

STUDI KOMPARASI ALGORITMA NAÏVE BAYES DAN K-NN UNTUK KLASIFIKASI PENERIMAAN BEASISWA DI MI AL – ISLAMIYAH KARANGSAWAH

¹⁾Muslim Hidayat, ²⁾Afif Nazmi Fuadi, ³⁾Dimas Prasetyo Utomo, ⁴⁾Erna Dwi Astuti, ⁵⁾Dian Asmarajati

^{1,2,3,4,5)}Universitas Sains Al-Qur'an

¹⁾muslim_h@unsiq.ac.id, ²⁾afifnzm@gmail.com, ³⁾vikiran.dpu@gmail.com, ⁴⁾ernada@unsiq.ac.id,
⁵⁾asmarajati@unsiq.ac.id

INFO ARTIKEL

Riwayat Artikel :

Diterima : 23 November 2023

Disetujui : 29 November 2023

Kata Kunci :

Beasiswa, Naïve Bayes, K-Nearest Neighbor, Studi Komparasi

ABSTRAK

Pemberian beasiswa dilakukan agar para siswa dapat tetap melanjutkan sekolah, dalam menyeleksi siswa parameter yang digunakan terdiri dari jarak, tanggungan, pekerjaan orang tua, pendapatan orang tua, kelengkapan keluarga dan kelayakan. Dikarenakan belum ada metode untuk menentukan penerima beasiswa maka sering salah sasaran dalam memberikan beasiswa. Oleh karena itu diperlukan klasifikasi penerima beasiswa yang tepat dan akurat. Salah satunya data mining dengan metode deskriptif analitis. Bisa dikatakan penelitian deskriptif analitis mengambil masalah atau memperhatikan masalah- masalah yang ada saat penelitian kemudian diolah untuk mendapatkan sebuah kesimpulan. Berdasarkan hasil analisis dan pembahasan studi komparasi algoritma naïve bayes dan K-NN untuk klasifikasi penerimaan beasiswa di MI AL-Islamiyah, dari 186 data siswa yang terdiri dari 150 data *training* dan 36 data *testing* diperoleh Hasil klasifikasi dengan Naïve Bayes dan K-Nearest Neighbor diperoleh masing-masing 91,67% dan 75,00%. Berdasarkan nilai akurasi yang diperoleh dari dua algoritma tersebut, maka akurasi termasuk *excellent classification*., dan algoritma Naïve Bayes lebih baik dalam klasifikasi penerimaan beasiswa dibandingkan algoritma K-Nearest Neighbor.

ARTICLE INFO

Article History :

Received : Nov 23, 2023

Accepted : Nov 29, 2023

Keywords:

Scholarships, Naïve Bayes, K-Nearest Neighbor, Comparative Studies

ABSTRACT

Scholarships are given so that students can continue their studies. In selecting students, the parameters used consist of distance, dependents, parents' occupation, parents' income, family completeness and eligibility. Because there is no method for determining scholarship recipients, it is often the wrong target in awarding scholarships. Therefore, a precise and accurate classification of scholarship recipients is needed. One of them is data mining with analytical descriptive method. It can be said that descriptive analytical research takes the problem or pays attention to the problems that exist when the research is then processed to get a conclusion. Based on the results of the analysis and discussion of the comparative study of the naïve Bayes and K-NN algorithms for the classification of scholarship recipients

at MI AL-Islamiah, from 186 student data consisting of 150 training data and 36 testing data obtained Classification results with Naïve Bayes and K-Nearest Neighbor obtained respectively 91.67% and 75.00%. Based on the accuracy values obtained from the two algorithms, the accuracy is included in the excellent classification, and the Naïve Bayes algorithm is better at classifying scholarship recipients than the K-Nearest Neighbor algorithm.

1. PENDAHULUAN

Beasiswa adalah hibah yang diberikan kepada pelajar atau mahasiswa sebagai bantuan belajar. Beasiswa diharapkan memungkinkan siswa untuk berpartisipasi dalam kegiatan belajar dengan biaya yang lebih murah. Dengan bantuan hibah ini, jumlah putus sekolah karena masalah keuangan dapat dikurangi. Alasan ekonomi adalah salah satu alasan khusus mengapa orang miskin tidak menyekolahkan anaknya sampai mereka menyelesaikan program pemerintah. Program beasiswa dapat menekan angka putus sekolah, mendorong siswa miskin untuk kembali bersekolah dan mendapatkan pendidikan yang layak (Ikhsan, 2023).

MI Karangsawah adalah sekolah swasta di Kecamatan Tonjong, Kabupaten Brebes. MI Karangsawah merupakan salah satu sekolah yang memberikan beasiswa miskin kepada siswanya. Murid MI Karangsawah berasal dari seluruh Karangsawah. Bentuk hibah di sekolah-sekolah ini harus memfasilitasi perlengkapan sekolah bagi para siswa.

Adanya informasi ketidak merataan beasiswa di mi al-islamiyah, maka perlu dilakukan klasifikasi data dengan tujuan agar di dapatkan suatu informasi atau data mengenai penerimaan beasiswa di mi al-islamiyah yang dapat digunakan oleh pihak sekolah ataupun pemerintah untuk dapat merancang strategi dalam meningkatkan kesejahteraan siswa, dan upaya sekolah atau pemerintah dalam mengurangi angka ketidak merataan dapat berjalan tepat sasaran. Klasifikasi tersebut menggunakan algoritma Naive Bayes dan K-Nearest Neighbor yang merupakan algoritma klasifikasi pada teknik data mining. Data mining merupakan proses mengerjakan satu atau lebih teknik pembelajaran komputer (machine learning) untuk menganalisis pengetahuan secara otomatis (Amalia, 2018).

Pengklasifikasi data siswa ini memegang peranan penting karena banyak hal yang harus

diperhatikan dalam membuat rekomendasi beasiswa seperti parameter yang digunakan yaitu jarak rumah siswa dari sekolah, tanggungan orang tua, gaji orang tua, pekerjaan dan status perkawinan. dari orang tua. (orang tua penuh/yatim piatu). Oleh karena itu, sekolah memerlukan cara yang efektif dan efisien dalam menggunakan waktu dan tenaga untuk meningkatkan pelayanan. Klasifikasi tersebut menggunakan algoritma Naive Bayes dan K-Nearest Neighbor yang merupakan algoritma klasifikasi yang digunakan dalam teknik data mining. Penambahan data menggunakan satu atau lebih teknik pembelajaran komputer (pembelajaran mesin) untuk menganalisis data secara otomatis (Sidik, 2020).

Algoritma Naive Bayes adalah klasifikasi yang dikemukakan oleh peneliti Inggris, yaitu Thomas Bayes, yang menggunakan metode probabilistik dan statistik untuk memprediksi kemungkinan masa depan berdasarkan pengalaman masa lalu (Pebdika, 2023). Algoritma KNearest Neighbor adalah metode klasifikasi data tergantung pada tipe data dari dataset yang ada. K-nearest neighbor mengklasifikasikan data berdasarkan nilai k yang diberikan (Silalahi, 2023).

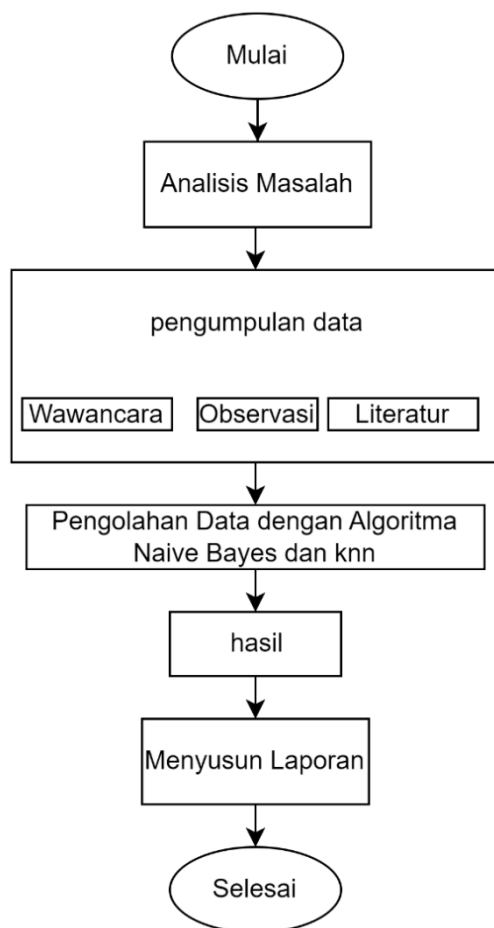
Kelebihan dari algoritma Naive Bayes antara lain pengolahan data kuantitatif dan diskrit, efisiensi dan efisiensi spasial, serta kemampuan untuk menggunakannya pada atribut yang tidak relevan. Kelebihan dari algoritma K-Nearest Neighbor antara lain adalah ketangguhannya terhadap data pelatihan dengan noise tinggi dan efisiensinya dalam pelatihan. Karena datanya besar, kelasnya kuat menurut ruang pencarian misalnya dan tidak perlu linier, yang dalam komputasi efisien dalam skala kecil. Presisi dari algoritma Naive Bayes yaitu nilai presisi yang diperoleh sebesar 80%, presisi sebesar 82,82%, dan recall sebesar 92,47% sedikit lebih tinggi dari algoritma KNN dengan

nilai presisi sebesar 78,97% dan akurasi sebesar 82,61%. , dan kembali 91,10% (Nabila, 2022).

Penelitian ini bertujuan untuk membandingkan hasil akurasi algoritma Naive Bayes dan K-Nearest Neighbor dalam klasifikasi di MI Karangasawah. Sehingga instansi terkait diharapkan dapat mempercepat proses dalam membuat keputusan calon penerima beasiswa.

2. METODE

Kerangka kerja penelitian merupakan tahapan-tahapan yang akan dilakukan dalam rangka penyelesaian masalah yang akan dibahas (Cahya, 2022). Adapaun kerangka kerja dari penelitian ini dapat dilihat pada gambar 1.



Gambar 1. Kerangka Kerja

Lokasi penelitian ini berada MI AL-ISLAMIAH Karangasawah yang beralamat di Dusun Karang Asem, Desa Kutamendala, Kec. Tonjong , Kab. Brebes, Jawa Tengah. Untuk menganalisis hasil akurasi algoritma Naive Bayes dan K-Nearest Neighbor dalam klasifikasi di MI Karangasawah, perlu adanya data yang

mendukung penelitian yaitu; Data primer adalah data yang diperoleh secara langsung dilapangan melalui survey diantaranya pendapatan orang tua, pekerjaan orang tua, tanggungan keluarga, jarak rumah, dan kelengkapan keluarga. Data *skunder* merupakan data yang telah tersedia. Data yang akan digunakan dalam penelitian penentuan penerima bantuan beasiswa adalah data *primer*.

3. HASIL DAN PEMBAHASAN

3.1. Pengujian Naïve Bayes

Naive Bayes adalah pendekatan pembelajaran mesin yang dkemukakan oleh Thomas Bayes yang memanfaatkan probabilitas dan perhitungan statistik untuk memprediksi probabilitas masa depan berdasarkan pengalaman sebelumnya. Persamaan menggambarkan perhitungan dengan cara ini:

$$P(H|X) = \frac{P(X|H).P(H)}{p(x)}$$

Di mana:

X : Data dengan class yang belum diketahui

H : Hipotesis data merupakan suatu class spesifik

P(H|X): Probabilitas hipotesis H berdasar kondisi X (posteriori probabilitas)

P(H) : Probabilitas hipotesis H (prior probabilitas)

P(X|H): Probabilitas X berdasarkan kondisi pada hipotesis H

P(X) : Probabilitas X

a. Data selection

Pada proses ini adalah proses pemilihan data dari data-data yang harus dilakukan sebelum menuju tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi ini nantinya akan digunakan dalam proses *Data Mining*. Data yang digunakan dalam penelitian ini yaitu data yang diperoleh melalui pengumpulan langsung dan wawancara dengan guru bimbingan konseling (BK) MI AL – ISLAMIAH Karangasawah.

Tabel 1. Data Awal

NO	Nama	JK	Jarak	Penghasilan	T	YT	Pekerjaan	KELAYAKAN
1	ADRY PRADANA P	L	2 km	Rp1,000,000	2	-	Buruh	LAYAK
2	AISYAH NUR L	P	1 km	Rp2,000,000	3	-	pedagang	TIDAK LAYAK
3	GHANI YAZID K	L	1 km	Rp1,800,000	1	-	wiraswasta	TIDAK LAYAK
4	GISKA SYAKILA R	P	2,8 km	Rp2,000,000	2	-	petani	TIDAK LAYAK
5	KENSO PUTRA W	L	1 km	Rp500,000	1	Ytm	Buruh	LAYAK
6	M. HENDRI NOVAL P	L	1 km	Rp1,200,000	2	-	petani	TIDAK LAYAK
7	M.NAJIH ASSUVI	L	< 1 km	Rp1,500,000	1	-	petani	TIDAK LAYAK
8	M RIZKI NUR A	L	1 km	Rp1,500,000	3	-	petani	TIDAK LAYAK
9	M. IBRAHIM	L	2 km	Rp600,000	3	-	Buruh	LAYAK
10	MAFA AL GHONIYU	L	1 km	Rp2,000,000	2	-	pedagang	TIDAK LAYAK
11	M. FAIQ ABBASY	L	5 km	Rp1,500,000	2	-	wiraswasta	TIDAK LAYAK
12	RAFA DWI P	L	< 1 km	Rp1,700,000	1	-	wiraswasta	TIDAK LAYAK
13	RANIYA REZKI N	P	3 km	Rp1,500,000	3	-	Buruh	LAYAK
14	SYAFIRA AYU L	P	2 km	Rp1,400,000	2	-	wiraswasta	TIDAK LAYAK
15	YUNA KALISYA P	P	1 km	Rp1,200,000	2	-	petani	TIDAK LAYAK
16	ZAHIRA MAHYA A.S	P	2 km	Rp1,000,000	3	-	Buruh	LAYAK
17	ASYA MAULIDIYAH	P	3 km	Rp1,300,000	1	-	petani	TIDAK LAYAK
18	ANINDIA RISTIANA	P	2 km	Rp1,000,000	1	-	Buruh	LAYAK
19	ADINDA ZAHIRANY	P	2 km	Rp1,130,000	2	-	petani	TIDAK LAYAK
20	SALSA RIZQY R	P	2,5 km	Rp960,000	2	Ytm	Buruh	LAYAK
21	NAILA AFIFA A	P	< 1 km	Rp900,000	2	-	Buruh	LAYAK
...	...	...	...	...	...	...	...	...
150	SHERLLY MARSYA	P	1 km	Rp1,300,000	3	-	wiraswasta	LAYAK

Semua atribut yang ada pada data seleksi beasiswa diantaranya Nama, Jenis Kelamin, Jarak rumah ke sekolah, Penghasilan orang tua, Tanggungan orang tua, kelengkapan keluarga (yatim/piatu), pekerjaan orang tua dan kelayakan hanya untuk proses *Knowledge Discovery in Database*

b. *Preprocessing*

Tahap *preprocessing* harus diselesaikan sebelum proses data mining. Tahap ini melibatkan proses menggabungkan data dari beberapa database. Selain itu, pembersihan data dilakukan menghasilkan dataset yang cocok untuk penambahan data.

c. *Transformation*

Tahapan dimana data yang telah dipilih diubah agar cocok untuk proses data mining adalah yang satu ini. Dalam KDD, proses transformasi merupakan proses kreatif yang sangat dipengaruhi oleh jenis atau pola informasi yang akan dicari dalam database. Kami mengonversi beberapa data dalam bentuk angka menjadi berbagai interval, termasuk jarak, selama proses ini.

Tabel 2. Data Hasil Transformasi

N o	Jar ak	Pendap atan	T	Y/ P	Pekerj aan	Kelaya kan
1	2	2	2	1	1	Layak
2	2	3	3	1	3	Tidak Layak
3	2	3	1	1	4	Tidak Layak

4	3	3	2	1	2	Tidak Layak
5	2	1	1	3	1	Layak
6	2	2	2	1	2	Tidak Layak
7	1	2	1	1	2	Tidak Layak
8	2	2	3	1	2	Tidak Layak
9	2	1	3	1	1	Layak
10	2	3	2	1	3	Tidak Layak
11	4	2	2	1	4	Tidak Layak
12	1	3	1	1	4	Tidak Layak
13	3	2	3	1	1	Layak
14	2	2	2	1	4	Tidak Layak
15	2	2	2	1	2	Tidak Layak
16	2	2	3	1	1	Layak
17	3	2	1	1	2	Tidak Layak
18	3	5	2	3	1	Layak
19	2	1	2	1	1	Layak
...	...	...	...	...	...	...
150	2	2	3	1	4	Layak

d. Analisis Algoritma *Naïve Bayes*

Setelah menghitung probabilitas setiap atribut selanjutnya dilakukan pengujian pada data testing yang berjumlah 36 data. Untuk mengujinya kita masukkan data training ke rumusan *Naïve bayes* dimana kriteria diambil dari data kemudian kita hitung setiap kriteria tersebut dengan probabilitas sebelumnya. Selanjutnya dilakukan perhitungan:

1. Jarak = 3, pendapatan = 3, tanggungan = 1, kelengkapan keluarga = 1, dan pekerjaan = 1
 $P(\text{layak}|X) = (0,1973 \times 0,0268 \times 0,0704 \times 0,8732 \times 0,619) \times 0,4733$
 $P(\text{layak}|X) = \underline{0,000099879}$
 $P(\text{tidak layak}|X) = (0,0759 \times 0,3164 \times 0,2658 \times 0,9873 \times 0,1646) \times 0,5267$
 $P(\text{tidak layak}|X) = \underline{0,00054636}$
 $P(\text{tidak layak}|X)$ adalah tertinggi yaitu sebesar 0,00054636 sehingga prediksinya adalah kelayakan TIDAK LAYAK.
2. Jarak = 3, pendapatan = 3, tanggungan = 3, kelengkapan keluarga = 1, dan pekerjaan = 1
 $P(\text{layak}|X) = (0,1973 \times 0,0268 \times 0,2676 \times 0,8732 \times 0,619) \times 0,4733$
 $P(\text{layak}|X) = \underline{0,0003796569}$
 $P(\text{tidak layak}|X) = (0,0759 \times 0,3164 \times 0,2532 \times 0,9873 \times 0,1646) \times 0,5267$
 $P(\text{tidak layak}|X) = \underline{0,00052045}$
 $P(\text{tidak layak}|X)$ adalah tertinggi yaitu sebesar 0,00052045 sehingga prediksinya adalah kelayakan TIDAK LAYAK.
3. Jarak = 3, pendapatan = 3, tanggungan = 3, kelengkapan keluarga = 1, dan pekerjaan = 3
 $P(\text{layak}|X) = (0,1973 \times 0,0268 \times 0,2676 \times 0,8732 \times 0,423) \times 0,4733$
 $P(\text{layak}|X) = \underline{0,00323030}$
 $P(\text{tidak layak}|X) = (0,0759 \times 0,3164 \times 0,2532 \times 0,9873 \times 0,1013) \times 0,5267$
 $P(\text{tidak layak}|X) = \underline{0,0032030}$
 $P(\text{tidak layak}|X)$ adalah tertinggi yaitu sebesar 0,0032030 sehingga prediksinya adalah kelayakan TIDAK LAYAK.
4. Jarak = 3, pendapatan = 1, tanggungan = 3, kelengkapan keluarga = 1, dan pekerjaan = 1
 $P(\text{layak}|X) = (0,1972 \times 0,4211 \times 0,2676 \times 0,8732 \times 0,619) \times 0,4733$

- $P(\text{layak}|X) = \underline{0,0060788}$
 $P(\text{tidak layak}|X) = (0,0759 \times 0,0127 \times 0,2532 \times 0,9873 \times 0,1646) \times 0,5267$
 $P(\text{tidak layak}|X) = \underline{0,0002089}$
 $P(\text{layak}|X)$ adalah tertinggi yaitu sebesar 0,0060788 sehingga prediksinya adalah kelayakan LAYAK.
5. Jarak = 3, pendapatan = 2, tanggungan = 2, kelengkapan keluarga = 1, dan pekerjaan = 1
 $P(\text{layak}|X) = (0,1972 \times 0,5263 \times 0,3662 \times 0,8732 \times 0,619) \times 0,4733$
 $P(\text{layak}|X) = \underline{0,01039683}$
 $P(\text{tidak layak}|X) = (0,0759 \times 0,4684 \times 0,4430 \times 0,9873 \times 0,1646) \times 0,5267$
 $P(\text{tidak layak}|X) = \underline{0,00131541}$
 $P(\text{layak}|X)$ adalah tertinggi yaitu sebesar 0,010396837 sehingga prediksinya adalah kelayakan LAYAK.
6. Jarak = 1, pendapatan = 2, tanggungan = 1, kelengkapan keluarga = 1, dan pekerjaan = 1
 $P(\text{layak}|X) = (0,2394 \times 0,5263 \times 0,0704 \times 0,8732 \times 0,619) \times 0,4733$
 $P(\text{layak}|X) = \underline{0,00242646}$
 $P(\text{tidak layak}|X) = (0,03418 \times 0,4684 \times 0,2658 \times 0,987 \times 0,1646) \times 0,5267$
 $P(\text{tidak layak}|X) = \underline{0,00355421}$
 $P(\text{tidak layak}|X)$ adalah tertinggi yaitu sebesar 0,00355421 sehingga prediksinya adalah kelayakan TIDAK LAYAK.
7. Jarak = 1, pendapatan = 3, tanggungan = 3, kelengkapan keluarga = 1, dan pekerjaan = 4
 $P(\text{layak}|X) = (0,2394 \times 0,0263 \times 0,2676 \times 0,8732 \times 0,1268) \times 0,4733$
 $P(\text{layak}|X) = \underline{0,000088294}$
 $P(\text{tidak layak}|X) = (0,03418 \times 0,3164 \times 0,253 \times 0,9634 \times 0,3924) \times 0,5267$
 $P(\text{tidak layak}|X) = \underline{0,005452192}$

$P(\text{tidak layak}|\mathbf{X})$ adalah tertinggi yaitu sebesar 0,005452192 sehingga prediksinya adalah kelayakan TIDAK LAYAK.

8. Jarak = 2, pendapatan = 3, tanggungan = 1, kelengkapan keluarga = 1, dan pekerjaan = 4
 $P(\text{layak}|\mathbf{X}) = (0,5352*0,0263*0,0704*0,8732*0,1268)*0,4733$
 $P(\text{layak}|\mathbf{X}) = \underline{0,0000519294}$
 $P(\text{tidak layak}|\mathbf{X}) = (0,5696*0,3164*0,2658*0,9634*0,3924)*0,5267$
 $P(\text{tidak layak}|\mathbf{X}) = \underline{0,009538067}$

$P(\text{tidak layak}|\mathbf{X})$ adalah tertinggi yaitu sebesar 0,009538067 sehingga prediksinya adalah kelayakan TIDAK LAYAK.

9. Jarak = 2, pendapatan = 1, tanggungan = 3, kelengkapan keluarga = 1, dan pekerjaan = 1
 $P(\text{layak}|\mathbf{X}) = (0,5352*0,4211*0,2676*0,8732*0,6619)*0,4733$
 $P(\text{layak}|\mathbf{X}) = \underline{0,016497953}$
 $P(\text{tidak layak}|\mathbf{X}) = (0,5696*0,0127*0,2632*0,9634*0,1646)*0,5267$
 $P(\text{tidak layak}|\mathbf{X}) = \underline{0,000152980}$

$P(\text{layak}|\mathbf{X})$ adalah tertinggi yaitu sebesar 0,016497953 sehingga prediksinya adalah kelayakan LAYAK.

10. Jarak = 2, pendapatan = 1, tanggungan = 4, kelengkapan keluarga = 3, dan pekerjaan = 1
 $P(\text{layak}|\mathbf{X}) = (0,5352*0,4211*0,2958*0,0986*0,6619)*0,4733$
 $P(\text{layak}|\mathbf{X}) = \underline{0,0002059232}$
 $P(\text{tidak layak}|\mathbf{X}) = (0,5696*0,0127*0,2632*0,9634*0,1646)*0,5267$
 $P(\text{tidak layak}|\mathbf{X}) = \underline{0,00000029074}$

$P(\text{layak}|\mathbf{X})$ adalah tertinggi yaitu sebesar 0,0002059232 sehingga prediksinya adalah kelayakan LAYAK.

Kemudian dengan cara yang sama semua data di data *testing* dimasukkan ke rumus

Naïve Bayes seperti di atas dan dihasilkan data seperti berikut ini:

Tabel 3. Hasil Data Testing

N o	Jar ak	Pendap atan	T	Y/ P	Pekerj aan	Kelaya kan
1	3	3	1	1	1	Tidak Layak
2	3	3	3	1	1	Tidak Layak
3	3	3	3	1	3	Tidak Layak
4	3	1	3	1	1	Layak
5	3	2	2	1	1	Layak
6	1	2	1	1	1	Tidak Layak
7	1	3	3	1	4	Tidak Layak
8	2	3	1	1	4	Tidak Layak
9	2	1	3	1	1	Layak
10	2	1	4	3	1	Layak
11	2	3	2	1	4	Tidak Layak
12	2	3	1	1	3	Tidak Layak
13	2	2	3	1	1	Layak
14	2	2	4	1	2	Layak
15	2	2	3	1	4	Tidak Layak
16	1	2	2	1	1	Tidak Layak
17	3	3	2	1	1	Tidak Layak
18	2	1	4	3	4	Layak
19	1	3	3	1	1	Tidak Layak
20	2	1	2	1	1	Layak
21	2	1	2	1	1	Layak
22	1	3	3	1	1	Tidak Layak
23	1	2	2	1	4	Layak

24	2	1	1	1	1	Tidak Layak
25	2	3	2	1	1	Tidak Layak
26	2	1	2	1	2	Tidak Layak
27	1	2	3	1	1	Tidak Layak
28	3	3	1	1	1	Tidak Layak
29	2	2	3	1	4	Layak
30	2	3	4	1	2	Tidak Layak
31	2	3	1	1	3	Tidak Layak
32	2	2	3	1	4	Layak
33	3	3	2	1	1	Tidak Layak
34	3	2	3	1	1	Layak
35	1	2	2	1	2	Tidak Layak
36	2	3	1	1	4	Tidak Layak

3.2. Pengujian dengan K-Nearest Neighbor

Proses pengklasifikasian dengan algoritma *K-Nearest Neighbor* juga dibagi menjadi 2, Yaitu proses *training* dan *testing*. Tabel 4.6 adalah data *training* yang terdiri dari 150. Sebelum memulai perhitungan *Euclidian Distance* dilakukan, nilai K ditentukan, pada penelitian ini dipilih nilai K=11.

Mencari nilai K:

$$K = \sqrt{n}$$

Dimana n = 150, maka

$$K = \sqrt{150} = 12,24744 \text{ 87 dibulatkan menjadi 12,}$$

maka nilai K = 12

Dan karna algoritma K-NN ini jika data yang di miliki peneliti genap maka peneliti mengambil nilai K ganjil yang lebih dekat dari hasil di atas.

Tabel 5. Nilai Jarak

No	Jarak	Penghasi lan	T	Y / P	Peke rjaan	Kela yaka n	Eucli edian
----	-------	-----------------	---	----------	---------------	-------------------	----------------

							<i>Distance</i>
1	2	2	2	1	1	Laya k	3,74 1657
2	2	3	3	1	3	Tida k Laya k	5,65 6854
3	2	3	1	1	4	Tida k Laya k	5,56 7764
4	3	3	2	1	2	Tida k Laya k	5,19 6152
5	2	1	1	3	1	Laya k	4
6	2	2	2	1	2	Tida k Laya k	4,12 3106
7	1	2	1	1	2	Tida k Laya k	3,31 6625
8	2	2	3	1	2	Tida k Laya k	4,69 0416
9	2	1	3	1	1	Laya k	4
10	2	3	2	1	3	Tida k Laya k	5,19 6152
11	4	2	2	1	4	Tida k Laya k	6,40 3124
12	1	3	1	1	4	Tida k Laya k	5,29 1503
13	3	2	3	1	1	Laya k	4,89 8979
14	2	2	2	1	4	Tida k Laya k	5,38 5165

1 5	2	2	2	1	2	Tida k Laya k	4,12 3106
1 6	2	2	3	1	1	Laya k	4,35 8899
1 7	3	2	1	1	2	Tida k Laya k	4,35 8899
1 8	2	2	1	1	1	Laya k	3,31 6625
1 9	2	2	2	1	2	Tida k Laya k	4,12 3106
...	...		...	...			
1 5 0	4	3	2	1	44	Laya k	6,78 233

Perhitungan Eucliediance Distance di urutkan dari yang terkecil menghasilkan pada tabel 4.17. berdasarkan tabel 4.17 dengan nilai K =11 maka data *testing* masuk kedalam kategori Tidak Layak mendapatkan beasiswa.

3.3. Pengujian Naive Bayes dengan Rapidminer

Pengujian di software Rapidminer, sebelum data *training* dan data *testing* disimpan dalam bentuk .exe atau excel agar bisa di import ke aplikasi Rapidminer. Pada hasil pengujian dengan aplikasi Rapidminer diperoleh hasil Accuracy sebesar 91,67%. Pengujian diatas dibuktikan dengan rumus menghitung akurasi sebagai berikut:

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

$$Accuracy = \frac{22+11}{(22+11)+(1+2)} \times 100\%$$

$$Accuracy = \frac{33}{36} \times 100\%$$

$$Accuracy = 0,916667 \times 100\% = 91,67\%$$

Hasil akurasi pengujian yaitu 91,67%

3.4. Pengujian K-Nearest Neighbor dengan Rapidminer

Tahapan dalam pengujian K-Nearest Neighbor sama seperti pada *Naive Bayes*, perhitungan manual pada K-Nearest

Neighbor nilai K ditentukan terlebih dahulu, yaitu 12. Gambar 2 menunjukkan hasil klasifikasi algoritma K-Nearest Neighbor dengan akurasi yang diperoleh yaitu sebesar 75,00%.

accuracy: 75.00%

	true TIDAK LAYAK	true LAYAK	class precision
pred. TIDAK LAYAK	16	2	88.89%
pred. LAYAK	7	11	61.11%
class recall	69.57%	84.62%	

Gambar 2. Hasil Klasifikasi K-Nearest Neighbor

Pengujian diatas dibuktikan dengan rumus menghitung akurasi sebagai berikut:

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

$$Accuracy = \frac{16+11}{(16+11)+(7+2)} \times 100\%$$

$$Accuracy = \frac{27}{36} \times 100\%$$

$$Accuracy = 0,75 \times 100\% = 75,00\%$$

Hasil akurasi pengujian yaitu 75,00%

3.5. Komparasi Akurasi Naive Bayes dan K-Nearest Neighbor

Berdasarkan pengujian akurasi *Naive Bayes* dan K-Nearest Neighbor menggunakan Rapidminer, diperoleh akurasi *Naive Bayes* yaitu 91,67%, sedangkan pada K-Nearest Neighbor diperoleh akurasi 75,00% Hasil akurasi tersebut menunjukkan bahwa untuk mengklasifikasi keluarga miskin, algoritma *Naive Bayes* lebih baik dibandingkan dengan algoritma K-Nearest Neighbor.

4. PENUTUP

4.1. Kesimpulan

Analisis dan pembahasan klasifikasi penerima beasiswa di MI Al-Islamyah Karangsawah dengan menggunakan algoritma *Naive Bayes* mendapatkan hasil dengan akurasi yang cukup tinggi yaitu 91,67 persen, sedangkan data yang sama memberikan hasil dengan akurasi sebesar 75,00 persen menggunakan algoritma K-NN. Dapat disimpulkan bahwa Algoritma *Naive Bayes* lebih unggul dari Algoritma K-Nearest Neighbor dalam hal presisi sehingga cocok digunakan dalam klasifikasi penerima beasiswa. Hasilnya, proses pemberian

beasiswa dapat diselesaikan dengan lebih cepat dan akurat.

4.2. Saran

Adapun saran yang disampaikan berdasarkan penelitian yang dilakukan untuk peneliti selanjutnya adalah sebagai berikut:

1. Penelitian ini perlu dikembangkan dengan menambahkan beberapa atribut lain yang memengaruhi penentuan pemberian beasiswa salah satunya menambahkan atribut “pendapatan saudara kandung” (kakak kandung), ataupun tagihan token listrik dengan menggunakan metode penelitian dan alat pengumpulan data yang lebih maksimal.
2. Penelitian selanjutnya bisa dikembangkan dengan membuat sistem pengambil keputusan

Body Shop” Menggunakan Metode Naïve Bayes Classifier Dan Support Vector Machine: Studi Kasus Di Female Daily.

Cahya, D. N., Hidayat, M., & Asnawi, M. F. (2022). Implementasi Algoritma K-Means Clustering untuk Menentukan Calon Penerima Beasiswa UKT di Universitas Sains Al-Qur'an. *Journal of Engineering and Informatic*, 1(1), 26-32.

5. DAFTAR PUSTAKA

- Ikhsan, A. N., Subarkah, P., & Alifian, R. S. (2023). Komparasi Algoritme K-NN, Naïve Bayes, dan Cart untuk Memprediksi Penerima Beasiswa. *JST (Jurnal Sains dan Teknologi)*, 12(2).
- Amalia, Y. R. (2018). *Penerapan Data Mining Untuk Prediksi Penjualan Produk Elektronik Terlaris Menggunakan Metode K-Nearest Neighbor (Studi Kasus: PT. Bintang Multi Sarana Palembang)* (Doctoral dissertation, UIN RADEN FATAH PALEMBANG).
- Sidik, A. D. W. M., Kusumah, I. H., Suryana, A., Artiyasa, M., & Junfithrana, A. P. (2020). Gambaran Umum Metode Klasifikasi Data Mining. *FIDELITY: Jurnal Teknik Elektro*, 2(2), 34-38.
- Pebdika, A., Herdiana, R., & Solihudin, D. (2023). Klasifikasi Menggunakan Metode Naive Bayes Untuk Menentukan Calon Penerima PIP. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(1), 452-458.
- Silalahi, A. P., Simanullang, H. G., & Hutapea, M. I. (2023). Supervised Learning Metode K-Nearest Neighbor Untuk Prediksi Diabetes Pada Wanita. *METHOMIKA: Jurnal Manajemen Informatika & Komputerisasi Akuntansi*, 7(1), 144-149.
- NABILA, A. (2022). Analisis Sentimen Ulasan Produk Toner Pada Beauty Brand “The