

Penerapan Metode Naïve Bayes dan Weighted Product untuk Prediksi Lanjut Studi Peserta Didik

Muhammad Dahlan Kurnia¹, Tukiyyat², and Makhsun³

Teknik Informatika, Universitas Pamulang, Jl. Raya Puspittek No.46, Buaran, Serpong, Tangerang Selatan, Banten, Indonesia, 15310

e-mail: ¹emdeka1001@gmail.com, ²dosen02711@unpam.ac.id, ³makhsun_toha@yahoo.com

Submitted Date: August 21st, 2023

Reviewed Date: September 27th, 2023

Revised Date: October 19th, 2023

Accepted Date: October 22nd, 2023

Abstract

The low interest of students at the Hidayaturrohman Teluknaga Foundation to continue their education from MTs Hidayaturrohman to Hiro High School makes it necessary to look for the factors causing this lack of interest. This study aims to combine the application of advanced study predictions of students using the naïve Bayes method and the application of ranking with the weighted product method. The data for this research object are graduates of MTs Hidayaturrohman in 2022. The research sample is 322 data. The data collection method is in the form of secondary data, namely students graduating from MTs Hidayaturrohman in 2022. The attributes used to assess factors for graduates of MTs Hidayaturrohman to continue their studies at Hiro High School consist of 5 attributes, namely hobbies, modes of transportation, parents' income, distance from home to school and school test scores. In the study, 322 data were divided by 85% (273 data) for training data and 15% (49 data) for testing data. The results showed that the Naïve Bayes method could be applied in predicting the further study of students from MTs Hidayaturrohman to Hiro High School. This is evidenced by the accuracy test using the confusion matrix with an accuracy value of 71%. Where from 49 testing data it is predicted that 34 data with advanced results and 15 data with moving results. Furthermore, data ranking using a weighted product was carried out on 316 data, where 50% of the data (158 data) with the highest vector value v entered advanced ranking and the rest entered moving ranking. The 50% figure is in accordance with the expectations of Hiro High School, namely that as many as 50% of MTs Hidayaturrohman graduates continue on to Hiro High School. Then the highest vector v value is 0.005945284 for parent number 19207207 and the lowest vector v value is 0.001552376 for parent number 19207219.

Keywords: Prediksi, Yayasan Pendidikan, Naïve Bayes, Weighted Product, Confusion Matrix.

Abstrak

Penerimaan peserta didik baru yang berasal dari MTs Hidayaturrohman ke SMA Hiro masih rendah. Berdasarkan data penerimaan peserta didik baru (PPDB) SMA Hiro tahun pelajaran 2022/2023, persentase peserta didik lanjut dari MTs Hidayaturrohman ke SMA Hiro hanya mencapai 37%, sehingga perlu diprediksi lebih awal peserta didik MTs Hidayaturrohman yang akan lanjut atau pindah studi ke SMA Hiro. Untuk memprediksi hal tersebut, penelitian ini mengkombinasikan penerapan prediksi lanjut studi peserta didik dengan metode *naïve bayes* dan penerapan pemeringkatan dengan metode *weighted product*. Data obyek penelitian ini adalah para lulusan MTs Hidayaturrohman tahun 2022. Sampel penelitian sebanyak 322 data. Metode pengumpulan data berupa data sekunder yaitu siswa lulusan MTs Hidayaturrohman tahun 2022. Adapun atribut yang digunakan untuk menilai faktor-faktor pada lulusan MTs Hidayaturrohman untuk melanjutkan ke SMA hiro terdiri dari 5 atribut yaitu hobi, moda transportasi, penghasilan orang tua, jarak dari rumah ke sekolah dan nilai ujian sekolah. Dalam penelitian, 322 data dibagi sebanyak 85% (273 data) untuk data training dan 15% (49 data) untuk data testing. Hasil penelitian menunjukkan bahwa metode *naïve bayes* dapat diterapkan dalam memprediksi lanjut studi peserta didik dari MTs Hidayaturrohman ke

SMA Hiro. Hal ini dibuktikan dengan uji akurasi menggunakan *confusion matrix* dengan nilai akurasi 71%. Dari 49 data testing diprediksi 34 data dengan hasil lanjut dan 15 data dengan hasil pindah. Selanjutnya pemeringkatan data menggunakan *weighted product* dilakukan terhadap 316 data. Dari 50% data (158 data) dengan nilai vektor v tertinggi masuk pemeringkatan lanjut dan sisanya masuk pemeringkatan pindah. Adapun angka 50%, sesuai dengan harapan pihak SMA Hiro yaitu sebanyak 50% lulusan MTs Hidayaturrohman melanjutkan ke SMA Hiro. Kemudian diperoleh hasil nilai vektor v tertinggi 0,005945284 untuk nomor induk 19207207 dan nilai vektor v terendah 0,001552376 untuk nomor induk 19207219. Dari kombinasi penerapan metode *naïve bayes* dan *weighted product* dapat memprediksi mengenai studi lanjut peserta didik dari MTs Hidayaturrohman ke SMA Hiro.

Keywords: Prediksi, Yayasan Pendidikan, *Naïve Bayes*, *Weighted Product*, *Confusion Matrix*.

1. Pendahuluan

Penerimaan peserta didik baru yang berasal dari MTs Hidayaturrohman ke SMA Hiro masih rendah. Berdasarkan data penerimaan peserta didik baru (PPDB) SMA Hiro tahun pelajaran 2022/2023, persentase peserta didik lanjut dari MTs Hidayaturrohman ke SMA Hiro hanya mencapai 37%, sehingga perlu diprediksi lebih awal peserta didik MTs Hidayaturrohman yang akan lanjut atau pindah studi ke SMA Hiro. Pada akhirnya, panitia PPDB SMA Hiro dapat menentukan strategi terbaik agar lulusan MTs Hidayaturrohman dapat melanjutkan ke SMA Hiro.

Panitia PPDB SMA Hiro sudah berupaya agar lulusan MTs Hidayaturrohman yang melanjutkan ke SMA Hiro dapat bertambah. Diantaranya dengan memberikan potongan biaya masuk dan memberikan bonus untuk para kolektor. Justru hal ini menambah pengeluaran biaya, yaitu untuk potongan biaya masuk dan bonus kolektor. Dan di sisi lain tidak menambah persentase lulusan MTs Hidayaturrohman yang melanjutkan ke SMA Hiro.

Maka untuk menambah persentase lulusan MTs Hidayaturrohman yang melanjutkan ke SMA Hiro, perlu dilakukan prediksi lanjut studi peserta didik dari MTs Hidayaturrohman ke SMA Hiro menggunakan teknik data mining. Data mining memiliki fungsi sebagai fungsi prediksi, fungsi deskripsi, fungsi klasifikasi dan fungsi asosiasi. Selain itu fungsi dari data mining sendiri ialah sebagai analisis asosiasi, klasifikasi dan prediksi, analisis cluster, analisis outlier, dan analisis trend dan evolusi (Hidayah & Rozi, 2021).

Metode data mining yang digunakan untuk penelitian ini yaitu metode *naïve bayes*. Algoritma ini menggunakan suatu fungsi yang memetakan *input* ke *output* yang diinginkan. Kualitas prediksi lanjut studi sangat bergantung pada kesesuaian

input dan *output* yang diberikan untuk menjadi *data training*. Dengan demikian, user sangat berperan penting dalam memvalidasi *input* dan *output* tersebut. Dengan kata lain bahwa yang dinamakan *Supervised Learning* adalah pembelajaran terawasi, *Supervised Learning* digunakan untuk masalah klasifikasi (Mafakhir & Solichin, 2020). Dalam penelitian ini metode *naïve bayes* akan mengklasifikasikan peserta didik dalam kelas “lanjut” atau “pindah”.

Sebelum diproses menggunakan metode *naïve bayes*, mesti dilakukan *preprocessing data* terlebih dahulu. *Preprocessing data* adalah tahapan awal dari data mining untuk menghasilkan analisis yang lebih akurat dalam pemakaian teknik-teknik *machine learning* (Aspiah & Taghfirul Azhima Yoga Siswa, 2022). Setelah dihitung menggunakan metode *naïve bayes*, selanjutnya menghitung nilai akurasi data dengan harapan >70%.

Setelah prediksi lanjut studi peserta didik dari MTs Hidayaturrohman ke SMA Hiro diklasifikasi menggunakan metode *naïve bayes* maka selanjutnya hasil klasifikasi tersebut akan dikombinasikan dengan metode sistem penunjang keputusan. Sistem pendukung keputusan merupakan sebuah sistem informasi yang menggunakan model keputusan, sebuah database dan sebuah wawasan dari pembuat keputusan dalam sebuah proses pemodelan yang *ad hoc* dan interaktif untuk mencapai sebuah keputusan yang spesifik (Sahadi et al., 2020). Selanjutnya pemeringkatan data menggunakan *weighted product* dilakukan terhadap seluruh data, dari 50% data dengan nilai vektor v tertinggi masuk pemeringkatan lanjut dan sisanya masuk pemeringkatan pindah. Adapun angka 50%, sesuai dengan harapan pihak SMA Hiro yaitu sebanyak 50% lulusan MTs Hidayaturrohman melanjutkan ke SMA Hiro.

2. Metode Penelitian

2.1 Metode Naïve Bayes (NB)

Metode NB menggunakan dasar penerapan teorema Bayes dalam melakukan klasifikasi. Algoritma klasifikasi metode NB sangat efektif dalam mendapatkan hasil yang akurat dan efisien karena mampu memanfaatkan proses penalaran dari input yang ada dengan cara yang relatif cepat. Algoritma ini bertujuan untuk melakukan klasifikasi data pada kelas tertentu (Wibowo, 2020).

Langkah-langkah metode NB (Maricar & Dian Pramana, 2019):

a. Hitung probabilitas bersyarat/*likelihood*

$$P(X|C_i) = \prod_{k=1}^n P(x_k|C_i) = P(x_1|C_i) \times P(x_2|C_i) \times \dots \times P(x_n|C_i)$$

dengan:

C = class

x = vektor dari nilai atribut n

$P(x_i|C_j)$ = proporsi dokumen dari class C yang mengandung nilai atribut x_i

b. Hitung probabilitas *prior* untuk kelas yang ada

$$P(C_i) = \frac{N^j}{N}$$

dengan:

N^j = jumlah dokumen dalam suatu class

N = jumlah total dokumen

c. Hitung probabilitas *posterior*

$$\text{Posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{evidence}}$$

2.2 Metode Weighted Product (WP)

Sistem pendukung keputusan sebagai sistem informasi interaktif yang menyediakan informasi, pemodelan, dan manipulasi data dalam pengambilan keputusan (Aldisa & Purba, 2023). Dalam tahapan membuat perancangan sistem menggunakan salah satu metode pengambilan keputusan *Multiple Attribute Decision Making* (MADM) yang digunakan untuk menetapkan alternatif terbaik dari sejumlah alternatif yang ada berdasarkan beberapa kriteria tertentu yang telah ditentukan (Pamungkas et al., 2022).

Weighted product merupakan pengambilan keputusan berdasarkan analisis multi kriteria yang sangat terkenal dan merupakan metode pengambilan keputusan multi kriteria (Pariaman, 2019). Metode *weighted product* melakukan perkalian untuk menghubungkan ranting atribut, dan ranting setiap atribut harus dipangkatkan

terlebih dulu dengan bobot atribut yang sudah ditentukan sebelumnya. Proses ini bukan hanya mirip, tetapi sama dengan proses normalisasi. Preferensi untuk alternatif A_i ditunjukkan pada persamaan 2 sebagai berikut (Kartika et al., 2017):

1. Jenis dan bobot kriteria

Disesuaikan dengan kondisi pada objek penelitian

2. Normalisasi nilai bobot

$$W_j = \frac{w_j}{\sum w_j}$$

Jika jenis kriteria *cost* maka nilai bobot diberikan tanda minus. Jika jenis kriteria *benefit* maka nilai bobot tidak diberikan tanda minus.

3. Hitung nilai vector S

$$S_i = \prod_{j=1}^n X_{ij}^{W_j}$$

dengan:

S = Preferensi alternatif dianalogikan sebagai vektor S

X = Nilai kriteria

W = Bobot Kriteria/Sub kriteria

i = Alternatif

j = Kriteria

n = Banyaknya kriteria

$\sum_{j=1}^n W_j = 1$ adalah pangkat bernilai positif yang diperoleh dari perbaikan bobot pada setiap kriteria

4. Hitung nilai vector V

$$V_i = \frac{\prod_{j=1}^n X_{ij}^{W_j}}{\prod_{j=1}^n (X_{j*})^{W_j}}$$

dengan:

V = Preferensi alternatif sebagai vektor V

X = Nilai kriteria

W = Bobot Kriteria/Sub kriteria

i = Alternatif

j = Kriteria

n = Banyaknya kriteria

$*$ = Banyaknya kriteria yang telah dinilai pada vektor S

5. Urutkan nilai vector V

Nilai vector V diurutkan dari nilai terbesar sampai nilai terkecil.

3. Hasil dan Pembahasan

Tidak semua atribut dataset digunakan dalam penelitian, karena sebagian besar hanya berupa label/pengenal saja. Berikut atribut/kriteria yang digunakan dalam penelitian.

Tabel 1. Atribut Relevan

No	Nama Atribut/Kriteria	Keterangan
1	NO. INDUK	Kepanjangan dari Nomor Induk. Merupakan Nomor Induk peserta didik yang berisi data jenis numerik, tetapi hanya berfungsi sebagai simbol atau pengenalan.
2	Jarak Rumah Siswa Ke Madrasah	Merupakan jarak SMA Hiro dengan tempat tinggal peserta didik. Berisi data jenis numerik dalam satuan kilometer.
3	Hobi	Merupakan hobi yang dimiliki peserta didik. Berisi data teks terkait hobi peserta didik. Data berupa numerik ordinal sebagai berikut: 1 = Tidak Terakomodir (Tidak ada kriteria terpenuhi) 2 = Kurang Terakomodir (1 kriteria terpenuhi) 3 = Cukup Terakomodir (2 kriteria terpenuhi) 4 = Terakomodir (3 kriteria terpenuhi) 5 = Sangat Terakomodir (4 kriteria terpenuhi) Dengan ketentuan sebagai berikut: <ul style="list-style-type: none"> • Ada dalam intrakurikuler • Ada dalam ekstrakurikuler • Ada dalam pembiasaan • Sarana prasarana tersedia
4	Rata-rata Nilai US	Merupakan nilai hasil ujian madrasah/ujian sekolah mata pelajaran Bahasa Indonesia, Matematika, dan IPA pada jenjang MTs Hidayaturohman. Berisi data numerik.
5	Moda Transportasi dari Rumah ke Madrasah	Merupakan alat transportasi yang digunakan peserta didik datang ke sekolah. Data berupa data numerik ordinal sebagai berikut: 1 = jalan kaki 2 = angkutan umum

		3 = antar jemput 4 = angkutan umum 5 = motor/mobil
6	Rata-Rata Penghasilan	Rata-rata penghasilan orang tua peserta didik per bulan, dengan rentang 0: <= Rp. 500.000 1: Rp. 500.001 - Rp. 1.000.000 2: Rp. 1.000.001 - Rp. 1.500.000 3: Rp.1.500.001 - Rp. 2.000.000 4: Rp.2.000.001 - Rp.2.500.000 5: Rp. 2.500.001 - Rp. 3.000.000 6: Rp. 3.000.001 - Rp. 3.500.000 7: Rp. 3.500.001 - Rp. 4.000.000 8: Rp.4.000.001 - Rp.4.500.000 9: Rp. 4.500.001 - Rp. 5.000.000 10. >Rp.5.000.000.
7	Prediksi	Merupakan atribut target yang berisi kondisi nyata peserta didik. Jika bernilai 1 maka lanjut dan bernilai 0 maka pindah.

Berikut tampilan beberapa data yang digunakan dalam penelitian.

NO. INDUK	Jarak Rumah Siswa Ke Madrasah	Hobi	Rata-rata Nilai US	Moda Transportasi Dari Rumah Ke Madrasah	Rata-Rata Penghasilan Orangtua per Bulan	Prediksi
19207001	1	2	81	1	3	0
19207004	2	4	74	1	4	1
19207305	2	4	58	3	5	0
19207046	3	4	78	4	2	1
19207263	1	5	84	1	6	0

Gambar 1. Dataset Siap Digunakan

3.1. Preprocessing Data untuk NB

Suatu data yang ingin diolah, terutama dalam metode klasifikasi data mining memerlukan suatu label atau target, agar data tersebut dapat diolah (Bari et al., 2022). Agar memudahkan dalam proses data berikutnya perlu dilakukan transformasi data berupa label atribut. Dengan ketentuan sebagai berikut.

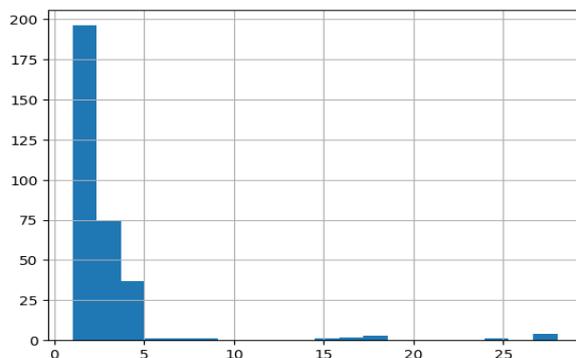
Tabel 2. Label Atribut

Label Atribut	Nama Atribut
Jarak	Jarak Rumah Siswa Ke Madrasah
Hobi	Hobi
Nilai	Rata-rata Nilai US
Transportasi	Moda Transportasi Dari Rumah Ke Madrasah
Penghasilan	Rata-Rata Penghasilan Orangtua per Bulan
Kondisi Riil	Prediksi

Tabel 3. Tampilan Label Atribut

Jarak	Hobi	Nilai	Transportasi	Penghasilan	Kondisi_Riil	
0	1	2	81.0	1	3	0
1	2	4	74.0	1	4	1
2	2	4	58.0	3	5	0
3	3	4	78.0	4	2	1
4	1	5	84.0	1	6	0
5	3	4	85.0	4	2	1
6	1	4	71.0	3	1	0
7	1	4	78.0	4	7	1

Untuk meningkatkan akurasi, digunakan teknik *binning* pada atribut Jarak, yaitu membagi atribut Jarak berdasarkan kelompok-kelompok tertentu. Berikut tampilan atribut Jarak dalam bentuk histogram.



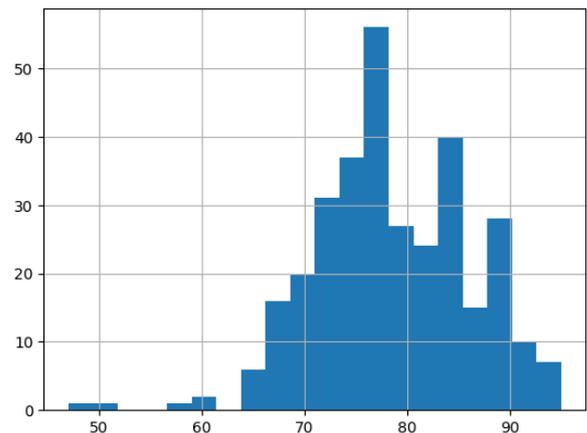
Gambar 2. Sebaran Data Atribut Jarak

Berdasarkan sebaran data di atas, tampak bahwa sebagian besar nilai atribut Jarak berada diantara 1-5 kilometer dan akan coba dibagi dalam 6 kelompok rentang, yaitu 0-1 km, 1-2 km, 2-3 km, 3-4 km, 4-5 km dan >5 km. Sehingga diperoleh hasil sebagai berikut:

Tabel 4. Hasil Teknik Binning pada Atribut Jarak

Jarak	Hobi	Nilai	Transportasi	Penghasilan	Kondisi_Riil	RentangJarak	JarakBaru	
0	1	2	81.0	1	3	0	0-1	1
1	2	4	74.0	1	4	1	1-2	2
2	2	4	58.0	3	5	0	1-2	2
3	3	4	78.0	4	2	1	2-3	3
4	1	5	84.0	1	6	0	0-1	1
5	3	4	85.0	4	2	1	2-3	3
6	1	4	71.0	3	1	0	0-1	1
7	1	4	78.0	4	7	1	0-1	1
8	2	4	89.0	3	8	0	1-2	2
9	27	5	78.0	2	9	1	>5	6

Selanjutnya teknik *binning* diterapkan pada atribut Nilai. Berikut tampilan atribut Nilai dalam bentuk histogram.



Gambar 3. Sebaran Data Atribut Nilai

Berdasarkan sebaran data pada histogram di atas, atribut Nilai akan coba dibagi dalam 5 kelompok, yaitu kelompok <70, 70-75, 75-80, 80-85, dan >80. Sehingga diperoleh hasil sebagai berikut:

Tabel 5. Hasil Teknik Binning pada Atribut Nilai

Jarak	Hobi	Nilai	Transportasi	Penghasilan	Kondisi_Riil	RentangJarak	JarakBaru	RentangNilai	NilaiBaru	
0	1	2	81.0	1	3	0	0-1	1	80-85	4
1	2	4	74.0	1	4	1	1-2	2	70-75	2
2	2	4	58.0	3	5	0	1-2	2	<70	1
3	3	4	78.0	4	2	1	2-3	3	75-80	3
4	1	5	84.0	1	6	0	0-1	1	80-85	4
5	3	4	85.0	4	2	1	2-3	3	80-85	4
6	1	4	71.0	3	1	0	0-1	1	70-75	2
7	1	4	78.0	4	7	1	0-1	1	75-80	3
8	2	4	89.0	3	8	0	1-2	2	85-100	5
9	27	5	78.0	2	9	1	>5	6	75-80	3

Maka hasil dari *preprocessing data*, atribut yang digunakan untuk proses selanjutnya adalah JarakBaru, Hobi, NilaiBaru, Transportasi, Penghasilan dan Kondisi_Riil.

3.2. Perhitungan Metode NB

Dalam prediksi lanjut studi peserta didik dari MTs Hidayaturrohman ke SMA Hiro terdiri dari 322 data dengan kondisi nyata sebanyak 150 lanjut dan 172 pindah.

$$\text{Probabilitas Prior } P(C_i) = \frac{N^j}{N} :$$

- $P(\text{Lanjut}) = 150/322 = 0,47$
- $P(\text{Pindah}) = 172/322 = 0,53$

Probabilitas *likelihood* atribut JarakBaru

- $P(\text{Jarak}=1 | \text{Lanjut}) = 64/150 = 0,43$
- $P(\text{Jarak}=2 | \text{Lanjut}) = 64/150 = 0,19$
- $P(\text{Jarak}=3 | \text{Lanjut}) = 37/150 = 0,25$
- $P(\text{Jarak}=4 | \text{Lanjut}) = 20/150 = 0,13$
- $P(\text{Jarak}=5 | \text{Lanjut}) = 0/150 = 0$

- $P(\text{Jarak}=6 \mid \text{Lanjut}) = 1/150 = 0,01$
- $P(\text{Jarak}=1 \mid \text{Pindah}) = 75/172 = 0,44$
- $P(\text{Jarak}=2 \mid \text{Pindah}) = 29/172 = 0,17$
- $P(\text{Jarak}=3 \mid \text{Pindah}) = 38/172 = 0,22$
- $P(\text{Jarak}=4 \mid \text{Pindah}) = 15/172 = 0,09$
- $P(\text{Jarak}=5 \mid \text{Pindah}) = 2/172 = 0,01$
- $P(\text{Jarak}=6 \mid \text{Pindah}) = 13/172 = 0,08$

Probabilitas *likelihood* atribut Hobi

- $P(\text{Hobi}=1 \mid \text{Lanjut}) = 17/150 = 0,11$
- $P(\text{Hobi}=2 \mid \text{Lanjut}) = 5/150 = 0,03$
- $P(\text{Hobi}=3 \mid \text{Lanjut}) = 7/150 = 0,05$
- $P(\text{Hobi}=4 \mid \text{Lanjut}) = 107/150 = 0,71$
- $P(\text{Hobi}=5 \mid \text{Lanjut}) = 14/150 = 0,09$
- $P(\text{Hobi}=1 \mid \text{Pindah}) = 16/172 = 0,09$
- $P(\text{Hobi}=2 \mid \text{Pindah}) = 23/172 = 0,13$
- $P(\text{Hobi}=3 \mid \text{Pindah}) = 10/172 = 0,06$
- $P(\text{Hobi}=4 \mid \text{Pindah}) = 109/172 = 0,63$
- $P(\text{Hobi}=5 \mid \text{Pindah}) = 14/172 = 0,08$

Probabilitas *likelihood* atribut NilaiBaru

- $P(\text{Nilai}=1 \mid \text{Lanjut}) = 26/150 = 0,17$
- $P(\text{Nilai}=2 \mid \text{Lanjut}) = 31/150 = 0,21$
- $P(\text{Nilai}=3 \mid \text{Lanjut}) = 43/150 = 0,29$
- $P(\text{Nilai}=4 \mid \text{Lanjut}) = 24/150 = 0,16$
- $P(\text{Nilai}=5 \mid \text{Lanjut}) = 26/150 = 0,17$
- $P(\text{Nilai}=1 \mid \text{Pindah}) = 21/172 = 0,12$
- $P(\text{Nilai}=2 \mid \text{Pindah}) = 37/172 = 0,22$
- $P(\text{Nilai}=3 \mid \text{Pindah}) = 40/172 = 0,23$
- $P(\text{Nilai}=4 \mid \text{Pindah}) = 40/172 = 0,23$
- $P(\text{Nilai}=5 \mid \text{Pindah}) = 34/172 = 0,20$

Probabilitas *likelihood* atribut Transportasi

- $P(\text{Transportasi}=1 \mid \text{Lanjut}) = 62/150 = 0,41$
- $P(\text{Transportasi}=2 \mid \text{Lanjut}) = 12/150 = 0,08$
- $P(\text{Transportasi}=3 \mid \text{Lanjut}) = 38/150 = 0,25$
- $P(\text{Transportasi}=4 \mid \text{Lanjut}) = 28/150 = 0,19$
- $P(\text{Transportasi}=5 \mid \text{Lanjut}) = 10/150 = 0,07$
- $P(\text{Transportasi}=1 \mid \text{Pindah}) = 61/172 = 0,35$
- $P(\text{Transportasi}=2 \mid \text{Pindah}) = 13/172 = 0,08$
- $P(\text{Transportasi}=3 \mid \text{Pindah}) = 52/172 = 0,30$
- $P(\text{Transportasi}=4 \mid \text{Pindah}) = 35/172 = 0,20$
- $P(\text{Transportasi}=5 \mid \text{Pindah}) = 11/172 = 0,06$

Probabilitas *likelihood* atribut Penghasilan

- $P(\text{Penghasilan}=1 \mid \text{Lanjut}) = 31/150 = 0,21$
- $P(\text{Penghasilan}=2 \mid \text{Lanjut}) = 41/150 = 0,27$
- $P(\text{Penghasilan}=3 \mid \text{Lanjut}) = 16/150 = 0,11$
- $P(\text{Penghasilan}=4 \mid \text{Lanjut}) = 8/150 = 0,05$
- $P(\text{Penghasilan}=5 \mid \text{Lanjut}) = 8/150 = 0,05$

- $P(\text{Penghasilan}=6 \mid \text{Lanjut}) = 9/150 = 0,06$
- $P(\text{Penghasilan}=7 \mid \text{Lanjut}) = 14/150 = 0,09$
- $P(\text{Penghasilan}=8 \mid \text{Lanjut}) = 11/150 = 0,07$
- $P(\text{Penghasilan}=9 \mid \text{Lanjut}) = 8/150 = 0,05$
- $P(\text{Penghasilan}=10 \mid \text{Lanjut}) = 1/150 = 0,01$
- $P(\text{Penghasilan}=1 \mid \text{Pindah}) = 33/172 = 0,19$
- $P(\text{Penghasilan}=2 \mid \text{Pindah}) = 15/172 = 0,09$
- $P(\text{Penghasilan}=3 \mid \text{Pindah}) = 18/172 = 0,10$
- $P(\text{Penghasilan}=4 \mid \text{Pindah}) = 17/172 = 0,10$
- $P(\text{Penghasilan}=5 \mid \text{Pindah}) = 15/172 = 0,09$
- $P(\text{Penghasilan}=6 \mid \text{Pindah}) = 10/172 = 0,06$
- $P(\text{Penghasilan}=7 \mid \text{Pindah}) = 13/172 = 0,08$
- $P(\text{Penghasilan}=8 \mid \text{Pindah}) = 16/172 = 0,09$
- $P(\text{Penghasilan}=9 \mid \text{Pindah}) = 16/172 = 0,09$
- $P(\text{Penghasilan}=10 \mid \text{Pindah}) = 0/172 = 0$

Probabilitas *posterior* merupakan probabilitas *likelihood* dari beberapa atribut. Misalkan data dengan nomor induk 19207001 dengan atribut Jarak = 1, Hobi = 2, Nilai = 4, Transportasi = 1 dan Penghasilan = 3.

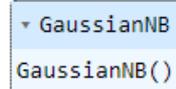
- $P(\text{Lanjut} \mid 19207001)$
 $= 0,47 \times 0,43 \times 0,03 \times 0,16 \times 0,41 \times 0,11$
 $= 0,00005$
- $P(\text{Pindah} \mid 19207001)$
 $= 0,53 \times 0,44 \times 0,13 \times 0,23 \times 0,35 \times 0,09$
 $= 0,00022$

Karena $P(\text{Lanjut} \mid 19207001) < P(\text{Pindah} \mid 19207001)$, maka nomor induk 19207001 diprediksi Pindah. Berikut sebagian hasil perhitungan NB.

3.3. Penerapan Metode NB dengan Python

Dataset dibagi dalam 2 variabel, yaitu variabel prediksi dan variabel target. Variabel prediksi (*var_pred*) terdiri dari atribut JarakBaru, Hobi, NilaiBaru, Transportasi dan Penghasilan, sedangkan variabel target (*var_tar*) terdiri dari atribut Kondisi_Riil. Selanjutnya kedua variabel tersebut akan dipecah sebanyak 85% untuk data training dan 15% untuk data testing. Memecah data sebanyak 85% dalam variabel prediksi untuk training (*pred_train*) dan variabel target untuk training (*tar_train*) serta 15% dalam variabel prediksi untuk testing (*pred_test*) dan variabel target untuk testing (*tar_test*). Diketahui jumlah data testing adalah 49 data yang diperoleh dari 15% x 322 data. Selanjutnya data siap dihitung menggunakan metode NB.

```
from sklearn.naive_bayes import GaussianNB
hiroNB = GaussianNB()
hiroNB.fit(pred_train, tar_train)
```



Gambar 4. Fungsi Metode NB pada Python

Untuk mengetahui kualitas hasil perhitungan NB maka perlu dilakukan uji akurasi data.

```
import sklearn.metrics as met
akurasi = met.accuracy_score(tar_test, prediksi)
print('Akurasi=', akurasi)
```

Akurasi= 0.7142857142857143

Gambar 5. Hasil Uji Akurasi

Tampak pada gambar di atas, variabel prediksi berisi atribut JarakBaru, Hobi, NilaiBaru, Transportasi dan Penghasilan. Sedangkan variabel target berisi atribut Kondisi_Riil. Selanjutnya kedua variabel tersebut diproses menggunakan metode NB dengan hasil akurasi 71%. Berikut Sebagian hasil metode NB terhadap 49 data testing.

Tabel 6. Hasil Metode NB

JarakBaru	Hobi	NilaiBaru	Transportasi	Penghasilan	Hasil NB
0	4	4	1	3	7 Lanjut
1	1	4	4	1	1 Lanjut
2	1	1	3	1	5 Lanjut
3	2	4	4	1	3 Lanjut
4	1	4	3	1	2 Lanjut
5	1	1	3	1	8 Lanjut
6	3	4	2	3	2 Lanjut
7	1	3	2	3	2 Lanjut
8	3	1	3	3	8 Lanjut
9	1	4	4	1	5 Lanjut

Karena akurasi sudah >70% maka Langkah berikutnya membuat model prediksi NB dengan memanfaatkan *script python*.

```
# Inputan data baru
Jarak_Ke_Hiro = 4 #ordinal 1 sampai 6
Hobi_Siswa = 4 #ordinal 1 sampai 5
Nilai_US = 1 #ordinal 1 sampai 5
Moda_Transportasi = 3 #ordinal 1 sampai 5
Penghasilan_Ortu = 7 #ordinal 1 sampai 10
Data_Baru = [[Jarak_Ke_Hiro, Hobi_Siswa, Nilai_US,
               Moda_Transportasi, Penghasilan_Ortu]]
```

Gambar 6. Model Prediksi NB dengan Python (1)

```
# Melakukan prediksi data baru
new_tar_prediksi = hiroNB.predict(Data_Baru)

# Mencetak hasil prediksi data baru
if new_tar_prediksi == 0:
    hasil = "Pindah"
elif new_tar_prediksi == 1:
    hasil = "Lanjut"
else :
    hasil = "error"

print ("Hasil Prediksi Lanjut Studi Peserta Didik = ", hasil)
```

Hasil Prediksi Lanjut Studi Peserta Didik = Lanjut

Gambar 7. Model Prediksi NB dengan Python (2)

3.4. Perhitungan Metode WP

Dataset penelitian yang digunakan sama dengan dataset yang digunakan pada metode naive bayes, dengan menambahkan atribut NO.INDUK yang nanti akan berfungsi sebagai label alternatif. Kemudian memberikan kode untuk setiap kriteria, dengan C1 = Jarakbaru, C2 = Hobi, C3 = Nilaibaru, C4 = Transportasi dan C5 = Penghasilan. Setelah dataset disiapkan, maka selanjutnya mengecek data dengan nilai 0. Karena hal ini akan membuat nilai vektor S dan vektor V menjadi 0. Jika ada data dengan nilai 0, maka data tersebut sebaiknya dihapus. Dari 322 data pada dataset terdapat 6 data dengan nilai 0, berarti data yang digunakan 322-6 = 316 data. Berikut dataset untuk metode WP yang memiliki isi data yang sama persis dengan dataset metode NB, dengan perbedaan pada label atribut saja.

Tabel 7. Tampilan Dataset Metode WP

Alternatif	C1	C2	C3	C4	C5	
0	19207001	1	2	4	1	3
1	19207004	2	4	2	1	4
2	19207305	2	4	1	3	5
3	19207046	3	4	3	4	2
4	19207263	1	5	4	1	6
...

Berikut jenis kriteria dan bobot dari masing-masing kriteria yang diperoleh dari hasil wawancara dengan panitia PPDB SMA Hiro.

Tabel 8. Jenis dan Bobot Kriteria

Nama Kriteria	Jenis Kriteria	Bobot Kriteria
Jarak	Cost	5
Hobi	Benefit	4
Nilai	Benefit	4
Transportasi	Cost	2
Penghasilan	Benefit	5



Selanjutnya melakukan normalisasi terhadap nilai bobot. Nilai bobot di normalisasi sehingga nilai total bobot = 1. Jenis kriteria = cost, maka nilai bobot = minus. Berikut perhitungannya:

$$W_j = \frac{w_j}{\sum w_j} \quad W_{C3} = \frac{4}{20} = 0,2$$

$$W_{C1} = \frac{5}{20} = -0,25 \quad W_{C4} = \frac{2}{20} = -0,1$$

$$W_{C2} = \frac{4}{20} = 0,2 \quad W_{C5} = \frac{5}{20} = 0,25$$

Selanjutnya menghitung nilai vector S. Sebagai contoh akan dihitung nilai vektor S alternatif 19207001, dengan nilai sebagai berikut: C1 = 1, C2 = 2, C3 = 4, C4 = 1, C5 = 3
 $S_{19207001} = (1^{-0,25})(2^{0,2})(4^{0,2})(1^{-0,1})(3^{0,25}) = 1,9948$

Selanjutnya menghitung nilai vector V. Sebagai contoh akan dihitung nilai vektor V alternatif 19207001, dengan nilai sebagai berikut:
 $V_{19207001} = \frac{1,9948}{538,503} = 0,003704$

Setelah nilai vektor V setiap alternatif berhasil dihitung, maka alternatif diurutkan berdasarkan nilai vektor V terbesar sampai nilai vektor V terkecil sehingga ditemukan hasil pemeringkatan dengan WP.

3.5. Penerapan Metode WP dengan Python

Setelah dataset siap diproses, maka langkah berikutnya adalah membuat variabel yang dibutuhkan dalam metode WP.

```
alternatif = datahiro_WP.iloc[:, 0].values
# Membuat variabel kriteria
kriteria = ["Jarak", "Hobi", "Nilai", "Transportasi", "Penghasilan"]
# Membuat variabel costbenefit, berisi jenis kriteria
costbenefit = ["Cost", "Benefit", "Benefit", "Cost", "Benefit"]
# Membuat variabel bobot, berisi nilai bobot kriteria dari hasil wawancara
bobot = [5, 4, 4, 2, 5]
# Membuat variabel nilaiAlternatif, berisi data dari setiap alternatif
# yang berasal dari atribut JarakBaru, Hobi, NilaiBaru, Transportasi
# dan penghasilan
nilaiAlternatif = datahiro_WP.iloc[:, [1,2,3,4,5]].values
```

Gambar 8. Variabel WP

Selanjutnya melakukan normalisasi nilai bobot dan menentukan jenis kriteria serta bobot kriteria.

```
# Menghitung jumlah bobot
jumlahBobot = 0
for i in range(len(kriteria)):
    jumlahBobot = jumlahBobot + bobot[i]
print("Jumlah Bobot = ", jumlahBobot)
```

Jumlah Bobot = 20

Gambar 9. Menghitung Jumlah Bobot

```
# Menghitung nilai bobot = bobot/jumlah bobot
nilaiBobot = []
for i in range(len(kriteria)):
    nilaiBobot.append(bobot[i]/jumlahBobot)
print("Nilai Bobot = ", nilaiBobot)
```

Nilai Bobot = [0.25, 0.2, 0.2, 0.1, 0.25]

Gambar 10. Normalisasi Nilai Bobot

```
#mengatur bobot dengan jenis kriteria cost=minus/benefit=plus
pangkatBobot = []
for i in range(len(kriteria)):
    #pangkatBobot.append(0)
    if costbenefit[i] == 'Cost':
        pangkatBobot.append(-1 * nilaiBobot[i])
    else:
        pangkatBobot.append(1 * nilaiBobot[i])
print ("Nilai Bobot Jenis Kriteria = ", pangkatBobot)
```

Nilai Bobot Jenis Kriteria = [-0.25, 0.2, 0.2, -0.1, 0.25]

Gambar 11. Menentukan Jenis Kriteria

Selanjutnya menghitung nilai vector S.

```
# Menghitung Nilai Vektor S
nilai_s=[]
total_s=0
for i in range(len(alternatif)):
    nilai_s.append(1)
    for j in range(len(kriteria)):
        nilai_s[i]=nilai_s[i]*(nilaiAlternatif[i,j]**pangkatBobot[j])
    total_s = total_s+nilai_s[i]
```

Nilai Vektor S Setiap Alternatif =

[1.994795184185913, 1.8025009252216604, 1.4865683253759978, 1.2930261804857846, 2.3275807678400264, 2.3067943099730286, 1.774753604971929: 54, 1.6437518295172258, 1.5530857575164012, 1.8628022422836297, 0.96: 36746868385573, 1.6641663941002407, 1.9572543238901665, 1.2930265287: 296432, 1.1236128227400037, 1.6245047927124714, 1.6225795538797088, : 1.3307648126513985, 1.274560627319262, 1.751384987282098, 1.53649803: 68385573, 1.7055948510016794, 2.306794309973029, 2.0439758053539716, Total nilai vektor S = 538.502688807116

Gambar 12. Menghitung Nilai Vektor S

Setelah nilai vector S dan total vector S diketahui, maka selanjutnya menghitung nilai vector V setiap alternatif.

```
#menghitung nilai vektor V
hasil_V=[]
for i in range(len(alternatif)):
    hasil_V.append(nilai_S[i]/total_S)
```

Gambar 13. Menghitung Nilai Vektor V

Setelah nilai vektor V diketahui, maka selanjutnya adalah mengurutkan alternatif dengan nilai vektor V terbesar sampai nilai vektor V terkecil.

```
#mengurutkan nilai vektor V
alternatifRanking = []
hasilRanking = []
for i in range(len(alternatif)):
    hasilRanking.append(hasil_V[i])
    alternatifRanking.append(alternatif[i])
for i in range(len(alternatif)):
    for j in range(len(alternatif)):
        if j>i:
            if hasilRanking[j]>hasilRanking[i]:
                tmpHasil = hasilRanking[i]
                tmpAlternatif = alternatifRanking[i]
                hasilRanking[i] = hasilRanking[j]
                alternatifRanking[i] = alternatifRanking[j]
                hasilRanking[j] = tmpHasil
                alternatifRanking[j] = tmpAlternatif
```

Gambar 14. Mengurutkan Nilai Vektor V

Adapun nilai pemeringkatan vektor V semua alternatif terdapat pada lampiran. Hasil terbaik diperoleh alternatif 19207207 dengan nilai vektor V = 0.005945.

Tabel 9. Hasil Pemeringkatan WP Katagori Lanjut

Alternatif	Vektor V	Hasil WP	
0	19207207	0.005945	Lanjut
1	19207032	0.005856	Lanjut
2	19207214	0.005686	Lanjut
3	19207155	0.005686	Lanjut
4	19207200	0.005438	Lanjut
...
153	19207157	0.003052	Lanjut
154	19207121	0.003047	Lanjut
155	19207076	0.003045	Lanjut
156	19207317	0.003029	Lanjut
157	19207285	0.003029	Lanjut

158 rows x 3 columns

Tabel 10. Hasil Pemeringkatan WP Katagori Pindah

Alternatif	Vektor V	Hasil WP	
158	19207088	0.003029	Pindah
159	19207096	0.003027	Pindah
160	19207060	0.003025	Pindah
161	19207258	0.003025	Pindah
162	19207137	0.003019	Pindah
...
311	19207061	0.001794	Pindah
312	19207063	0.001678	Pindah
313	20218006	0.001637	Pindah
314	19207122	0.001585	Pindah
315	19207219	0.001552	Pindah

158 rows x 3 columns

4. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan dalam penerapan metode NB dan WP untuk prediksi lanjut studi peserta didik dengan studi kasus dari MTs Hidayaturohman ke SMA Hiro, dapat disimpulkan sebagai berikut:

1. Prediksi lanjut studi peserta didik dari MTs Hidayaturohman ke SMA Hiro dengan metode naïve bayes dapat diterapkan dengan nilai akurasi 71% dan atribut yang memiliki pengaruh terhadap prediksi lanjut studi peserta didik dari MTs Hidayaturohman ke SMA Hiro yaitu atribut hobi siswa, moda transportasi yang digunakan, penghasilan orang tua, nilai ujian sekolah dan jarak rumah ke sekolah. Adapun sebagai kombinasi dapat dilakukan pemeringkatan terhadap data yang ada menggunakan metode weighted product, yaitu dengan melakukan pemeringkatan terhadap 316 data, dan data tersebut diurutkan berdasarkan nilai vektor V terbesar sampai terkecil. Setelah diurutkan, maka 50% dari 316 data dengan nilai vektor V tertinggi masuk pemeringkatan lanjut dan sisanya masuk pemeringkatan pindah. Adapun angka 50% diperoleh dari target panitia penerimaan peserta didik baru SMA Hiro yang mengharapkan 50% lulusan MTs Hidayaturohman lanjut ke SMA Hiro.
2. Model prediksi terbaik untuk memprediksi lanjut studi peserta didik dari MTs Hidayaturohman ke SMA Hiro yaitu dengan melakukan pemeringkatan terhadap semua



dataset. Selanjutnya data yang masuk pemeringkatan lanjut diprediksi menggunakan model prediksi naïve bayes. Dari 322 data yang ada, sebanyak 316 data dapat diranking menggunakan weighted product, maka 50% dari 316 data yaitu 158 data akan masuk pemeringkatan lanjut. Data yang masuk pemeringkatan lanjut akan diprediksi menggunakan model prediksi naïve bayes. Diharapkan model kombinasi antara weighted product dengan naïve bayes dapat memperkuat hasil prediksi.

Hasil penerapan metode naïve bayes dan weighted product untuk memprediksi lanjut studi peserta didik dari MTs Hidayaturrohman ke SMA Hiro dapat digunakan sebagai dasar pengambilan keputusan untuk meningkatkan jumlah peserta didik di SMA Hiro.

Referensi

- Aldisa, R. T., & Purba, R. A. (2023). Penerapan Metode Weight Product (WP) untuk Rekomendasi Pemilihan Jasa Paket Laundry Terbaik. *Journal of Information System Research (JOSH)*, 4(2), 676–682. <https://doi.org/10.47065/josh.v4i2.2925>
- Aspiah, R., & Taghfirul Azhima Yoga Siswa. (2022). Implementasi Correlation Based Feature Selection (Cfs) Untuk Peningkatan Akurasi Algoritma C4.5 Dalam Prediksi Performa Akademik Mahasiswa Berbasis Learning Management System. *Jurnal Ilmiah Betrik*, 13(2), 199–207. <https://doi.org/10.36050/betrik.v13i2.523>
- Bari, M., Majid, A., Cani, Y. M., & Enri, U. (2022). Penerapan Algoritma K-Means dan Decision Tree Dalam Analisis Prestasi Siswa Sekolah Menengah Kejuruan. *Jurnal Sistem Komputer Dan Informatika (JSON) Hal: 355–, 364(2)*, 355–364. <https://doi.org/10.30865/json.v4i2.5299>
- Hidayah, A. N. Z., & Rozi, A. F. (2021). Penerapan Data Mining Dalam Menentukan Kinerja Karyawan Terbaik Dengan Menggunakan Metode Algoritma C4.5 (Studi Kasus : Universitas Mercu Buana Yogyakarta). *Journal Of Information System And Artificial Intelligence*, 1(2), 73–80.
- Kartika, J. I., Santoso, E., & Sutrisno. (2017). Penentuan Siswa Berprestasi Menggunakan Metode K-Nearest Neighbor dan Weighted Product (Studi Kasus: SMP Negeri 3 Mejayan). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 1(5), 352–360.
- Mafakhir, A. Z., & Solichin, A. (2020). Penerapan Metode Naïve Bayes Classifier Untuk Penjurusan Siswa Pada Madrasah Aliyah Al-Falah Jakarta Abstrak. 5(1), 1–6.
- Maricar, M. A., & Dian Pramana. (2019). Perbandingan Akurasi Naïve Bayes dan K-Nearest Neighbor pada Klasifikasi untuk Meramalkan Status Pekerjaan Alumni ITB STIKOM Bali. *Jurnal Sistem Dan Informatika (JSI)*, 14(1), 16–22. <https://doi.org/10.30864/jsi.v14i1.233>
- Pamungkas, Y. A., Suryadi, L., Informasi, T., Informasi, S., Luhur, U. B., Jawab, T., & Jawab, T. (2022). Penerapan Metode Weighted Product Untuk Pemilihan Siswa Berprestasi Pada SMAIT Insan Madani 8 Application Of The Weighted Product Method For The Selection Of Achieving Student At SMAIT Insan Madani 8. *Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI)*, September, 1921–1929.
- Pariaman, P. M. A. N. (2019). Sistem Pendukung Keputusan Pemilihan Guru Terbaik dengan Metode Weighted Product. 6(2), 310–321.
- Sahadi, S., Ardiansyah, M., & Husain, T. (2020). Sistem Pendukung Keputusan Pemilihan Siswa/i Kelas Unggulan Menggunakan Metode AHP dan TOPSIS. *Jurnal Teknologi Sistem Informasi*, 1(2), 153–167. <https://doi.org/10.35957/jtsi.v1i2.513>
- Wibowo, G. W. N. (2020). Prediksi Kelanjutan Studi Siswa Ke Perguruan Tinggi Dengan Naive Bayes. *Jurnal DISPROTEK*, 11(1), 41–46. <https://doi.org/10.34001/jdpt.v11i1.1159>

