, data yang digunakan pada penelitian ini adalah data penerima beasiswa Program Indonesia Pintar (PIP) dari tahun 2019 sampai dengan 2023.

3.3. Pengolahan Data Awal

Pada tahap ini akan dilakukan pembersihan data (*cleaning data*) yang dimana akan menghilangkan beberapa data tidak penting atau tidak dipakai untuk data yang tidak konsisten akan diperbaiki dan diperiksa, sehingga menjadi data yang siap untuk diteliti atau dijadikan objek penelitian (Resmiati & Arifin, 2021). Selanjutnya akan dilakukan sebuah transformasi data yang akan diolah langsung menggunakan model yang telah ditentukan pada Microsoft Excel.

a. *Data Cleaning*

Pada proses pembersihan (*cleaning*) akan dilakukan pembersihan data yang tidak sesuai dan akan memperbaiki kesalahan data pada penulisan. Atribut data yang tidak akan digunakan dan akan dihilangkan, yaitu: NIPD, NISN, tempat lahir, tanggal lahir, NIK, agama, kebutuhan khusus, alamat, RT/RW, dusun, kelurahan, kecamatan, kode pos, jenis tinggal, telepon, hp, email, SKHUN, penerima KPS, nomor KPS, nama orang tua dan wali murid, tahun lahir orang tua dan wali murid, jenjang pendidikan orang tua dan wali murid, kebutuhan khusus orang tua, penghasilan dan pekerjaan ibu dan wali murida ujian nasional.

b. *Data Transformasi*

Pada tahap ini akan dilakukan transformasi data yang dimana data asli akan diubah menjadi bentuk yang sesuai untuk mempermudah dalam proses penambangan atau pengolahan data. Atribut-atribut yang berjenis alphabet seperti alat transportasi, pekerjaan orang tua, penghasilan orang tua, status penerima PIP tahun sebelumnya pada *data mining* akan di *transformasi* terlebih dahulu ke dalam bentuk angka atau numerikal. Berikut ini merupakan tabel transformasi data ditunjukkan pada Tabel III.3.

Tabel III.3. Transformasi Data

No	Atribut	Keterangan	Nilai Variabel
1	Alat Transportasi	Lainnya	1
		Angkutan Umum Bus/Pete-Pete	2
		Sepeda	3
		Jalan Kaki	4
		Sepeda Motor	5
2	Pekerjaan Ayah	Karyawan Swasta	1
		Sudah meninggal	2
		PNS/TNI/Polri	3
		Pedagang besar	4
		Wirasaha	5
		Buruh	6
		Lainnya	7
		Nelayan	8
		Pedagang kecil	9
		Petani	10
		Wiraswasta	11
3	Penghasilan Ayah	0	1
		<500000	2
		500000-999999	3
		<1000000	4
		1000000-1999999	5
		1000000-2000000	6
		2000000-4999999	7
4	Pekerjaan Ibu	Karyawan Swasta	1
		Sudah Meninggal	2
		Nelayan	3
		Buruh	4
		Lainnya	5
		Petani	6
		Pedagang Kecil	7
		Wiraswasta	8
		Tidak Berkerja	9
5	Penghasilan Ibu	0	1
		<500000	2
		500000-999999	3
		1000000-1999999	4
		2000000-4999999	5
		<1000000	6
6	Status Penerima PIP Tahun Sebelumnya	Ya Menerima	1
		Tidak Menerima	2
Keterangan Penerima PIP	Ya		
	Tidak		

c. *Data selection*

Seleksi data adalah tahap untuk memutuskan atribut-atribut yang diperlukan untuk digunakan pada tahap *data mining*. Variabel independen yang akan digunakan seperti nama,

jenis kelamin (JK), alat transportasi, pekerjaan orang tua, penghasilan orang tua, layak penerima PIP tahun sebelumnya. Sedangkan untuk keterangan penerima beasiswa PIP termasuk dalam variabel dependen dengan output yang dikeluarkan yaitu YA atau TIDAK (Mulyani, 2020).

d. Normalisasi Data

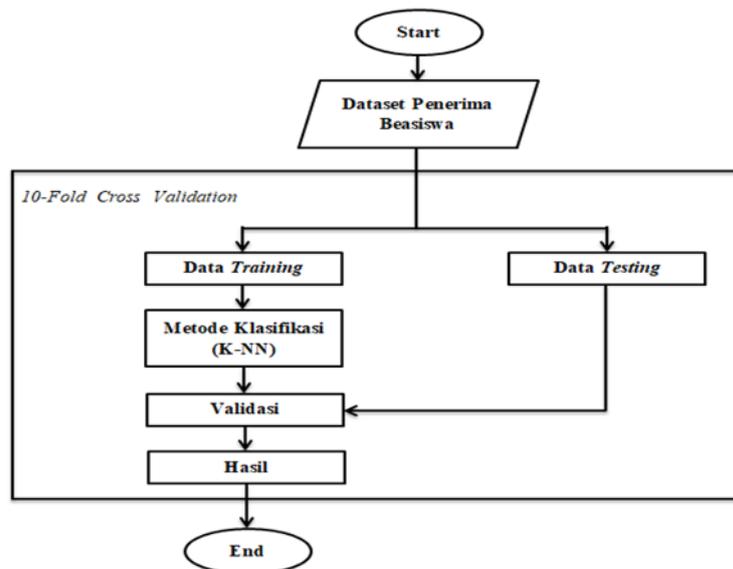
Pada penelitian ini untuk memperkecil range data perlu dilakukan sebuah normalisasi data dengan menggunakan min-max normalization dengan melakukan transformasi linier data asli yang menghasilkan keseimbangan nilai perbandingan antara data sebelum dan sesudah proses normalisasi. Di bawah ini merupakan rumus dari normalisasi data min-max normalization (Mulyani, 2020):

$$V' = \frac{V - \min \text{kolom } 1}{\max \text{kolom } 1 - \min \text{kolom } 1} \cdot (\text{new max } A - \text{new min } A) + \text{new min } A$$

Keterangan :

- V' = Nilai dari data baru hasil dari normalisasi
- V = Nilai dari data sebelum di normalisasi
- New max A = Batas nilai maksimum terbaru
- New min A = Batas nilai minimum terbaru
- Max A = Nilai maksimum pada kolom
- Min A = Nilai minimum pada kolom

3.4. Model yang diusulkan



Sumber : (Arifin & Ariesta, 2019)

Gambar III.2. Model Yang diusulkan

3.5. Pengujian Model

Berikut ini adalah langkah-langkah dari pengujian model dalam penelitian ini sebagai berikut:

1. Pada langkah pertama perlu menyiapkan sebuah dataset yang nantinya akan diuji dan dataset tersebut adalah data penerimaan beasiswa PIP.
2. Pada langkah kedua akan melakukan sebuah proses pembersihan data atau data *cleaning* data yang tidak diperlukan atau data kosong akan dihapus pada dataset tersebut.
3. Pada langkah ketiga akan dilakukan suatu proses transformasi atribut yang dimana sebuah data yang awalnya berbentuk alfabet akan diubah menjadi bentuk numerikal agar dapat diolah menggunakan *data mining*.
4. Pada langkah keempat akan dilakukan suatu sampling untuk membagi *dataset* penerimaan beasiswa PIP menjadi data *training* dan data *testing*.
5. Pada langkah kelima data tersebut akan diuji dengan menggunakan algoritma *K-Nearest Neighbor* sehingga diharapkan dapat menghasilkan nilai akurasi yang tinggi.

3.6. Evaluasi dan Validasi Hasil

Setelah pengolahan data melalui rapidminer dengan menggunakan algoritma *K-Nearest Neighbor* dan melakukan pengujian dengan menggunakan *10-Flod Cross Validation*, langkah selanjutnya akan di evaluasi dan validasi terhadap hasil dengan menggunakan *Confusion Matrix* sebagai performace dengan hasil nilai *Accuary*.

HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

Pada penelitian ini dilakukan pengumpulan data untuk penelitian yang diambil dari SMK Karya Medika, data yang diambil yaitu data siswa penerima beasiswa PIP kelas XII dari tahun 2019 sampai dengan 2023. Data yang diperoleh sebanyak 360 data dalam bentuk Ms. Excel sehingga dapat mempermudah dalam pengolahan data. Data tersebut terdiri dari 340 data yang sudah diketahui kategori penerimanya dan 20 data yang belum diketahui kategori penerimanya.

4.2. Pengolahan Data Awal

Pada tahapan pengumpulan data diperoleh data sebanyak 360 dengan jumlah atribut awal sebanyak 44 atribut. Kemudian data tersebut akan dilakukakn Pre-Processing data seperti di bawah ini:

a. Penyusunan dan Transformasi Data

Pada tahapan ini dilakukan sebuah transformasi terhadap atribut data yang berbentuk alfabet diubah menjadi bentuk numerikal atau angka berdasarkan transformasi data yang sudah ditentukan dengan menggunakan software microsoft excel. Sehingga dapat diperoleh hasil sebagai berikut:

No	NAMA	JL	ALAT TRANSPORTASI	PEKERJAAN AYAH	PENGHASILAN AYAH	PEKERJAAN IBU	PENGHASILAN IBU	PI
1	A. FAISAL	L		11	3		3	
2	ALFI MAULANA	L	5	11	3		9	
3	ALI NURSIDIK	L	2	7	3		7	2
4	ALFIA WIDIA PRAMESTI	P	3	7	3		5	
5	AFTON RANGGA PERMANA	L	5	11	3		9	
6	ANI NUR INDAH SARI	P	5	6	2		9	1
7	ANIFATUN NASIFAH	L	2	8	3		6	2
8	AYU KUSUMAWATI	L	2	9	2		5	
9	AYU PIPIT KARMILA	L	5	11	3		8	3
10	CUCU NOVIA	P	3	9	2		5	
11	CUCUN JUWITA	P	5	10			9	
12	DELIA RAHMAWATI	P	3	6	2		5	
13	DEWI SAPITRI	L	5	6	2		4	2
14	DIMAS KURNIAWAN	L	3	6	2		4	2
15	EVI SILVIANI	P	2	11	3		9	
16	FADIA KOMALASARI	P	5	1	5		8	2
17	FAIMI GUNTUR SUPRAPTO	L	5	11	5		8	4
18	FAJAR TEGUH MULYAWAN	L		11	3		9	1
19	FINA MARSELA	P	2	11	7		7	4
20	HERMAWAN	L		6	3		9	1
21	HUMAEDAH	P	4	9	2		8	2
22	ILHAM CAHYA FADILLAH	L	2	10	5		6	3
23	IMAS MASITOH	P	3	6	2		7	2

Gambar IV.1. Hasil Penyusunan dan Transformasi Data

b. Seleksi Atribut

Pada tahap ini akan dilakukan seleksi atribut dari semua atribut awal yang kemudian akan diseleksi untuk digunakan pada penelitian. Pada penelitian ini atribut yang akan digunakan yaitu nama, jenis kelamin, alat transportasi, pekerjaan dan penghasilan ayah, pekerjaan dan penghasilan ibu, status penerima PIP tahun sebelumnya dan keterangan penerima PIP dengan keluaran YA atau TIDAK menerima. Atribut yang akan dilakukan perhitungan di *RapidMiner* yaitu alat transportasi, pekerjaan dan penghasilan ayah, pekerjaan dan penghasilan ibu, status penerima PIP tahun sebelumnya.

c. Pembersihan Data (*Missing Value*)

Pada tahap ini akan dilakukan akan dilakukan pembersihan atau menghilangkan data yaitu alat transportasi, pekerjaan dan penghasilan ayah, pekerjaan dan penghasilan ibu, status penerima PIP tahun sebelumnya. Sedangkan atribut keterangan penerima PIP tidak dilakukan pembersihan data karena dalam keterangan terdapat *missing label* sebagai data *testing* yang akan dilakukan pengujian.

d. Normalisasi Data

Normalisasi data akan dilakukan terhadap atribut alat transportasi, pekerjaan dan penghasilan ayah, status layak PIP tahun sebelumnya dengan menggunakan perhitungan *min-*

max normalization pada *software RapidMiner*. Berikut ini terdapat rumusan dan pengertian dari perhitungan *min-max normalization* (Mulyani, 2020).

(4.5)

$$V' = \frac{v - \min A}{\max A - \min A} \cdot (\text{new max } A - \text{new min } A) + \text{new min } A$$

Keterangan :

V' = Nilai dari data baru hasil dari normalisasi

V = Nilai dari data sebelum di normalisasi

New max A = Batas nilai maksimum terbaru

New min A = Batas nilai minimum terbaru

Max A = Nilai maksimum pada kolom

Min A = Nilai minimum pada kolom

Tabel IV.2. Data Training dan Menyusun Kategori

No	Nama	B1	B2	B3	B4	B5	B6
1	A. FAISAL	4	11	3	8	3	2
2	ALFI MAULANA	5	11	3	9	2	2
3	ALI NURSIDIK	2	7	3	7	2	2
4	ALIFIA WIDIA PRAMESTI	3	7	3	5	2	2
5	AFTON RANGGA PERMANA	5	11	3	9	2	1
6	ANI NUR INDAH SARI	5	6	2	9	1	2
7	ANIFATUN NASIFAH	2	8	3	6	2	1
8	AYU KUSUMAWATI	2	9	2	5	2	1
9	AYU PIPIT KARMILA	5	11	3	8	3	1
10	CUCU NOVIA	3	9	2	5	2	1
11	CUCUN JUWITA	5	10	2	9	2	1
12	DELIA RAHMAWATI	3	6	2	5	2	1
13	DEWI SAPITRI	5	6	2	4	2	1
14	DIMAS KURNIAWAN	3	6	2	4	2	2
.....
356	HALYATUL MAHFUDOH	5	11	5	7	5	1
357	SELVIANA PUTRI	5	7	3	9	1	1
358	SINTA	5	7	3	9	1	1
359	TRIA ALPIANA	5	9	3	9	1	1
360	WULAN KRISNAWATI	5	10	3	9	1	1

Keterangan:

B1 = Alat Transportasi

B2 = Pekerjaan Ayah

B3 = Penghasilan Ayah

B4 = Pekerjaan Ibu

B5 = Penghasilan Ibu

B6 = Status Penerima PIP Tahun Sebelumnya

Tabel IV.3. Nilai Maximun dan Nilai Minimum

Atribut	Maximum	Minimum
Alat Tansportasi	5	1
Pekerjaan Ayah	11	1
Penghasilan Ayah	7	1
Pekerjaan Ibu	9	1
Penghasilan Ibu	6	1
Status Layak Penerima PIP Tahun Sebelumnya	2	1

Perhitungan manual baris pertama yaitu sebagai berikut ini:

- **Baris 1 kolom 1**

$$\begin{aligned}
 V' &= \frac{v - \min \text{kolom } 1}{\max \text{kolom } 1 - \min \text{kolom } 1} \cdot (\text{new max } A - \text{new min } A) + \text{new min } A \\
 &= \frac{4 - 1}{5 - 1} \cdot (1 - 0) + 0 \\
 &= 0.750
 \end{aligned}$$

- **Baris 1 kolom 2**

$$\begin{aligned}
 V' &= \frac{v - \min \text{kolom } 2}{\max \text{kolom } 2 - \min \text{kolom } 2} \cdot (\text{new max } A - \text{new min } A) + \text{new min } A \\
 &= \frac{11 - 1}{11 - 1} \cdot (1 - 0) + 0 \\
 &= 1
 \end{aligned}$$

- **Baris 1 kolom 3**

$$\begin{aligned}
 V' &= \frac{v - \min \text{kolom } 3}{\max \text{kolom } 3 - \min \text{kolom } 3} \cdot (\text{new max } A - \text{new min } A) + \text{new min } A \\
 &= \frac{3 - 1}{7 - 1} \cdot (1 - 0) + 0 \\
 &= 0.333
 \end{aligned}$$

- **Baris 1 kolom 4**

$$\begin{aligned}
 V' &= \frac{v - \min \text{kolom } 4}{\max \text{kolom } 4 - \min \text{kolom } 4} \cdot (\text{new max } A - \text{new min } A) + \text{new min } A \\
 &= \frac{8 - 1}{9 - 1} \cdot (1 - 0) + 0 \\
 &= 0.875
 \end{aligned}$$

- **Baris 1 kolom 5**

$$\begin{aligned}
 V' &= \frac{v - \min \text{kolom } 5}{\max \text{kolom } 5 - \min \text{kolom } 5} \cdot (\text{new max } A - \text{new min } A) + \text{new min } A \\
 &= \frac{3 - 1}{6 - 1} \cdot (1 - 0) + 0 \\
 &= 0.400
 \end{aligned}$$

- **Baris 1 kolom 6**

$$\begin{aligned} V' &= \frac{v - \min \text{kolom } 6}{\max \text{kolom } 6 - \min \text{kolom } 6} \cdot (\text{new max } A - \text{new min } A) + \text{new min } A \\ &= \frac{2 - 1}{2 - 1} \cdot (1 - 0) + 0 \\ &= 1 \end{aligned}$$

Perhitungan manual baris kedua yaitu sebagai berikut ini:

- **Baris 2 kolom 1**

$$\begin{aligned} V' &= \frac{v - \min \text{kolom } 1}{\max \text{kolom } 1 - \min \text{kolom } 1} \cdot (\text{new max } A - \text{new min } A) + \text{new min } A \\ &= \frac{5 - 1}{5 - 1} \cdot (1 - 0) + 0 \\ &= 1 \end{aligned}$$

- **Baris 2 kolom 2**

$$\begin{aligned} V' &= \frac{v - \min \text{kolom } 2}{\max \text{kolom } 2 - \min \text{kolom } 2} \cdot (\text{new max } A - \text{new min } A) + \text{new min } A \\ &= \frac{11 - 1}{11 - 1} \cdot (1 - 0) + 0 \\ &= 1 \end{aligned}$$

- **Baris 2 kolom 3**

$$\begin{aligned} V' &= \frac{v - \min \text{kolom } 3}{\max \text{kolom } 3 - \min \text{kolom } 3} \cdot (\text{new max } A - \text{new min } A) + \text{new min } A \\ &= \frac{3 - 1}{7 - 1} \cdot (1 - 0) + 0 \\ &= 0.333 \end{aligned}$$

- **Baris 2 kolom 4**

$$\begin{aligned} V' &= \frac{v - \min \text{kolom } 4}{\max \text{kolom } 4 - \min \text{kolom } 4} \cdot (\text{new max } A - \text{new min } A) + \text{new min } A \\ &= \frac{9 - 1}{9 - 1} \cdot (1 - 0) + 0 \\ &= 1 \end{aligned}$$

- **Baris 1 kolom 5**

$$\begin{aligned} V' &= \frac{v - \min \text{kolom } 5}{\max \text{kolom } 5 - \min \text{kolom } 5} \cdot (\text{new max } A - \text{new min } A) + \text{new min } A \\ &= \frac{2 - 1}{6 - 1} \cdot (1 - 0) + 0 \\ &= 0.200 \end{aligned}$$

• **Baris 1 kolom 6**

$$V' = \frac{v - \text{min kolom } 6}{\text{max kolom } 6 - \text{min kolom } 6} \cdot (\text{new max } A - \text{new min } A) + \text{new min } A$$

$$= \frac{2 - 1}{2 - 1} \cdot (1 - 0) + 0$$

$$= 1$$

Perhitungan di atas juga berlaku untuk perhitungan baris kedua dan selanjutnya.

4.3. Pengujian Model Split Validation dan Cross Validation

Pada tahap ini akan dilakukan perbandingan terhadap dua pengujian yaitu antara menggunakan pengujian *Split Validation* dan *10-Cross Validation*, maka nanti dapat disimpulkan pengujian mana yang menghasilkan kinerja baik dan akurat.

a. Pengujian Model *Split Validation*

Pada tahap penelitian ini pengujian model dilakukan menggunakan *Split Validation* sebanyak 5 kali pengujian terhadap nilai K=3, untuk melihat hasil perbandingan dari 50:50 sampai dengan 90:10 nilai akurasi yang lebih tinggi pada algoritma *K-Nearest Naighbor*. Di bawah ini merupakan hasil dari pengujian model yang telah dilakukan menggunakan *Split Validation*, yaitu:

Tabel IV.4. Pengujian Hasil Akurasi Split Validation

	Perbandingan	Jumlah K	Nilai Akurasi
Pengujian ke-1	50:50	K=1	64.58%
		K=2	64.58%
		K=3	72.92%
		K=4	72.92%
		K=5	72.92%
		K=6	72.92%
		K=7	72.92%
		K=8	72.92%
		K=9	72.92%
		K=10	72.92%
Pengujian ke-2	60:40	K=1	66.67%
		K=2	66.67%
		K=3	74.36%
		K=4	74.36%
		K=5	69.23%
		K=6	71.79%
		K=7	71.79%
		K=8	71.79%
		K=9	74.36%
		K=10	74.36%
Pengujian ke-3	70:30	K=1	58.62%
		K=2	58.62%
		K=3	65.52%
		K=4	65.52%
		K=5	62.07%
		K=6	62.07%
		K=7	68.97%
		K=8	68.97%
		K=9	68.97%
		K=10	68.97%
Pengujian ke-4	80:20	K=1	50.00%
		K=2	50.00%
		K=3	60.00%
		K=4	55.00%
		K=5	55.00%
		K=6	55.00%
		K=7	65.00%
		K=8	65.00%
		K=9	65.00%
		K=10	65.00%
Pengujian ke-5	90:10	K=1	77.78%
		K=2	77.78%
		K=3	66.67%
		K=4	77.78%
		K=5	55.56%
		K=6	55.56%
		K=7	55.56%
		K=8	55.56%
		K=9	55.56%
		K=10	55.56%

Dapat disimpulkan nilai K=3 yang paling tinggi akurasinya terdapat pada perbandingan 60:40 dengan nilai akurasi 74.36 %.

b. Pengujian Model *10-Cross Validation*

Pada pengujian ini *RapidMiner* akan membagi *dataset* menjadi dua bagian, yaitu data *training* dan data *testing* dan akan dilakukan pengujian sebanyak 10 kali dan akan menghasilkan nilai akurasi yang lebih akurat. Di bawah ini merupakan hasil dari pengujian model yang telah dilakukan menggunakan *10-Cross Validation*, yaitu:

Tabel IV.5. Pengujian Hasil Akurasi *10-Fold Cross Validation*

	Jumlah K	Nilai Akurasi
Pengujian ke-1	K=1	80.00%
Pengujian ke-2	K=2	80.00%
Pengujian ke-3	K=3	77.06%
Pengujian ke-4	K=4	79.12%
Pengujian ke-5	K=5	72.94%
Pengujian ke-6	K=6	75.00%
Pengujian ke-7	K=7	73.24%
Pengujian ke-8	K=8	72.94%
Pengujian ke-9	K=9	74.41%
Pengujian ke-10	K=10	72.65%

Dapat disimpulkan nilai akurasi tertinggi terdapat pada K=1 dan K=2 akan tetapi untuk perhitungan algoritma *K-Nearest Neighbor* yang akan dilakukan adalah menggunakan nilai K=3 dengan akurasi 77.06% agar mendapatkan hasil perhitungan yang lebih akurat.

4.4. Pengujian Algoritma *K-Nearest Neighbor*

Berikut ini merupakan sebuah *dataset* penelitian yang akan diimplementasikan dengan algoritma *K-Nearest Neighbor*.

Tabel IV.6. Dataset Penelitian

NO	NAMA	B1	B2	B3	B4	B5	B6	KETERANGAN PENERIMA PIP
1	A. FAISAL	0.750	1	0.333	0.875	0.400	1	TIDAK
2	ALFIA MULLANA	1	1	0.333	1	0.200	1	TIDAK
3	ALINURSIDIK	0.250	0.600	0.333	0.750	0.200	1	TIDAK
4	ALFIA WIDIA PRAMESTI	0.500	0.600	0.333	0.500	0.200	1	TIDAK
5	AFTON FANGGA PERLIANA	1	1	0.333	1	0.200	0	YA
6	ANI NUR INDAH SARI	1	0.500	0.167	1	0	1	TIDAK
7	ANFATUN NASIFAH	0.250	0.700	0.333	0.625	0.200	0	YA
8	AYU KUSUMAWATI	0.250	0.800	0.167	0.500	0.200	0	YA
9	AYU PIPIT KEMILA	1	1	0.333	0.875	0.400	0	YA
10	CUCU NOVIA	0.500	0.800	0.167	0.500	0.200	0	YA
11	CUCUN JURITA	1	0.900	0.167	1	0.200	0	YA
12	DELLA RAHMAWATI	0.500	0.500	0.167	0.500	0.200	0	YA
13	DEWI SAPITRI	1	0.500	0.167	0.375	0.200	0	YA
14	DMAS KURNIAWAN	0.500	0.500	0.167	0.375	0.200	1	TIDAK
15	EVI SILVIANI	0.250	1	0.333	1	0.200	1	TIDAK
16	FADIA KOKALASARI	1	0	0.667	0.875	0.200	1	TIDAK
17	FABMI GUNTUR SUPRAPTO	1	1	0.667	0.875	0.600	1	TIDAK
18	FAJAR TEGUH MULYAWAN	0.750	1	0.333	1	0	0	YA
19	FINA MARSELA	0.250	1	1	0.750	0.600	1	TIDAK
20	HERMAWAN	0.750	0.500	0.333	1	0	0	YA
.....
339	ZALFA ARIZTYA	1	1	0.333	0.625	0.200	0	YA
340	ZIRKA INDAH MANIA	1	0.700	0.333	0.125	0	0	YA
341	ANGGIT BINTORO	0	0.800	0.333	0.875	0.200	0	?
342	BAGUS ARYANTO	1	0.500	0.667	1	0.200	0	?
343	DEDEH SETIA NINGSIH	1	0.900	0.500	0.625	1	0	?
344	NUR FAIZAH	0.250	0.800	0.167	0.250	0.200	0	?
345	SELY SELVIANA	0.750	1	0.333	1	0	0	?
346	ABDUL HOLIL	1	0.900	0.333	0.625	0.200	0	?
347	AMELI MONIKA	1	0.700	0.333	0.750	0.200	0	?
348	APRIYANTO	1	0.900	0.333	0.625	0.200	0	?
349	ICH TRISNAWATI	1	0.600	0.333	0.625	0.200	0	?
350	JUMADAH	1	0.800	0.167	0.500	0	0	?
351	LUDA	1	0.600	0.333	0.625	0.200	0	?
352	NUR HALIZAH	1	0.600	0.333	1	0	0	?
353	YANI KHOLIPAH TUN	1	0	0.333	1	0	0	?
354	LUSI FEBRIANI	1	0.100	0	1	0	0	?
355	DIVA KARTIKA	1	0.800	0.333	0.750	0.200	0	?
356	HALYATUL MAHFUDOH	1	1	0.667	0.75	0.800	0	?
357	SELVIANA PUTRI	1	0.800	0.333	1	0	0	?
358	SINTA	1	0.600	0.333	1	0	0	?
359	TRIA ALPIANA	1	0.800	0.333	1	0	0	?
360	WULAN KRISNAWATI	1	0.9000	0.333	1	0	0	?

Keterangan:

B2 = Pekerjaan Ayah

B3 = Penghasilan Ayah

B4 = Pekerjaan Ibu

B5 = Penghasilan Ibu

B6 = Status Penerima PIP Tahun Sebelumnya

Berikut ini merupakan sebuah hasil perhitungan euclidean distace menggunakan perhitungan manual algoritma *K-Nearest Neighbor*.

Tabel IV.7. Hasil Perhitungan Euclidean Distance

NO	NAMA	EUCLIDEAN DISTANCE	KETERANGAN
1	A. FAISAL	1.281	TIDAK
2	ALFI MAULANA	1.164	TIDAK
3	ALI NURSIDIK	1.182	TIDAK
4	ALIFIA WIDIA PRAMESTI	1.172	TIDAK
5	AFTON RANGGA PERMANA	1.070	YA
6	ANI NUR INDAH SARI	3.428	TIDAK
7	ANIFATUN NASIFAH	1.041	YA
8	AYU KUSUMAWATI	3.428	YA
9	AYU PIPIT KARMILA	3.641	YA
10	CUCU NOVIA	3.943	YA
....
339	ZALFA ARIZTYA	1.324	YA
340	RIKHA INDAH MANIA	1.284	YA
341	ANGGIT BINTORO	-	?

a. Menentukan Kelas Mayoritas

Berdasarkan hasil perhitungan di atas menggunakan rumus *euclidean distance*, maka kelas mayoritas data *testing* pertama yang diurutkan dari ranking terkecil sampai dengan terbesar. Dapat dilihat sebagai berikut ini:

Tabel IV.8. Penentuan Kelas Mayoritas

EUCLIDEAN DISTANCE	RANKING	KETERANGAN
0	1	TIDAK
0.230	2	TIDAK
0.243	3	TIDAK
-	-	TIDAK

Dari tabel penentuan kelas mayoritas di atas, maka disimpulkan data *testing* pertama atas nama **Anggit Bintoro** termasuk ke dalam kategori siswa yang TIDAK beasiswa PIP berdasarkan hasil dari penerapan algoritma *KNN*.

Tabel IV.9. Hasil Klasifikasi *Data Testing*

NO	NAMA	A1	A2	A3	A4	A5	A6	KETERANGAN
1	ANGGIT BINTORO	0	0.800	0.333	0.875	0.200	0	TIDAK
2	BAGUS ARYANTO	1	0.500	0.667	1	0.200	0	TIDAK
3	DEDEH SETIA NINGSIH	1	0.900	0.500	0.625	1	0	TIDAK
4	NUR FAIZAH	0.250	0.800	0.167	0.750	0.200	0	TIDAK
5	SELY SELVIANA	0.750	1	0.333	1	0	0	TIDAK
6	ABDUL HOLIL	1	0.900	0.333	0.625	0.200	0	TIDAK
7	AMELI MONIKA	1	0.700	0.333	0.750	0.200	0	TIDAK
8	APRIYANTO	1	0.900	0.333	0.625	0.200	0	TIDAK
9	ICIH TRISNAWATI	1	0.600	0.333	0.625	0.200	0	TIDAK
10	JUMAIDAH	1	0.600	0.167	0.500	0	0	TIDAK
11	LUDIA	1	0.600	0.333	0.625	0.200	0	TIDAK
12	NUR HALIZAH	1	0.600	0.333	1	0	0	TIDAK
13	YANI KHOLIPAH TUN	1	0	0.333	1	0	0	TIDAK
14	LUSI FEBRIANI	1	0.100	0	1	0	0	TIDAK
15	DIVA KARTIKA	1	0.800	0.333	0.750	0.200	0	TIDAK
16	HALYATUL MAHFUDOH	1	1	0.667	0.75	0.800	0	TIDAK
17	SELVIANA PUTRI	1	0.600	0.333	1	0	0	TIDAK
18	SINTA	1	0.600	0.333	1	0	0	TIDAK
19	TRIA ALPIANA	1	0.800	0.333	1	0	0	TIDAK
20	WULAN KRISNAWATI	1	0.9000	0.333	1	0	0	TIDAK

Dari tabel penentuan kelas mayoritas di atas, maka disimpulkan data *testing* pertama atas nama Anggit Bintoro termasuk ke dalam kategori siswa yang TIDAK beasiswa PIP berdasarkan hasil dari penerapan algoritma *KNN*. Perhitungan data *testing* selanjutnya sama seperti perhitungan di atas. Berikut ini hasil dari perhitungan yang telah dilakukan *RapidMiner* versi 10.1, yaitu:

4.5. Evaluasi dan Validasi

Pada penelitian ini *confusion matrix* digunakan untuk pengujian validitas sebagai *performance* yang akan menghasilkan nilai *accuracy*, dalam perhitungannya dapat menggunakan rumus sebagai berikut ini:

$$Accuracy = \frac{FN + FP}{FN + FP + TP + TN} \times 100\% \quad (4.6)$$

Tabel IV.10. Performance Vector

	True TIDAK	True YA
Pred. TIDAK	226	50
Pred. YA	28	36

Pada tabel di atas menunjukkan hasil *performance* vector, maka dapat disimpulkan dengan menggunakan algoritma *K-Nearest Neighbor* dalam 340 data, sebanyak 226 siswa tidak menerima beasiswa PIP dan 36 siswa menerima beasiswa PIP. Kemudian 28 siswa tidak menerima diklasifikasikan menerima dan 50 siswa menerima diklasifikasikan tidak menerima. Berikut ini nilai akurasi yang diperoleh, yaitu:

$$Accuracy = \frac{226 + 36}{226 + 36 + 28 + 50} \times 100\% = 77.06\%$$

Dapat disimpulkan dari hasil uji validitas di atas dengan tingkat akurasi menggunakan algoritma *K-Neares Neighbo* adalah 77.06%.

KESIMPULAN DAN SARAN

Maka dapat disimpulkan bahwa dalam pengolahan data penentuan calon penerima beasiswa PIP dengan adanya penerapan *data mining* pada sistem informasi maka dapat digunakan untuk pengolahan data peserta penerima beasiswa PIP. Karena akan menghasilkan nilai akurasi yang lebih akurat sehingga penerimaan beasiswa PIP tepat sasaran. Metode *K-Nearest Neighbor* jika diterapkan pada seleksi penerimaan beasiswa PIP di SMK Karya Medika belum mencapai hasil yang maksimal. Karena nilai akurasi yang diperoleh lebih sedikit dari hasil jurnal yang dijadikan acuan dalam penelitian ini.

Maka dalam pengolahan data penentuan calon penerima beasiswa PIP penulis memberikan saran untuk penelitian selanjutnya diharapkan menambahkan atribut lebih lengkap dalam menentukan penerima beasiswa PIP. Untuk penelitian selanjutnya dapat dikembangkan menggunakan metode *Neural Network*, *Desicision Tree*, *Naive Bayes* dan lain sebagainya. Untuk penelitian selanjutnya dapat menggunakan pengujian lain seperti *10-Fold Cross Validation*.

DAFTAR REFERENSI

- Andini, A. D., Arifin, T., Sanjaya, A. R., Sanjaya, A. R., Coefficient, S., & Pasien, P. (2020). *Implementasi Algoritma K-Medoids Untuk Klasterisasi Data Penyakit Pasien*. 2(2), 128–138.
- Arifin, T., & Ariesta, D. (2019). Prediksi Penyakit Ginjal Kronis Menggunakan Algoritma Naive Bayes Classifier Berbasis Particle Swarm Optimization. *Jurnal Tekno Insentif*, 13(1), 26–30. <https://doi.org/10.36787/Jti.V13i1.97>
- Arifin, T., & Herliana, A. (2020). Optimizing Decision Tree Using Particle Swarm Optimization To Identify Eye Diseases Based On Texture Analysis. *Jurnal Teknologi Dan Sistem Komputer*, 8(1), 59–63. <https://doi.org/10.14710/Jtsiskom.8.1.2020.59-63>
- Homepage, J., Cholil, S. R., Handayani, T., Prathivi, R., & Ardianita, T. (2021). Ijcit (Indonesian Journal On Computer And Information Technology) Implementasi Algoritma Klasifikasi K-Nearest Neighbor (Knn) Untuk Klasifikasi Seleksi Penerima Beasiswa. *Ijcit (Indonesian Journal On Computer And Information Technology)*, 6(2), 118–127.
- Iriane, R., & Nurfaizah, N. (2023). Klik: Kajian Ilmiah Informatika Dan Komputer Penerapan Data Mining Untuk Prediksi Penjualan Produk Pangan Hewan Menggunakan Metode K-Nearest Neighbor. *Media Online*, 3(5), 509–515. <https://djournals.com/klik>

- Karepesina, F., & Zahrotun, L. (2023). Penerapan Data Mining Untuk Penentuan Penerima Beasiswa Dengan Metode K-Nearest Neighbor (K-Nn). *Techno (Jurnal Fakultas Teknik, Universitas Muhammadiyah Purwokerto)*, 24(1), 1. <https://doi.org/10.30595/Techno.V24i1.9084>
- Koda, A., Rahayu, P., Pratama, A., Rafly, A., & Kaslani. (2020). Penentuan Bonus Karyawan Dengan Menggunakan Algoritma K-Nearest Neighbor. *Kopertip: Jurnal Ilmiah Manajemen Informatika Dan Komputer*, 4(1), 14–20. <https://doi.org/10.32485/Kopertip.V4i1.115>
- Mulyani, N. K. (2020). *Penerapan Algoritma K-Nearest Naighbor Dalam Menentukan Penerima Beasiswa Peningkatan Prestasi Akademik Di Unisnu Jepara*. <https://doi.org/https://eprints.unisnu.ac.id/id/eprint/658/>
- Nata, A., & Suparmadi, S. (2022). Analisis Sistem Pendukung Keputusan Dengan Model Klasifikasi Berbasis Machine Learning Dalam Penentuan Penerima Program Indonesia Pintar. *Journal Of Science And Social Research*, 5(3), 697. <https://doi.org/10.54314/Jssr.V5i3.1041>
- Noviana, D., Susanti, Y., & Susanto, I. (2019). Analisis Rekomendasi Penerima Beasiswa Menggunakan Algoritma K-Nearest Neighbor (K-Nn) Dan Algoritma C4.5. *Seminar Nasional Penelitian Pendidikan Matematika (Snp2m) 2019 Umt*, 79–87.
- Pauji, A., Aisyah, S., Surip, A., Saputra, R., & Ali, I. (2020). Implementasi Algoritma K-Nearest Neighbor Dalam Menentukan Penerima Bantuan Langsung Tunai. *Kopertip: Jurnal Ilmiah Manajemen Informatika Dan Komputer*, 4(1), 21–27. <https://doi.org/10.32485/Kopertip.V4i1.114>
- Putry, N. M., & Sari, B. N. (2022). Komparasi Algoritma Knn Dan Naïve Bayes Untuk Klasifikasi Diagnosis Penyakit Diabetes Mellitus. *Evolusi: Jurnal Sains Dan Manajemen*, 10(1). <https://doi.org/10.31294/Evolusi.V10i1.12514>
- Resmiati, R., & Arifin, T. (2021). Klasifikasi Pasien Kanker Payudara Menggunakan Metode Support Vector Machine Dengan Backward Elimination. *Sistemasi*, 10(2), 381. <https://doi.org/10.32520/Stmsi.V10i2.1238>
- Soleha, R. A., Wijaya, H. O. L., & Daulay, N. K. (2023). Klasifikasi Kemampuan Ekonomi Calon Siswa Baru Dengan Metode K-Nearest Neighbor Pada Sma Negeri 1 Musi Rawas. *Jurasik (Jurnal Riset Sistem Informasi Dan Teknik Informatika)*, 8(1), 62–69. <http://tunasbangsa.ac.id/ejurnal/index.php/jurasik/article/view/542>
- Zaman, B., Rifai, A., & Hanif, M. B. (2021). Komparasi Metode Klasifikasi Batik Menggunakan Neural Network Dan K-Nearest Neighbor Berbasis Ekstraksi Fitur Tekstur. *Journal Of Information Systems And Informatics*, 3(4), 582–595. <https://doi.org/10.51519/Journalisi.V3i4.213>
- Zulfallah, F. H. (2022). *Implementasi Algoritma Knn Dalam Mengukur Ketepatan Kelulusan Mahasiswa Uin Syarif Hidayatullah Jakarta*.