

---

---

## PREDIKSI DALAM PENENTUAN BEASISWA DI SMP NEGERI 1 GEDUNG SURIAN MENGGUNAKAN ALGORITMA NAIVE BAYES

**Kuswandari<sup>1\*</sup>, Febri Sugandi<sup>2</sup>**

**Univeritas Dharmawacana, Lampung**

email: Brizky383@gmail.com, fsugandi87@gmail.com

**Abstract:** *Data mining is one approach that can be used to assist in the decision-making process, including in predicting scholarship recipients. One method in data mining that is quite popular and effective is the Naive Bayes algorithm. This study aims to apply the Naive Bayes algorithm in predicting scholarship recipients at SMP Negeri 1 Gedung Surian using the RapidMiner software tool. Data collection is carried out before carrying out the Algorithm calculation process. The data used in this study are student data at SMP Negeri 1 Gedung Surian, West Lampung in 2025. The results of the study showed that predictions using the Naive Bayes algorithm on scholarships at SMP Negeri 1 Gedung Surian, West Lampung, accuracy reached 97.84%, precision 100% and Recall 97.64% so that it can be interpreted, the Naive Bayes algorithm model is a good algorithm in predicting, this shows that the advantage of the Naive Bayes method is based on Bayes' theorem.*

**Keywords:** *SMP Negeri 1 Gedung Surian, prediction, Naive Bayes algorithm.*

**Abstrak:** Data mining merupakan salah satu pendekatan yang dapat digunakan untuk membantu dalam proses pengambilan keputusan, termasuk dalam prediksi penerima beasiswa. Salah satu metode dalam data mining yang cukup populer dan efektif adalah algoritma Naive Bayes. Penelitian ini bertujuan untuk menerapkan algoritma Naive Bayes dalam memprediksi penerima beasiswa di SMP Negeri 1 Gedung Surian dengan menggunakan alat bantu perangkat lunak RapidMiner. Pengumpulan data dilakukan sebelum melakukan proses perhitungan Algoritma. Data yang digunakan dalam penelitian ini adalah data siswa sekolah SMP Negeri 1 Gedung Surian Lampung Barat pada tahun 2025. Hasil penelitian menunjukkan prediksi menggunakan algoritma Naive Bayes pada beasiswa pada Sekolah SMP Negeri 1 Gedung Surian Lampung Barat, akurasi mencapai 97,84%, presisi 100% dan Recall 97,64% sehingga dapat diartikan, model algoritma Naive Bayes merupakan algoritma yang baik dalam memprediksi, hal ini menunjukkan bahwa keunggulan dari metode Naive Bayes yaitu berdalil pada teorema bayes.

**Kata kunci :** SMP Negeri 1 Gedung Surian, prediksi, algoritma Naive Bayes.

### PENDAHULUAN

Pendidikan merupakan salah satu faktor utama dalam pembangunan suatu negara. Salah satu upaya pemerintah dalam meningkatkan kualitas pendidikan adalah dengan menyediakan program beasiswa bagi siswa yang memenuhi kriteria tertentu. Program beasiswa diharapkan dapat membantu siswa dari keluarga kurang mampu atau yang berprestasi agar tetap dapat melanjutkan

pendidikan tanpa hambatan biaya.

Namun, dalam proses penentuan penerima beasiswa, seringkali terdapat permasalahan seperti subjektivitas penilaian, keterbatasan waktu, dan keterbatasan sumber daya manusia. Oleh karena itu, dibutuhkan suatu sistem yang dapat membantu dalam proses pengambilan keputusan pemberian beasiswa secara objektif dan efisien.

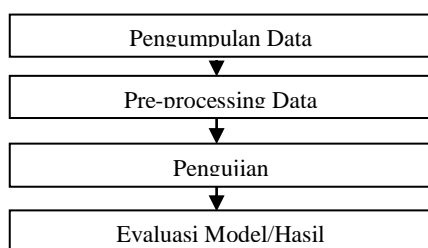
Data mining merupakan salah satu pendekatan yang dapat digunakan untuk

membantu dalam proses pengambilan keputusan, termasuk dalam prediksi penerima beasiswa. Salah satu metode dalam data mining yang cukup populer dan efektif adalah algoritma Naive Bayes. Algoritma ini mampu melakukan klasifikasi data berdasarkan probabilitas kemunculan data tertentu.

Penelitian ini bertujuan untuk menerapkan algoritma Naive Bayes dalam memprediksi penerima beasiswa di SMP Negeri 1 Gedung Surian dengan menggunakan alat bantu perangkat lunak RapidMiner.

## METODE

Adapun kerangka penelitian adalah sebagai berikut :



Gambar 1 kerangka penelitian

## Sumber Data

Data diperoleh dari SMP Negeri 1 Gedung Surian Lampung Barat, adapun data yang diperoleh berupa, Pekerjaan Orang Tua, Penghasilan orang tua. Jumlah tanggungan Keluarga, Status keluarga serta Status penerima beasiswa (label).

## Teknik Pengumpulan Data

Dalam melakukan pengumpulan data, penulis meminta sejumlah data di Sekolah SMP Negeri 1 Gedung Surian Lampung Barat dengan beberapa teknik yaitu, Wawancara dengan pihak sekolah, Dokumentasi arsip data siswa dan Observasi data dari sistem administrasi sekolah.

## Teknik Analisa Data

Dalam hal ini teknik analisa data pada penelitian yang digunakan dalam analisis data dilakukan dengan langkah-

langkah, Preprocessing data (normalisasi, encoding), Splitting data (training dan testing), Penerapan algoritma Naive Bayes, Evaluasi hasil menggunakan confusion matrix dan akurasi

## Pengujian Metode Naive Bayes

Naive Bayes adalah salah satu algoritma klasifikasi yang didasarkan pada Teorema Bayes dengan asumsi bahwa setiap fitur (atribut) saling bebas atau independen satu sama lain. Algoritma ini termasuk dalam metode klasifikasi probabilistik yang sederhana namun cukup efektif dan sering digunakan dalam berbagai kasus data mining, termasuk klasifikasi teks, filtering email spam, prediksi penyakit, dan penentuan keputusan seperti seleksi beasiswa.

Teorema Bayes adalah dasar matematis dari algoritma Naive Bayes. Teorema ini menyatakan:

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

Keterangan:

- $P(H|X)P(H|X)P(H|X)$ : Probabilitas hipotesis H yang benar diberikan data X (posterior probability)
- $P(X|H)P(X|H)P(X|H)$ : Probabilitas data X muncul jika hipotesis H benar (likelihood)
- $P(H)P(H)P(H)$ : Probabilitas awal dari hipotesis (prior probability)
- $P(X)P(X)P(X)$ : Probabilitas dari data X (evidence)

Dalam konteks klasifikasi, hipotesis H adalah kelas tertentu (misalnya "Layak Beasiswa"), dan X adalah data fitur siswa (misalnya Pekerjaan Orang Tua, Penghasilan orang tua. Jumlah tanggungan Keluarga, Status keluarga serta Status penerima beasiswa (label).

## HASIL DAN PEMBAHASAN

### Persiapan Data

Pengumpulan data dilakukan

sebelum melakukan proses perhitungan Algoritma. Data yang digunakan dalam penelitian ini adalah data siswa sekolah SMP Negeri 1 Gedung Surian Lampung Barat pada tahun 2025. Pada raw data terdapat atribut seperti, Nama, NIPD, JK, NISN, Tempat lahir, tanggal lahir, NIK, Agama, Alamat, Nama Ayah, Nama Ibu, Pekerjaan Orang Tua, Penghasilan orang tua. Jumlah tanggungan Keluarga, Status keluarga.

redundant. Dimana atribut yang nilainya kosong akan dihilangkan.



**Gambar 2 Tahap Preprocessing Data menggunakan Rapid Miner**

**Tabel 1 Raw Data**

**Seleksi Data**

Seleksi data yaitu dengan memilih atribut-atribut yang diperlukan serta membuang atribut yang tidak dibutuhkan. Pada data siswa sekolah SMP Negeri 1 Gedung Surian Lampung Barat pada tahun 2025, adapun atribut yang tidak diperlukan adalah NIPD, JK, NISN, Tempat lahir, tanggal lahir, NIK, Agama, Alamat, Nama Ayah, Nama Ibu, seperti tabel 2 berikut:

**Preprocessing Data**

Menurut (Putra, A., & Sugandi, F. (2024), pre-processing adalah proses persiapan data. Proses ini sangat menentukan keberhasilan penambangan data. Pertama dilakukan pemilihan data, yaitu proses penghilangan atribut-atribut yang tidak digunakan. Maka pada penelitian ini dilakukan terlebih dahulu preprocessing data agar data menjadi berkualitas karena dari suatu data yang tidak berkualitas akan menghasilkan kualitas data mining yang tidak baik, kemudian didalam mendapatkan data yang berkualitas yaitu dengan cara mengubah, mengoreksi, atau menghapus data-data yang salah, tidak lengkap, tidak akurat, atau memiliki format yang salah dalam basis data guna menghasilkan data berkualitas tinggi.

**Tabel 2 Seleksi Data**

Atribut	Ket
Nama	√
NIPD	x
JK	x
NISN	x
Tempat Lahir	x
Tanggal Lahir	x
NIK	x
Agama	x
Nama Ayah	x
Nama Ibu	x
Pekerjaan Orang Tua	√
Penghasilan Orang Tua	√
Jumlah Tanggungan Keluarga	√
Status Keluarga	√
Beasiswa (label)	√

**Pembersihan Data**

Data cleaning yaitu melakukan pembersihan data terhadap noise yang ditemukan berupa missing value, inkonsisten data dan redundant data. Seluruh atribut akan diseleksi untuk mendapatkan atribut-atribut yang relevan, tidak missing value dan

Berdasarkan pada tabel 2 diatas maka tabel yang digunakan pada penelitian ini terdiri dari Nama, Pekerjaan Orang Tua, Penghasilan orang tua. Jumlah tanggungan Keluarga, Status keluarga.

**Import Data Rapidminer**

Pada impor data menggunakan rapidminer sebanyak 751 record data, dimana data training sebanyak 438 dan

data testing sebanyak 303, seperti gambar 3 dibawah ini :

Gambar 3 Import data

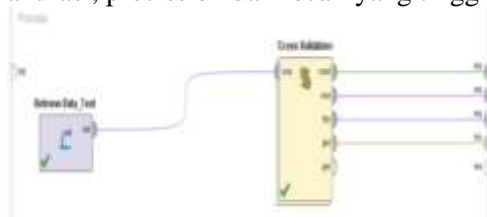
Perubahan type data pada 6 atribut diatas dapat dilihat seperti tabel 3 dibawah ini :

Tabel 3 Type data

Atribut	Tipe Atribut
Nama	Polynomial
Pekerjaan Orang Tua	Polynomial
Penghasilan Orang Tua	Polynomial
Jumlah Tanggungan Keluarga	Polynomial
Status Keluarga	Polynomial
Beasiswa	Label

### Cross Validation

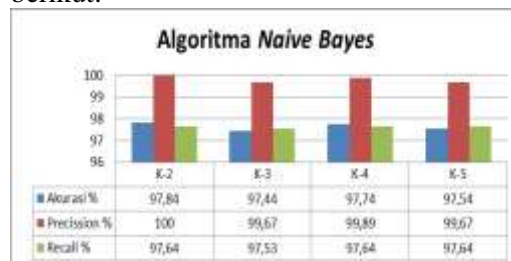
Metode statistik yang dapat digunakan untuk mengevaluasi kinerja model klasifikasi dimana data dipisahkan menjadi dua bagian yaitu data proses latih (training) dan data uji disebut k-fold cross validation. Didalam pengujian menggunakan cross validation pada algoritma Naive Bayes pada tools rapidminer, perlu dilakukan percobaan sebanyak nilai k fold terdahulu dengan bertujuan agar mendapatkan tingkat akurasi, precession dan recall yang tinggi.



Gambar 3 Cross Validation

Pada penelitian ini pengujian k fold cross validation dilakukan percobaan sebanyak 5 kali pada algoritma Naive

Bayes menggunakan tools rapidminer, dimana hasil dalam percobaan tersebut ditunjukkan pada gambar 5 sebagai berikut:

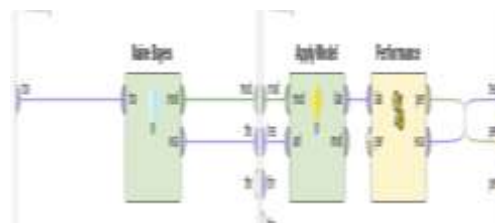


Gambar 5 Grafik k fold CV

Pada gambar 5 diatas akurasi tertinggi algoritma Naive Bayes terletak pada k = 2 fold cross validation yaitu dengan mencapai Akurasi sebesar 97,84%, Precision 100% dan Recall 97,64%.

### Pengujian Data

Pada pengolahan data menggunakan proses pengambilan data pada Rapidminer untuk Algoritma Naive bayes langsung pada format xls. Proses pengolahan data menggunakan Algoritma Naive bayes seperti gambar 6. dibawah ini:



Gambar 6 Proses pengolahan Data

### Confusion Matriks Naive Bayes

Hasil dari pengukuran kinerja klasifikasi menggunakan algoritma Naive bayes pada data testing sekolah SMP Negeri 1 Gedung Surian Lampung Barat dapat dilihat sebagai berikut:

Tabel 4 Hasil Performa

	Real Test Results	Real Total Test Results	Real Precision
Real Test Results	20	4	100%
Real Total Test Results	20	20	75%
Real Recall	100%	100%	

Dari hasil diatas membuktikan taraf akurasi mencapai 97,84%, Precision 100% dan Recall 97,64%, beasiswa pada Sekolah

SMP Negeri 1 Gedung Surian Lampung Barat. menggunakan Algoritma Naïve Bayes.

## SIMPULAN

Adapun hasil prediksi menggunakan algoritma Naïve Bayes pada beasiswa pada Sekolah SMP Negeri 1 Gedung Surian Lampung Barat, akurasi mencapai 97,84%, presisi 100% dan Recall 97,64% sehingga dapat diartikan, model algoritma Naive Bayes merupakan algoritma yang baik dalam memprediksi, hal ini menunjukkan bahwa keunggulan dari metode Naive Bayes yaitu berdalil pada teorema bayes yang juga bisa diartikan mampu dan dapat memprediksi peluang dimasa depan.

## DAFTAR PUSTAKA

Abdullah Baz, Fatima Alshareef, Ebtihal Alshareef, Hosam Alhakami, Tahani Alsubait. 2020. "Predicting Students' Academic Performance Using Naïve Bayes". *IJCSNS International Journal of Computer Science and Network Security*.

Artaye, Ketut. 2015. "Implementation of Naive Bayes Classifier Method to Predict Graduation Time of IBI Darmajaya Scholar". *International Conference on Information Technology and Business:1-2*.

Bonaccorso, Giuseppe. 2017. *Machine*

*Learning Algorithm*. Packt: Birmingham-Mumbai.

Han, J., & Kamber, M. (2012). *Data Mining: Concepts and Techniques*. Elsevier.

Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann.

Larose, D. T. (2005). *Discovering Knowledge in Data: An Introduction to Data Mining*. John Wiley & Sons.

RapidMiner Documentation. (2023). Retrieved from <https://docs.rapidminer.com>

Putra, A., & Sugandi, F. (2024). A PREDIKSI TINGKAT KELULUSAN MAHASISWA KEPERAWATAN DENGAN PENDEKATAN NAÏVE BAYES. *Jurnal SIMADA (Sistem Informasi Dan Manajemen Basis Data)*, 7(2), 138–146. <https://doi.org/10.30873/simada.v7i2.639>

Putri, A. D. (2020). Prediksi Penerima Beasiswa dengan Metode C4.5. *Jurnal Teknologi Informasi*, 5(2), 112-120.

Rahman, Y. (2021). Klasifikasi Calon Penerima Beasiswa Menggunakan Naive Bayes. *Jurnal Informatika*, 8(1), 55-63.

Siregar, H. (2019). Penggunaan RapidMiner dalam Analisis Klasifikasi. *Jurnal Sistem Informasi*, 7(2), 88-95.

Zhang, H. (2004). The Optimality of Naive Bayes. *AAAI Conference on Artificial Intelligence*.