

# Penerapan Metode *Naïve Bayes* Untuk Klasifikasi Penerima Bantuan Stimulan Perumahan Swadaya (BSPS)

Alda Veronika Jona<sup>1</sup>, Era Pratiwi<sup>2</sup>, Hat Syahputra<sup>3</sup>, Yulia Safitri<sup>4</sup>

<sup>1,2,3,4</sup> Mahasiswa Program Studi Sistem Informasi, Sekolah Tinggi Manajemen Informatika dan Komputer Royal

<sup>1</sup> aquaona28@gmail.com\*; <sup>2</sup> erapradiwi020501@gmail.com; <sup>3</sup> hatsyahputra2@gmail.com; <sup>4</sup> yuliasafitrii266@gmail.com

\* Email Koresponden

## INFO ARTIKEL

### Histori Artikel

Diterima: 26/Juli/2024

Ditinjau: 28/Juli/2024

Disetujui: 31/Juli/2024

## ABSTRAK

Bantuan Stimulan Perumahan Swadaya (BSPS) adalah program bedah rumah dari pemerintah yang ditujukan kepada masyarakat berpenghasilan rendah. Program ini bertujuan untuk meningkatkan keswadayaan dalam pembangunan serta memperbaiki kualitas rumah, sarana, prasarana, dan utilitas umum melalui prinsip gotong-royong. Penerima bantuan BSPS harus memenuhi beberapa kriteria seperti penghasilan, status kepemilikan rumah, ukuran rumah, jenis lantai, jenis dinding, jenis atap, dan sumber air minum. Untuk menyelesaikan masalah sesuai kriteria tersebut, digunakan teknik Data Mining dengan metode Naive Bayes. Penelitian ini menggunakan dataset penerima BSPS di Desa Air Genting, Kecamatan Air Batu, yang terdiri dari 88 sampel. Hasil klasifikasi dalam penerapan Naive Bayes didapatkan nilai precision sebesar 84%, nilai recall 81%, nilai fl-score sebesar 82%, dan nilai akurasi sebesar 81%. Tujuan dari penelitian ini ialah untuk mempermudah mengklasifikasi warga yang layak dan berhak menerima bantuan pemerintah berupa BSPS. Hasil dari penelitian ini diharapkan bisa menjadi alternatif solusi dalam menentukan penerima bantuan BSPS di Desa Air Genting Kecamatan Air Batu.

**Kata Kunci:** BSPS, Data Mining, Naive Bayes



This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

Copyright ©2024 by Author. Published by PT Beranda Teknologi Academia

## ABSTRACT

The Self-Help Housing Stimulant Assistance (BSPS) is a government house renovation program aimed at low-income communities. This program aims to enhance self-sufficiency in construction and improve the quality of houses, facilities, infrastructure, and public utilities through the principle of mutual cooperation. BSPS recipients must meet several criteria, such as income, house ownership status, house size, floor type, wall type, roof type, and water source. To address issues based on these criteria, Data Mining techniques using the Naive Bayes method were employed. This study utilized a dataset of BSPS recipients in Air Genting Village, Air Batu Sub-district, comprising 88 samples. The classification results from applying Naive Bayes yielded a precision value of 84%, a recall value of 81%, an F1-score of 82%, and an accuracy of 81%. The objective of this research is to facilitate the classification of eligible and rightful recipients of government assistance in the form of BSPS. The results of this study are expected to provide an alternative solution in determining BSPS recipients in Air Genting Village, Air Batu Sub-district.

**Keywords:** BSPS, Data Mining, Naive Bayes

## PENDAHULUAN

Bantuan Stimulan Perumahan Swadaya (BSPS) adalah inisiatif pemerintah yang memberikan bantuan renovasi rumah untuk keluarga berpenghasilan rendah. Tujuan program ini adalah untuk meningkatkan kemandirian dalam pembangunan dan memperbaiki kualitas rumah, serta fasilitas dan utilitas umum. Dengan adanya bantuan ini, diharapkan dapat terwujud rumah yang layak huni, sehat, aman, harmonis, tertata, dan berkelanjutan [1].

Program BSPS di Desa Air Genting Kecamatan Air Batu dinilai tidak tepat sasaran saat proses seleksi warga yang layak dan tidak layak menerima bantuan. Sehingga timbulnya kecemburuan sosial dalam lingkungan masyarakat. Sistem manual digunakan dalam penentuan penerima bantuan tersebut dan objektif yang tentunya tidak efisien waktu. tidak adanya sistem dalam proses klasifikasi penerima program BSPS. Untuk resiko lainnya sering terjadi kehilangan dokumen. Sehingga, penyaluran program BSPS ini dinilai kurang tepat.

Sehubungan masalah tersebut, Desa Air Genting Kecamatan Air Batu memerlukan sistem yang dapat mempermudah kelayakan penerima Bantuan Stimulan Perumahan Swadaya (BSPS) secara tepat. Masalah penentuan kelayakan dapat diatasi dengan penerapan data mining menggunakan metode Algoritma Naïve Bayes [2]. Berikut adalah ringkasan penelitian sebelumnya yang telah dilakukan: penelitian oleh [3] dengan judul “Penerapan Data Mining untuk Mengklasifikasikan Penerima Bantuan PKH Desa Wae Jare Menggunakan Metode Naïve Bayes” menunjukkan bahwa penerapan metode Naïve Bayes dalam data mining menghasilkan tingkat akurasi sebesar 82,14%.

Dalam penelitian lain yang dilakukan oleh [4] berjudul “Klasifikasi Masyarakat Miskin Menggunakan Metode Naïve Bayes,” ditemukan bahwa metode klasifikasi Naive Bayes pada dataset yang digunakan mencapai tingkat akurasi sebesar 73%, yang termasuk dalam kategori Klasifikasi Baik. Selain itu, penelitian oleh [5] dengan judul "Sistem Pendukung Keputusan Penerimaan Bantuan Sosial Program Keluarga Harapan (PKH) Menggunakan Metode Naïve Bayes Classifier (Studi Kasus di Balai Desa Bendungan Kraton Pasuruan)" menunjukkan bahwa penerapan metode Naïve Bayes Classifier pada 50 data latih dan 10 data uji menghasilkan akurasi sebesar 80%. Ini menandakan bahwa metode Naïve Bayes Classifier efektif dalam menentukan penerimaan bantuan sosial.

Berdasarkan dari beberapa penelitian sebelumnya, pada penelitian ini, yang dilakukan di Desa Air Genting Kecamatan Air Batu bertujuan untuk mengklasifikasi warga yang layak dan berhak menerima bantuan pemerintah berupa dana BSPS berdasarkan pada pengalaman di masa sebelumnya. Hasil dari penelitian ini diharapkan dapat menjadi alternatif solusi dalam menentukan penerima bantuan BSPS di Desa Air Genting Kecamatan Air Batu.

## METODE

Data mining adalah proses untuk mengidentifikasi korelasi atau hubungan baru dalam bentuk pola atau tren dari kumpulan data besar yang disimpan di repositori, dengan memanfaatkan teknologi penalaran pola serta teknik statistik dan matematika [6]. Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan pembelajaran mesin untuk menemukan pengetahuan berharga yang terdapat dalam data besar. Klasifikasi adalah salah satu teknik yang diterapkan dalam data mining.

Alur metode dalam penelitian ini yaitu *Cross-Standard Industry for Data Mining (CRISP-DM)*. Metode ini merupakan metodologi standar dalam penelitian data mining. Terdapat 6 tahap proses (CRISP-DM) yaitu:



Gambar 1. Tahapan CRISP-DM

**Business Understanding**

BSPS merupakan salah satu program pemerintah berupa dukungan dana bedah rumah untuk masyarakat yang berpenghasilan rendah (MBR) guna meningkatkan kualitas rumah swadaya berlandaskan gotong royong. Seperti pada warga yang ada di desa Air Genting Kecamatan Air Batu. Pada penelitian ini akan dilakukan analisis bagaimana membuat klasifikasi untuk menentukan warga yang layak maupun tidak layak yang menerima bantuan BSPS sehingga dana bantuan yang disalurkan tepat pada sasaran.

**Data Understanding**

Sumber data diperoleh dari Dinas Sosial Kabupaten Asahan, khususnya dari Desa Air Genting di Kecamatan Air Batu. Data tersebut mencakup atribut seperti Nama, Kepemilikan Rumah, Ukuran Rumah, Jenis Lantai, Jenis Dinding, Jenis Atap, Sumber Air Minum, dan Keputusan.

```
df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 88 entries, 0 to 87
Data columns (total 10 columns):
#   Column                Non-Null Count  Dtype
---  ---                ---
0   No                     88 non-null    int64
1   Nama                   88 non-null    object
2   Penghasilan            88 non-null    int64
3   Status Kepemilikan Rumah 88 non-null    object
4   Ukuran Rumah          88 non-null    object
5   Jenis Lantai           88 non-null    object
6   Jenis Dinding          88 non-null    object
7   Jenis Atap             88 non-null    object
8   Sumber Air Minum       88 non-null    object
9   Keputusan              88 non-null    object
dtypes: int64(2), object(8)
memory usage: 7.0+ KB
```

Gambar 2. Informasi Dataset

**Data Preparation**

Dalam persiapan data ada beberapa hal yang harus dilakukan yaitu mengumpulkan data, mengolah data, menggabungkan data dan melakukan pembersihan data.

```
df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 88 entries, 0 to 87
Data columns (total 8 columns):
#   Column                Non-Null Count  Dtype
---  ---                ---
0   Penghasilan            88 non-null    int64
1   Status Kepemilikan Rumah 88 non-null    int32
2   Ukuran Rumah          88 non-null    int32
3   Jenis Lantai           88 non-null    int32
4   Jenis Dinding          88 non-null    int32
5   Jenis Atap             88 non-null    int32
6   Sumber Air Minum       88 non-null    int32
7   Keputusan              88 non-null    object
dtypes: int32(6), int64(1), object(1)
memory usage: 3.6+ KB
```

Gambar 3. Data Yang Akan Diolah

**Modelling**

Teorema Naive Bayes, yang diperkenalkan oleh ilmuwan Inggris Thomas Bayes, digunakan untuk memprediksi kemungkinan di masa depan berdasarkan pengalaman sebelumnya dengan menerapkan teknik klasifikasi serta metode probabilitas dan statistik. Klasifikasi Naive Bayes mengasumsikan bahwa adanya atau tidak adanya fitur tertentu dalam suatu kelas tidak dipengaruhi oleh fitur dari kelas lain. Klasifikasi Bayesian terbukti efektif dengan akurasi dan kecepatan tinggi ketika diterapkan pada database dengan volume data yang besar. Teorema ini menggunakan metode Naive yang mengasumsikan kondisi antar atribut bersifat independen [7]. Dilanjutkan dengan pengecekan atribut yang paling berpengaruh dalam pengklasifikasian [8].

Persamaan dari *teorema Bayes* yaitu :

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

Keterangan :

- X : Data dengan kelas yang belum diketahui
- H : Hipotesis data merupakan suatu *class* spesifik
- P(H|X) : Probabilitas Hipotesis H berdasarkan kondisi X (posterior probabilitas)
- P(X|H) : Probabilitas X berdasarkan kondisi pada Hipotesis H
- P(H) : Probabilitas Hipotesis H (*prior probabilitas*)
- P(X) : Probabilitas X

Dari total 88 data yang ada, akan dilakukan *split* data yaitu membagi data menjadi 2 bagian. Data *training* dan data *testing* dengan porsi perbandingan 70 : 30

x_train.shape
(61, 7)
x_test.shape
(27, 7)

Gambar 4. *Split* Data

**Evaluation**

Pada langkah ini akan dilakukan pengukuran tingkat keakuratan hasil dan validasi yang telah dicapai dengan menggunakan *Confusion Matrix* [9]. *Confusion Matrix* meliputi true positive, false positive, true negative, dan false negative, yang digunakan untuk menghitung nilai akurasi, presisi, dan recall. Akurasi menilai seberapa dekat nilai prediksi dengan nilai sebenarnya. Presisi mengukur sejauh mana jawaban yang diberikan sistem sesuai dengan informasi yang diminta pengguna. Sedangkan recall menilai kemampuan sistem dalam mengidentifikasi kembali informasi yang relevan [10]. *Confusion matrix* akan menampilkan tabel yang menggambarkan data uji yang benar dan data uji yang salah [11].

Untuk menghitung *accuracy*, *precision*, dan *recall* dapat menggunakan persamaan berikut :

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

**Deployment**

Hasil penerapan algoritma dalam proses data mining divisualisasikan dengan menggunakan bahasa pemrograman sehingga dapat dipahami oleh masyarakat umum [12].

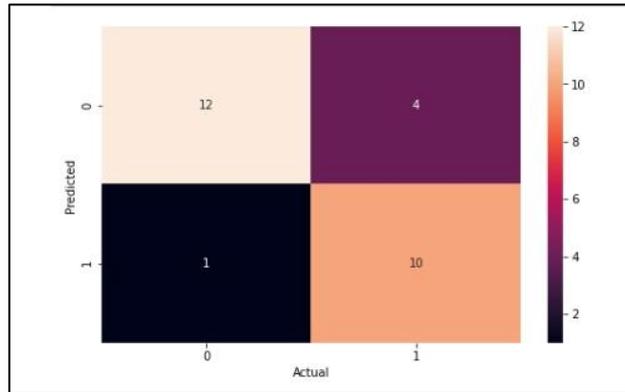
**HASIL DAN PEMBAHASAN**

Tahap pertama adalah analisis, yang melibatkan pengimporan modul dan data yang diperlukan. Kemudian lakukan *Encoding* pada dataset.

	Penghasilan	Status Kepemilikan Rumah	Ukuran Rumah	Jenis Lantai	Jenis Dinding	Jenis Atap	Sumber Air Minum	Keputusan
0	1500000	0	0	1	1	2	1	Layak
1	1400000	0	1	1	1	2	1	Layak
2	1400000	0	0	1	1	2	1	Layak
3	1400000	0	1	1	1	2	1	Layak
4	1500000	0	1	1	1	0	1	Layak

Gambar 5. Data *Encoding*

Setelah itu lakukan split data menjadi dua bagian yaitu data *training* dan data *testing* dengan porsi 70:30. Kemudian lakukan pengujian dengan data *testing*. Maka diperoleh *confusion matrix* dari perbandingan seperti pada gambar berikut :



Gambar 6. *Visualisasi Confusion Matrix*

	precision	recall	f1-score	support
Layak	0.92	0.75	0.83	16
Tidak Layak	0.71	0.91	0.80	11
accuracy			0.81	27
macro avg	0.82	0.83	0.81	27
weighted avg	0.84	0.81	0.82	27

Gambar 7. Tingkat Akurasi

Dengan tingkat akurasi berdasarkan pada Gambar 7 didapatkan hasil algoritma *Naive Bayes* dengan nilai *precision* untuk kelas positif sebesar 92%, untuk kelas negative 71%, untuk nilai *recall* kelas positif 75%, untuk kelas negative 91%, untuk nilai *f1-score* kelas positive sebesar 83%, kelas negative sebesar 80%, dan untuk nilai akurasi sebesar 81% dalam metode algoritma *Naive Bayes*.

## KESIMPULAN

Berdasarkan pembahasan di atas, penerapan algoritma Naive Bayes dan model Confusion Matrix menghasilkan nilai presisi sebesar 84%, recall sebesar 81%, F1-score sebesar 82%, dan akurasi sebesar 81%, yang termasuk dalam kategori Klasifikasi Baik. Dengan demikian, metode Naive Bayes dianggap efektif dalam memprediksi peluang berdasarkan pengalaman sebelumnya. Penerapan metode ini mempermudah proses klasifikasi calon penerima program dana bantuan BPS di Desa Air Genting, Kecamatan Air Batu.

## DAFTAR PUSTAKA

- [1] S. Margareth, I. Suharjo, and M. Eng, "SPK Penentuan Penerima Bantuan Stimulan Perumahan Swadaya di Kecamatan Maligano Menggunakan Metode SAW DSS For Determining Recipient Of Stimulant Assistance Of Selft-Help Housing In Maligano District Suing The Saw Method," *Smai*, pp. 160–167, 2020.
- [2] A. A. A. Arifin, W. Handoko, and Z. Efendi, "Implementasi Metode Naive Bayes Untuk Klasifikasi Penerima Program Keluarga Harapan," *J-Com (Journal Comput.*, vol. 2, no. 1, pp. 21–26, 2022, doi: 10.33330/j-com.v2i1.1577.
- [3] A. Ifon Purnama, A. Aziz, A. Sartika Wiguna, and K. Kunci, "Penerapan Data Mining Untuk Mengklasifikasi Penerima Bantuan PKH Desa Wae Jare Menggunakan Metode Naive Bayes," *Kurawal J. Teknol. Inf. dan Ind.*, vol. 3, pp. 1–8, 2020, [Online]. Available: <https://jurnal.machung.ac.id/index.php/kurawal>
- [4] H. Annur, "Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes," *Ilk. J. Ilm.*, vol. 10, no. 2, pp. 160–165, 2018, doi: 10.33096/ilkom.v10i2.303.160-165.
- [5] A. A. Saputro, "Sistem Pendukung Keputusan Penerimaan Bantuan Sosial Program Keluarga Harapan ( PKH ) dengan Menggunakan Metode Naive Bayes Classifier ( Studi Kasus di Balai Desa Bendungan Kraton Pasuruan )," *J. Ilm. Edutic*, vol. 9, no. 1, pp. 40–48, 2022.
- [6] E. Fitriani, "Perbandingan Algoritma C4.5 Dan Naive Bayes Untuk Menentukan Kelayakan Penerima Bantuan Program Keluarga Harapan," *Sistemasi*, vol. 9, no. 1, p. 103, 2020, doi: 10.32520/stmsi.v9i1.596.
- [7] I. P. Ninditama, I. P. Ninditama, W. Cholil, M. Akbar, and D. Antoni, "Klasifikasi Keluarga Sejahtera Study Kasus : Kecamatan Kota Palembang," *J. TEKNO KOMPAK*, vol. 15, no. 2, pp. 37–49, 2020, [Online]. Available: <https://ejurnal.teknokrat.ac.id/index.php/teknokompak/article/view/1156>
- [8] Meilaniwati Essy R, "Klasifikasi penduduk miskin penerima PKH menggunakan metode naive bayes dan KNN," *J. Kaji. dan Terap. Mat.*, vol. 8, pp. 114–128, 2023.
- [9] N. Alfiah, "Klasifikasi Penerima Bantuan Sosial Program Keluarga Harapan Menggunakan Metode Naive Bayes," *Respati*, vol. 16, no. 1, p. 32, 2021, doi: 10.35842/jtir.v16i1.386.
- [10] A. Damuri, U. Riyanto, H. Rusdianto, and M. Aminudin, "Implementasi Data Mining dengan Algoritma Naive Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako," *J. Ris. Komput.*, vol. 8, no. 6, pp. 219–225, 2021, doi: 10.30865/jurikom.v8i6.3655.
- [11] D. Utami, P. Aisyiyah, and R. Devi, "Klasifikasi kelayakan penerima bantuan program keluarga harapan (pkh) menggunakan metode weighted naive bayes dengan laplace smoothing," *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 7, pp. 1373–1384, 2022.
- [12] S. F. Wulandari, W. Handoko, and U. B. B. F. Agis, "Perbandingan Naive Bayes Dan C45 Dalam Klasifikasi Tes Kesehatan Mahasiswa Baru Akbid As-Syifa," *JUTSI (Jurnal Teknol. dan Sist. Informasi)*, vol. 2, no. 3, pp. 167–176, 2022.