

# Penentuan penerima beasiswa dengan metode *Naïve Bayes* pada SMK Nahdlatul Ulama Karanganyar

1<sup>st</sup> Yusril Budimansyah  
Universitas Buana Perjuangan  
Karawang  
Karawang, Indonesia  
[if18.yusrilbudimansyah@mhs.ubpkarawang.ac.id](mailto:if18.yusrilbudimansyah@mhs.ubpkarawang.ac.id)

2<sup>nd</sup> Amril Mutoi Siregar  
Universitas Buana Perjuangan  
Karawang  
Karawang, Indonesia  
[amrilmutoi@ubpkarawang.ac.id](mailto:amrilmutoi@ubpkarawang.ac.id)

3<sup>rd</sup> Dwi Sulistya Kusumaningrum  
Universitas Buana Perjuangan  
Karawang  
Karawang, Indonesia  
[dwi.sulistya@ubpkarawang.ac.id](mailto:dwi.sulistya@ubpkarawang.ac.id)

**Abstract**— Pendidikan memiliki pengaruh penting dalam menghasilkan anak didik yang berkarakter, inovatif dan berprestasi. Prestasi tentu dapat diraih berdasarkan kemampuan individu yang didasari pengetahuan yang baik dari tiap mata pelajaran. Salah satu keinginan murid meraih prestasi diluar maupun didalam sekolah adalah mendapatkan beasiswa. Namun pemberian bantuan beasiswa seringkali kurang tepat sasaran, maka dibuat penentuan penerimaan beasiswa di SMK Nahdlatul Ulama dengan tahapan pembersihan data, seleksi data, transformasi data, perhitungan manual maupun didalam program *python*, dan evaluasi pola menggunakan algoritma *Naïve Bayes*. Dengan hasil 25 data *True positive*, 96 data *True negative* dan 7 data *False positive*. Lalu evaluasi dari algoritma *Naïve Bayes* dengan menggunakan 31 data *training*, serta 128 data *testing* yang menghasilkan nilai akurasi 94,5%, presisi 78,1%, dan *recall* 100%.

**Kata kunci** — Prestasi, Beasiswa, *Naïve Bayes*, *python*.

## I. PENDAHULUAN

Pendidikan salah satu faktor yang memiliki kontribusi begitu berarti dalam menjalani aktivitas sebagai individu, dalam hal lain perkembangan zaman serta teknologi mendorong pendidikan menjadi sebuah kebutuhan. Pendidikan memiliki pengaruh penting dalam menghasilkan anak didik yang berkarakter, inovatif dan berprestasi[1]. Salah satu keinginan murid meraih prestasi diluar maupun didalam sekolah adalah mendapatkan beasiswa.

Beasiswa diperuntukan untuk mengurangi beban biaya pendidikan siswa yang menerimanya[2], bentuk beasiswa berupa bantuan uang dengan nominal yang sudah ditentukan dan akan diberikan kepada perorangan[3]. Pada tiap instansi pendidikan pasti terdapat beasiswa yang bertujuan membantu siswa siswi berprestasi yang kesulitan dalam biaya pendidikan dengan ketentuan yang sudah di tetapkan. Namun pemberian bantuan beasiswa seringkali kurang tepat sasaran, salah satu penyebab salahnya menentukan penerima beasiswa adalah karena banyaknya pengajuan beasiswa.

SMK Nahdlatul Ulama Karanganyar mempunyai program beasiswa yang diperuntukan guna pelajar yang berprestasi serta kurang berkecukupan. Beasiswa yang akan dibagikan terdapat kriteria meliputi pekerjaan orang tua, penghasilan, prestasi, jumlah tanggungan, serta peringkat kelas. Penentuan beasiswa dilakukan dengan cara seleksi administrasi dan komponen penilaian secara manual. Belum adanya perhitungan secara sistematis dan aplikasi dalam memecahkan masalah tersebut.

Adapun penelitian dalam klasifikasi penerima beasiswa data menggunakan algoritma *Naïve Bayes* diantaranya oleh Febrian Anggoro Harimurti (2017) tentang klasifikasi penerima beasiswa menggunakan metode *Naïve Bayes* ( studi kasus Universitas Trunojoyo Madura ). Hasil penelitian yang diperoleh bahwa penerapan Algoritma *Naïve Bayes* dengan jumlah 176 data mahasiswa untuk beasiswa PPA dan 97 data mahasiswa untuk beasiswa BBM dengan ketentuan 2/3 untuk data latih dan 1/3 data uji. Dimana Algoritma tingkat akurasi beasiswa PPA 83,05% dan beasiswa BBM 75%[4]. Penelitian yang selanjutnya dilakukan oleh Nining Suryani (2019) mengenai optimasi *Naïve Bayes* untuk prediksi penerimaan beasiswa pendidikan pada SMP Utama. Penelitian tersebut menghasilkan dari 200 data siswa yang digunakan diperoleh 154 siswa yang mendapatkan beasiswa dengan tingkat akurasi sebesar 83%[5].

Berdasarkan penelitian diatas algoritma *Naive Bayes* sangat efektif digunakan, dibandingkan Algoritma data mining lainnya. Penulis berharap dengan algoritma *Naive bayes* pada mengklasifikasi data sekolah mampu mempermudah memecahkan masalah dalam menentukan penerima beasiswa di SMK NU Karanganyar.

## II. METODE PENELITIAN

### A. Data Mining

Data mining merupakan sebuah teknik yang diterapkan secara otomatis dengan memanfaatkan teknik statistik, matematika, kecerdasan buatan dan mesin pembelajaran untuk diekstrak dan mendapatkan berupa informasi pengetahuan potensial dan konstruktif, kemudian dimasukkan dalam database yang luas. Informasi yang dihasilkan merupakan komponen dari proses *Knowledge Discovery in Database (KDD)* yang tersusun dari beberapa langkah diantaranya pemilihan data, prapengolahan, transformasi, data mining dan evaluasi hasil[6].

B. Beasiswa

Beasiswa merupakan salah satu program sekolah yang diberikan kepada peserta didik yang berprestasi dibidang akademik ataupun *non* akademik. Beasiswa bertujuan membantu mengurangi pengeluaran biaya sekolah bagi peserta didik. Menurut Rima Mawarni, Meri Parida (2020) beasiswa juga bisa dijadikan sebuah pendapatan bakal siswa yang akan menerimanya, sebab beasiswa sendiri mampu meningkatkan nilai ekonomis bagi yang menerimanya, sehingga beasiswa adalah salah satu penghasilan[7] Pada dasarnya, beasiswa adalah penghasilan bagi yang menerimanya.

C. Algoritma *Naïve Bayes*

Menurut Syarli dan Asrul Ashari Muin[8] Thomas Bayes ilmuwan dari inggris mengutarakan sebuah metode klasifikasi dengan mengaplikasikan probabilitas dan statistik yaitu algoritma *Naïve Bayes*. Algoritma tersebut menghitung satu *set* probabilitas dengan menambahkan frekuensi dan kombinasi nilai dari kumpulan data algoritma menggunakan teorema Bayes dan mengasumsikan bahwa semua atribut berhubungan atau tidak berhubungan dari nilai yang diberikan ke variabel kelas. *Naïve Bayes* memiliki pembangunan model yang cepat, kemampuan prediktif, dan juga menyediakan metode baru untuk mengeksplorasi dan memahami data.

Algoritma ini mempunyai keuntungan, karena metode ini hanya membutuhkan jumlah data pelatihan (training data) yang kecil untuk menetapkan jangka waktu parameter yang diperlukan dalam proses pengklasifikasian. Dengan kata lain, dalam *Naïve bayes*, model yang digunakan adalah “model fitur independen”.

$$P(X|H) = \frac{P(X|H)P(H)}{P(X)} \tag{1}$$

Dimana:

- P = Data dengan kelas yang tidak diketahui.
- H = Data Hopotesis X yaitu suatu class yang spesifik.
- P(H|X) = Probabilitas hipotesis H berdasarkan kondisi x (probabilitas posterior).
- P(H) = Probabilitas hipotesis H (probabilitas sebelumnya).
- P(X|H) = Probabilitas X berdasarkan kondisi hipotesis H.
- P(X) = Probabilitas X)

D. *Evaluasi*

Pada tahapan evaluasi peneliti akan melakukan beberapa percobaan guna ditentukannya perhitungan manakah yang nilai akurasi maksimal dalam menentukan penerima beasiswa. Evaluasi yang digunakan adalah *Confusion matrix* untuk mengetahui nilai tingkat akurasi[9]. *Confusion matrix* adalah pengujian model data mining untuk mengetahui nilai akurasi, berupa gambar hasil yang benar berkisar dari prediksi positif benar, prediksi positif yang meleset, prediksi negatif benar, dan prediksi negatif salah. Akurasi akan dihitung dari semua prediksi yang benar ( prediksi positif dan negative ). Semakin tinggi nilai akurasi, semakin baik model yang diperoleh.

E. Prosedur Penelitian

Prosedur penelitian dimulai dengan pengumpulan data dengan rincian sebagai berikut:



Gambar 1. Prosedur Penelitian

## III. HASIL DAN PEMBAHASAN

## A. Pengumpulan Data

Didapatkan data dari SMK NU Karanganyar Indramayu data siswa kelas X, XI, XII jurusan Teknik Komputer dan Jaringan serta Agribisnis Pengolahan Hasil Pertanian tahun pelajaran 2018-2019 sampai 2021-2022. Yang ditampilkan sebagai berikut :

Table 1 Biodata siswa

No	Nama	NIPD	JK	...	Peringkat
1	ABDUL KOHAR	202110001	L	...	11
2	ABDUL WAHID	192010001	L	...	15
3	ADEN SABILI	212210000	L	...	17
4	ADHITYA JANUAR SAMIAJI	202110002	L	...	20
5	AHMAD SAKURI	192010002	L	...	7
6	ALIMAH	212210001	P	...	2
7	ALISA	212210002	P	...	6
8	AMELIA PUTRI	192010004	P	...	8
9	ANA FITRIYAH	202110060	P	...	18
10	ANDRE HARDIYANTO	212210003	L	...	19
...	...	...	...	...	...
...	...	...	...	...	...
...	...	...	...	...	...
158	YEPI RIYANTO	202110051	L	...	24
159	ZAHROTUS TSANIYAH	212210054	P	...	28

## B. Seleksi data

Setelah dilakukan penyeleksian variable yang tidak di perlukan, yang serta di tambahan variabel kelas "Beasiswa". Penentuan *input* dalam kelas beasiswa diambil dari beberapa ketentuan variabel diantaranya Jumlah orang tua hidup, Saudara kandung, Penghasilan orang tua, serta untuk variabel Peringkat kelas diubah menjadi Prestasi.

Untuk mendapatkan bantuan beasiswa terdapat 4 kriteria dengan ketentuan, sebagai berikut:

- 1) 2 orang tua yang masih hidup, lebih diutamakan yatim atau piatu.
- 2) Jumlah tanggungan 2 atau lebih.
- 3) Gaji kedua orang tua tidak lebih Rp. 1.000.000
- 4) Memiliki prestasi di lingkungan sekolah atau lingkungan luar sekolah.

Sehingga didapatkan suatu data yang akan dijadikan sebagai kumpulan data yang akan digunakan dalam penelitian ini. Data set dapat dilihat berikut ini :

Table 2 Data set

Jumlah orang tua	Jumlah tanggungan orang tua	Penghasilan orang tua	Prestasi	Beasiswa
2	3	Menengah	Tidak	Tidak dapat
2	2	Menengah	Tidak	Tidak dapat
2	1	Atas	Tidak	Tidak dapat
2	5	Rendah	Tidak	Tidak dapat
1	3	Rendah	Tidak	Tidak dapat
1	1	Menengah	Ya	Dapat
2	1	Rendah	Ya	Tidak dapat
2	4	Rendah	Tidak	Tidak dapat
2	4	Rendah	Tidak	Tidak dapat
2	1	Atas	Tidak	Tidak dapat
1	3	Atas	Tidak	Tidak dapat
2	1	Menengah	Tidak	Tidak dapat
2	2	Atas	Tidak	Tidak dapat
...	...	...	...	...
...	...	...	...	...
...	...	...	...	...
2	3	Menengah	Tidak	Tidak Dapat
1	1	Atas	Tidak	Tidak Dapat

C. Klasifikasi

1. Perhitungan Manual

Hasil Dari perhitungan manual menggunakan model algoritma Naïve Bayes menggunakan Microsoft Exel 2013 menunjukan nilai akurasi sebesar 94,5% terhadap data testing, menggunakan data latih kisaran 159 dengan 128 data uji yang di ambil dari data latih.

Table 3 Data hasil prediksi

Kelas prediksi	Dapat	Tidak Dapat
Tidak Dapat	0	0,038539836

Kelas prediksi	Dapat	Tidak Dapat
Tidak Dapat	0	0,053898735
Tidak Dapat	0	0,04729889
Tidak Dapat	0	0,008698775
Tidak Dapat	0	0,02087914
Tidak Dapat	0	0,002525867
Tidak Dapat	0	0,009981247
Tidak Dapat	0	0,081846227
Tidak Dapat	0	0,038539836
Tidak Dapat	0	0,043917488
Tidak Dapat	0	0,010278481
Tidak Dapat	0	0,053898735
Tidak Dapat	0	0,029047466
-	-	-
-	-	-
-	-	-
Tidak Dapat	0	0,053898735
Tidak Dapat	0	0,01832251

2. Program Python

Pada tahap ini, Algoritma Naïve Bayes dihitung menggunakan bahasa pemrograman python.

```
prediction = NaiveBayes.predict(X_test)
accuracies = accuracy_score(y_test, prediction)

print(prediction)
print(accuracies)

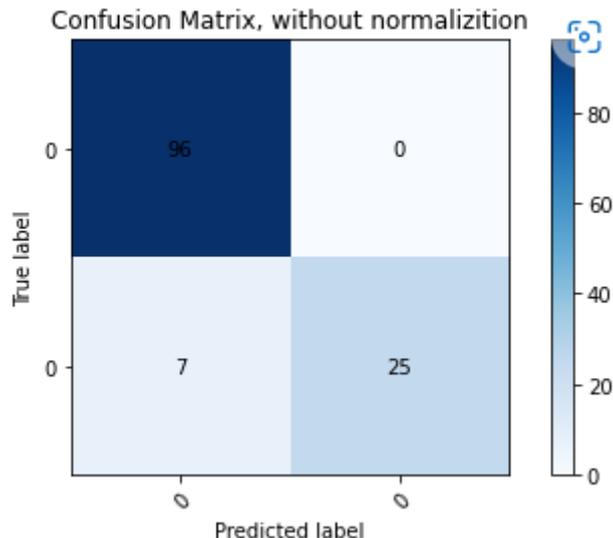
[0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0 1 1 0 0 1 1 0 1 0 1 0 0 0 0 0
1 0 0 0 1 0 0 0 0 0 0 1 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 1 0 0 1 1 0
1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 1 1 0 0 0 0 0 1 0 0 0 0 0 0 1
0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0]
0.9453125
```

Gambar 1 Hasil Perhitungan Program Python metode Naïve Bayes  
147

Hasil prediksi dari data testing sebanyak 128 data dengan keterangan 0 untuk “Tidak Dapat” dan 1 untuk “Dapat”. Dengan akurasi data testing yang terdiri dari 96 *True negative*, 25 *True positive* dan 7 *False positive* dengan akurasi data testing sebesar 94,5 %.

D. Evaluasi

Setelah melakukan perhitungan manual selanjutnya adalah melakukan evaluasi dengan menggunakan metode *Confusion Matrix*. Dalam metode ini kita akan mencari nilai performa akurasi, presisi, *recall* berdasarkan hasil visualisasi data testing yang digunakan:



Gambar 2 Visualisasi *Confusion Matrix*

Hasil prediksi dari Kemudian masukan data yang terdapat pada visual confusion matrix kedalam tabel confusion matrix untuk mencari nilai akurasi, seperti berikut datanya :

Table 4 Hasil *Confusion Matrix*

Klasifikasi	Prediksi	
	Dapat	Tidak dapat
Dapat	25	7
Tidak dapat	0	96

1. Menghitung nilai akurasi

$$Akurasi = \frac{25+96}{25+96+7+0} = 94,5\%$$

2. Menghitung presisi

$$Presisi = \frac{25}{25+7} = 78,1\%$$

3. Menghitung presisi

$$Recall = \frac{25}{25+0} = 100\%$$

IV. KESIMPULAN DAN SARAN

A. Kesimpulan

Sehubungan dengan penelitian ini, dapat disimpulkan beberapa hal sebagai berikut ::

1. Penerapan algoritma *Naïve Bayes* dalam perhitungan penerima beasiswa SMK Nahdlatul Ulama dengan tahapan pembersihan data, seleksi data, transformasi data, perhitungan manual maupun didalam program *python*, dan evaluasi pola berhasil di implementasikan.
2. Hasil perhitungan yang diperoleh dari penerapan algoritma *Naïve Bayes* dalam penerimaan beasiswa dalam bentuk file “*xlx*” dan program *python*. Dengan hasil 25 data *True positive*, 96 data *True negative* dan 7 data *False positive*.

Lalu evaluasi dari algoritma *Naïve Bayes* dari 31 data *training*, serta 128 data *testing* yang menghasilkan nilai akurasi 94,5%, presisi 78,1%, dan *recall* 100%..

#### B. Saran

Penelitian kedepannya diharapkan mampu menggunakan data set yang baik dan banyak serta melakukan pengujian menggunakan program lainya. Dalam melakukan klasifikasi data set juga diperlukan perhitungan manual yang sesuai dengan dataset yang diterapkan, sehingga hasil yang diperoleh lebih kompeten dan sesuai harapan.

#### Pengakuan

Karya ilmiah ini adalah sebagai dari penelitian Tugas Akhir milik Yusril Budimansyah yang dibimbing oleh Amril Mutoi Siregar dan Dwi Sulistya Kusumaningrum dengan judul Penentuan penerima beasiswa dengan motode *Naïve Bayes* pada SMK Nahdlatul Ulama Karanganyar.

#### DAFTAR PUSTAKA

- [1] P. Sd, N. Batu, J. Tangerang, and I. Pendahuluan, "Metode Ahp Dan Topsis," vol. 7, no. 2, 2020.
- [2] R. A. Saputra and S. Ayuningtias, "Penerapan Algoritma Naive Bayes Untuk Penentuan Calon Penerima Beasiswa Pada Smk Pasim Plus Sukabumi," *Swabumi*, vol. IV, no. 2, pp. 114–120, 2016.
- [3] A. Pratama, F. Ali, I. Ade, and R. Rinaldi, "Klasifikasi Penerima Beasiswa Dengan Menggunakan Algoritma K Nearest Neighbor," *J. Data Sci. Inform.*, vol. 1, no. 1, pp. 11–15, 2021, [Online].
- [4] F. A. Harimurti and E. Riksakomara, "Klasifikasi Penerimaan Beasiswa Menggunakan Metode *Naïve Bayes Classifier* (Studi Kasus Universitas Trunojoyo Madura)," 2017.
- [5] I. Wahyudi, S. Bahri, and P. Handayani, "Aplikasi Pembelajaran Pengenalan Budaya Indonesia," vol. V, no. 1, pp. 135–138, 2019.
- [6] R. Situmeang, "Implementasi Data Mining Pola Penyakit Pasien Menerapkan Decision Tree J48 (Studi Kasus : RS.Estomihi)," *J. Ris. Komput.*, vol. 6, no. 6, pp. 575–579, 2019, [Online].
- [7] R. Mawarni and M. Parida, "Perancangan Aplikasi Analytical Hierarchy Process Berbasis Web Penentuan Penerimaan Beasiswa Smk Negeri 1 Talangpadang," *Jtksi (Jurnal Teknol. Komput. Dan Sist. ...)*, vol. 03, no. 01, 2020.
- [8] A. A. Muin, "Metode Naive Bayes Untuk Prediksi Kelulusan," *J. Ilm. Ilmu Komput.*, vol. 2, no. 1, pp. 22–26, 2017.
- [9] M. B. Nugroho, "Bab Ii Tinjauan Pustaka 1.2.," *J. Chem. Inf. Model.*, vol. 53, no. 9, pp. 1689–1699, 2017.