



## Implementasi Data Mining Dengan Naïve Bayes Untuk Prediksi Penerima Dana BOS Di Sekolah X

Wulan Dari

Universitas Potensi Utama, Medan

\*Email : [ulandari2796@gmail.com](mailto:ulandari2796@gmail.com)

**Abstract:** Schools will receive assistance from recipients of boss funds at school X implementing a program from the government to provide BOS assistance to underprivileged school children in School X. The Bos Assistance Program is a government assistance program in the form of providing cash to underprivileged school children or families capable. Implementation of the program aims to improve the welfare of the less fortunate. Currently, the distribution of Bos Fund Assistance is considered to be still not on target and does not meet the criteria that are used as benchmarks for the people who will receive the assistance. The Naïve Bayes method is used in this case study to classify which children can and cannot afford school fees so that it can more easily select school children or families who are truly eligible to receive the Bos Fund Assistance. This study uses 20 data and 5 criteria including: name, age, gender, status, income, and beneficiary status. After conducting research with a total of 20 data and using the Naïve Bayes method, the results obtained were as many as 45% of underprivileged children and families who were eligible to receive Boss Fund Assistance and as many as 55% of children and families who were not eligible to receive Assistance Boss Fund..

**Keywords:** Data Mining, Naïve Bayes Algorithm, Bos Funding Assistance

**Abstrak** Sekolah akan mendapatkan bantuan penerima dana bos di sekolah X melaksanakan program dari pemerintah untuk memberikan Bantuan BOS kepada anak-anak sekolah yang kurang mampu yang ada di Sekolah X. Program Bantuan Bos adalah program bantuan pemerintah dalam bentuk pemberian uang tunai bagi anak sekolah atau keluarga kurang mampu. Pelaksanaan program tersebut bertujuan guna meningkatkan kesejahteraan masyarakat yang kurang mampu. Saat ini pembagian Bantuan Dana Bos dinilai masih tidak tepat sasaran dan belum memenuhi kriteria yang menjadi tolak ukur dari masyarakat yang akan menerima bantuan tersebut. Metode Naïve Bayes digunakan dalam studi kasus ini untuk klasifikasi mana anak yang mampu dan tidak mampu dalam biaya sekolah agar dapat lebih mudah menyeleksi anak-anak sekolah atau keluarga yang benar-benar layak untuk menerima Bantuan Dana Bos. Penelitian ini menggunakan 20 data dan 5 kriteria diantaranya :nama, usia, jenis kelamin, status, penghasilan, dan status penerima. Setelah dilakukan penelitian dengan jumlah data sebanyak 20 dan menggunakan metode Naïve Bayes, maka diperoleh hasil berupa sebanyak 45% anak-anak dan kluarga yang kurang mampu yang masuk kedalam kriteria layak menerima Bantuan Dana Bos dan sebanyak 55% anak dan keluarga tidak layak dalam menerima Bantuan Dana Bos.

**Kata Kunci:** Data Mining, Algoritma Naïve Bayes, Bantuan Dana Bos

## **PENDAHULUAN**

Dana Bantuan Operasional Sekolah yang biasa disebut dengan Dana BOS adalah program bantuan pemerintah dalam bentuk pemberian uang tunai bagi sekolah-sekolah di seluruh Indonesia. Mengutip laman Direktorat Sekolah Dasar, dana Bantuan Operasional Satuan Pendidikan atau biasa disebut dana BOS adalah dana alokasi khusus non fisik untuk mendukung biaya operasional non personalia bagi satuan pendidikan. Dana BOS diberikan kepada sekolah-sekolah baik yang dikelola oleh pemerintah maupun swasta. Dana ini digunakan untuk biaya operasional sekolah seperti gaji guru dan karyawan, kebutuhan belajar mengajar seperti buku dan alat tulis, serta keperluan lainnya seperti biaya listrik, air, dan perawatan gedung sekolah. Setiap sekolah di Indonesia berhak menerima dana BOS sesuai dengan jumlah siswa yang terdaftar di sekolah tersebut. Besaran dana yang diterima oleh setiap sekolah juga bervariasi tergantung dari tingkatan sekolah, yakni SD, SMP, atau SMA. [1].

Saat ini pembagian Dana Bos dinilai masih tidak tepat sasaran dan belum memenuhi kriteria yang menjadi tolak ukur dari masyarakat yang akan menerima bantuan tersebut. Pelaksanaan program tersebut bertujuan guna meningkatkan keunggulan pendidikan yang tidak mampu. Untuk meminimalisir angka kemiskinan yang ada, maka pemerintah menjalankan program pemberian dana Bos kepada sekolah dan anak didik yang kurang mampu di Sekolah X. Dalam melakukan kualifikasi bantuan dana Bos di Sekolah X pasti akan mengalami beberapa kendala dan kesulitan jika tidak menggunakan suatu metode dikarenakan banyaknya masyarakat yang membutuhkan bantuan tersebut[2].

Sekolah X terletak di Kecamatan Medan Skip, Kota Medan Sumatera Utara. Dengan penduduk sebanyak 500 anak, masih cukup banyak keluarga yang kurang mampu yang belum menerima bantuan dana Bos dikarenakan bantuan yang diberikan tidak menyeluruh atau bahkan penyaluran dana Bos tidak tepat sasaran yang tidak memenuhi keluarga penerima bantuan dana Bos tersebut.

Data mining merupakan sebuah cara menemukan gambaran otomatis atau semi otomatis melalui data yang sudah ada di dalam database untuk diselesaikan dengan berbagai teknik. Fungsi klasifikasi adalah salah satu teknik yang ada pada data mining. Fungsi klasifikasi bertujuan memudahkan dalam proses mengidentifikasi pihak yang termasuk kedalam kriteria masyarakat penerima bantuan. Teknik klasifikasi merupakan sebuah teknik yang ada dalam pembelajaran data mining yang cara kerjanya dengan memprediksi nilai dari suatu atribut kategori[3].

Metode Naive Bayes Classifier sangat berpotensi dalam mengklasifikasikan dokumen dibandingkan dengan metode pengklasifikasian yang lain dalam keakuratan dan efisiensi komputasi[10].

Penelitian ini bertujuan untuk memudahkan dalam mengklasifikasikan keluarga dan anak yang bersekolah di Sekolah X yang berhak menerima Bantuan dana Bos agar bantuan yang diberikan tepat sasaran kepada keluarga anak yang bersekolah di Sekolah X yang membutuhkan. Setelah dilakukannya penelitian, maka diperoleh hasil sebesar 45% anak yang layak menerima dan sebesar 55% anak yang masuk namun tidak memenuhi kriteria dalam menerima Bantuan dana Bos.

## **METODOLOGI**

### **A. Jenis Penelitian**

Penelitian ini menggunakan jenis penelitian deskriptif melalui data kualitatif. Penelitian deskriptif merupakan metode yang memanfaatkan data kualitatif berupa simbol angka untuk selanjutnya dijelaskan secara deskriptif[4]. Selain dengan menggunakan metode penelitian deskriptif, penelitian ini juga menggunakan metode klasifikasi dan Algoritma Naïve Bayes.

#### **1. Klasifikasi**

Klasifikasi merupakan sebuah teknik yang ada dalam pembelajaran data mining yang cara kerjanya dengan memprediksi nilai dari suatu atribut kategori[3]. Algoritma klasifikasi yang banyak digunakan secara luas, yaitu Decision/classification trees, Bayesian classifiers/ Naïve Bayes classifiers, Neural networks, Analisa Statistik, Algoritma Genetika, Rough sets, k-nearest neighbor, Metode Rule Based, Memory based reasoning, dan Support vector machines (SVM)[5].

#### **2. Algoritma Naïve Bayes**

Naïve bayes adalah cara klasifikasi yang ditemukan oleh Thomas Bayes, Ilmuan Inggris. Naïve bayes dapat memprediksi berbagai peluang yang akan terjadi pada masa mendatang berdasarkan kejadian ataupun pengalaman di masa lalu. Naïve bayes juga dikenal dengan Theorema Bayes [8]. Dasar Teorema Bayes adalah menangani permasalahan bersifat hipotesis dengan mendesain suatu klasifikasi untuk memisahkan objek[11]. Keuntungan yang dapat diperoleh dari penggunaan Naïve bayes adalah tidak membutuhkan data training yang besar untuk menentukan perkiraan kriteria dalam proses mengklasifikasi[6]. Bayesian classification terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar [7].

Naïve Bayes memiliki dasar persamaan sebagai berikut [8]:

Dimana :

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

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

$P(X)$  = Probabilitas X

## B. Jenis Pengumpulan Data

Pada penelitian ini metode yang digunakan untuk mengumpulkan data-data adalah Literatur Studi Pustaka Penelitian ini menggunakan literatur studi pustaka untuk dijadikan kajian teoritis sebagai sumber referensi untuk memperoleh data-data terkait seperti, jurnal, buku, makalah dan artikel yang membahas tentang Bantuan Dana Bos, data mining, klasifikasi, serta algoritma naïve bayes.

Dari seluruh data yang telah dikumpulkan, untuk menghasilkan data baru yang Layak atau tidaknya di sekolah X dalam menerima Bantuan Dana Bos, maka diperlukan Microsoft Excel untuk mencari probabilitas serta menghitung peluang pada masing-masing atribut[9]

## 1. HASIL DAN PEMBAHASAN

Untuk menentukan kelayakan penerima Bantuan Dana Bosdi Sekolah X dengan menggunakan Microsoft Excel beserta data dan kategori sebagai berikut:

**Tabel 1**

Nama	Usia	JK	Status	Penghasilan	Status Penerima
Adi	40	Pria	Tidak Bekerja	Rendah	Layak
Sulas	45	Wanita	Tidak Bekerja	Rendah	Layak
Bobi	34	Pria	Tidak Bekerja	Cukup	Tidak Layak
Ana	30	Wanita	Bekerja	Cukup	Tidak Layak
Ifan	30	Pria	Bekerja	Rendah	Layak
Nani	50	Wanita	Bekerja	Cukup	Layak
Kiki	45	Wanita	Bekerja	Rendah	Layak
Siti	35	Wanita	Bekerja	Cukup	Tidak Layak
Angga	35	Pria	Bekerja	Cukup	Tidak Layak
Rizal	30	Pria	Bekerja	Cukup	Tidak Layak
Mimi	40	Wanita	Tidak Bekerja	Rendah	Layak
Ratna	50	Wanita	Bekerja	Cukup	Tidak Layak
Budi	45	Pria	Tidak Bekerja	Rendah	Layak
Dodi	36	Pria	Bekerja	Cukup	Tidak Layak
Lia	36	Wanita	Tidak Bekerja	Rendah	Layak
Indra	30	Pria	Bekerja	Cukup	Tidak Layak
Ulan	34	Wanita	Bekerja	Cukup	Tidak Layak
Galih	35	Pria	Bekerja	Cukup	Tidak Layak
Wawan	30	Pria	Bekerja	Cukup	Tidak Layak
Tari	45	Wanita	Bekerja	Rendah	?

Pada tabel 1 terdapat 20 data yang berisikan nama, usia, jenis kelamin, status, penghasilan, dan status penerima sekolah X dalam menerima Bantuan Dana Bos. Namun dari 20 data yang ada, terdapat 1 data yang masih belum diketahui status penerima. Untuk menentukan kelayakan anak atau keluarga tersebut dalam menerima Bantuan Dana Bos, maka diperlukan software Microsoft Excel untuk menghitung jumlah class atau label serta mencari peluang pada masing-masing atribut/kriteria diatas.

Beberapa tahapan yang harus dilakukan diantaranya adalah :

**1. Menghitung Jumlah Class/Label**

Untuk menghitung jumlah class atau label dapat dilakukan dengan cara menghitung jumlah sel pada Status penerima Layak dan Tidak Layak yang memenuhi kriteria.

**Tabel 2**

Label	Jumlah data dan Jumlah seluruh data		Hasil
Layak	8	19	0.4
Tidak Layak	11	19	0.6

Pada saat akan menghitung jumlah class atau label yang digunakan adalah sel status penerima. Hal tersebut karena dari data-data keseluruhan pada tabel, yang belum diketahui kejelasannya adalah sel status penerima. Dari tabel 2 dapat dilihat bahwa dari 19 data keseluruhan pada kriteria status penerima terdapat 2 pembagian data berupa Layak sebanyak 8 data dan Tidak Layak sebanyak 11 data dengan masing-masing hasil 0.4 dan 0.6.

**2. Mencari Peluang Pada Masing-Masing Atribut/Kriteria**

Untuk mencari peluang pada masing-masing atribut kriteria dengan cara menerapkan kriteria ke sel Usia dengan menghitung kriteria yang dipenuhi pada sel Status penerima

**Tabel 3**

Usia	Layak	Tidak Layak	P(Layak)	P(TidakLayak)
50	1	1	0.13	0.09
45	3	0	0.38	0.00
40	2	0	0.25	0.00
36	1	1	0.13	0.09
35	0	3	0.00	0.27
34	0	2	0.00	0.18
30	1	4	0.13	0.36
Total	8	11	1.00	1.00

Pada tabel kriteria usia keluarga Sekolah X terbagi menjadi 7 data, yaitu : 50,45,40,36,35,34, dan 30. Pada data 50 terdapat 1 data layak dengan probabilitas 0.13 sedangkan data tidak layak terdapat 1 data dengan probabilitas 0.09. Pada data 45 terdapat 3 data layak dengan probabilitas 0.38 sedangkan data tidak layak terdapat 0 data dengan probabilitas 0.00. Pada data 40 terdapat 2 data layak dengan probabilitas 0.25 sedangkan data tidak layak terdapat 0 data dengan probabilitas 0.00. Pada data 36 terdapat 1 data layak dengan probabilitas 0.13 sedangkan data tidak layak terdapat 1 data dengan probabilitas 0.09. Pada data 35 terdapat 0 data layak dengan probabilitas 0.00 sedangkan data tidak layak terdapat 3 data dengan probabilitas 0.27. Pada data 34 terdapat 0 data layak dengan probabilitas 0.00 sedangkan data tidak layak terdapat 2 data dengan probabilitas 0.18. Pada data 30 terdapat 1 data layak dengan probabilitas 0.13 sedangkan data tidak layak terdapat 4 data dengan probabilitas 0.36. Total data usia pada layak sebanyak 8 dengan probabilitas 1.00 dan tidak layak sebanyak 11 data dengan probabilitas 1.00.

**Tabel 4**

JK	Layak	Tidak Layak	P(Layak)	P(TidakLayak)
Pria	3	7	0.38	0.64
Wanita	5	5	0.63	0.36
Total	8	11	1.00	1.00

Pada tabel kriteria jenis kelamin keluarga Sekolah X terbagi menjadi 2 data, yaitu : pria dan wanita. Pada data pria terdapat 3 data layak dengan probabilitas 0.38 sedangkan data tidak layak terdapat 7 data dengan probabilitas 0.64. Pada data wanita terdapat 5 data layak dengan probabilitas 0.63 sedangkan data tidak layak terdapat 5 data dengan probabilitas 0.36. Total data jenis kelamin pada layak sebanyak 8 dengan probabilitas 1.00 dan tidak layak sebanyak 11 data dengan probabilitas 1.00.

**Tabel 5**

Status	Layak	Tidak Layak	P(Layak)	P(TidakLayak)
Bekerja	5	6	0.63	0.55
Tidak Bekerja	3	5	0.38	0.45
Total	8	10	1.00	1.00

Pada tabel kriteria status keluarga Sekolah X terbagi menjadi 2 data, yaitu : Bekerja dan Tidak Bekerja. Pada data Bekerja terdapat 5 data layak dengan probabilitas 0.63 sedangkan data tidak layak terdapat 6 data dengan probabilitas 0.55. Pada data tidak bekerja terdapat 3 data layak dengan probabilitas 0.38 sedangkan data tidak layak terdapat 5 data dengan probabilitas 0.45. Total data status pada layak sebanyak 8 dengan probabilitas 1.00 dan tidak layak sebanyak 10 data dengan probabilitas 1.00.



**Tabel 6**

Penghasilan	Layak	Tidak Layak	P(Layak)	P(TidakLayak)
Cukup	1	11	0.13	1.00
Rendah	7	0	0.88	0.00
Total	8	11	1.00	1.00

Pada tabel kriteria penghasilan keluarga Sekolah X terbagi menjadi 2 data, yaitu : cukup dan rendah. Pada data cukup terdapat 1 data layak dengan probabilitas 0.13 sedangkan data tidak layak terdapat 11 data dengan probabilitas 1.00. Pada data rendah terdapat 7 data layak dengan probabilitas 0.88 sedangkan data tidak layak terdapat 0 data dengan probabilitas 0.00. Total data jenis kelamin pada layak sebanyak 8 dengan probabilitas 1.00 dan tidak layak sebanyak 11 data dengan probabilitas 1.00.

## 2. Menentukan Status Kriteria Akhir

Untuk menentukan status kriteria akhir, jumlahkan masing-masing kriteria yang diminta pada data terkait dengan pengkalian dan data dengan hasil terbesar adalah data yang menjadi status kriteria penerima.

**Tabel 7**

Layak	0.38	0.63	0.38	0.88	<b>0.08</b>
Tidak Layak	0.00	0.36	0.45	0.00	<b>0.00</b>

Pada tabel 7, sel yang digunakan untuk menentukan status kelayakan penerima adalah probabilitas layak dan tidak layak pada usia, jenis kelamin, status, dan penghasilan. Setelah keseluruhan data dijumlahkan dengan pengkalian, data terbesar dengan nilai 0.08 terletak pada kriteria layak sedangkan data terkecil dengan nilai 0.00 terletak pada kriteria tidak layak. Oleh karena itu status kriteria akhir penerima Bantuan Dana Bos warga dengan nama Tari, usia 45 tahun, jenis kelamin Wanita, status Single, serta penghasilan Rendah, memiliki status penerima LAYAK seperti tabel berikut :

Nama	Usia	JK	Status	Penghasilan	Status Penerima
Adi	40	Pria	Tidak Bekerja	Rendah	Layak
Sulas	45	Wanita	Tidak Bekerja	Rendah	Layak
Bobi	34	Pria	Tidak Bekerja	Cukup	Tidak Layak
Ana	30	Wanita	Bekerja	Cukup	Tidak Layak
Ifan	30	Pria	Bekerja	Rendah	Layak
Nani	50	Wanita	Bekerja	Cukup	Layak
Kiki	45	Wanita	Bekerja	Rendah	Layak
Siti	35	Wanita	Bekerja	Cukup	Tidak Layak
Angga	35	Pria	Bekerja	Cukup	Tidak Layak
Rizal	30	Pria	Bekerja	Cukup	Tidak Layak
Mimi	40	Wanita	Tidak Bekerja	Rendah	Layak
Ratna	50	Wanita	Bekerja	Cukup	Tidak Layak
Budi	45	Pria	Tidak Bekerja	Rendah	Layak
Dodi	36	Pria	Bekerja	Cukup	Tidak Layak
Lia	36	Wanita	Tidak Bekerja	Rendah	Layak
Indra	30	Pria	Bekerja	Cukup	Tidak Layak
Ulan	34	Wanita	Bekerja	Cukup	Tidak Layak
Galih	35	Pria	Bekerja	Cukup	Tidak Layak
Wawan	30	Pria	Bekerja	Cukup	Tidak Layak
<b>Tari</b>	<b>45</b>	<b>Wanita</b>	<b>Bekerja</b>	<b>Rendah</b>	<b>Layak</b>

## KESIMPULAN

Berdasarkan hasil dan pembahasan yang dijabarkan dapat disimpulkan bahwa data mining merupakan sebuah cara menemukan gambaran otomatis atau semi otomatis melalui data yang sudah ada di dalam database untuk diselesaikan dengan berbagai teknik. Penerapan Algoritma Naïve Bayes dalam menentukan status kelayakan penerima Bantuan Dana Bos dengan bantuan Microsoft Excel sangat memudahkan proses klasifikasi dibandingkan harus menghitung satu-persatu kriteria beserta probabilitas yang ada.

Dengan menerapkan Algoritma Naïve Bayes, diperoleh hasil berupa sebanyak 45% keluarga Sekolah X layak dalam menerima Dana Bos, sedangkan 55% keluarga Sekolah X tidak layak dalam menerima Dana Bos walaupun masuk kedalam kriteria namun tidak memenuhi secara keseluruhan. Metode Algoritma Naïve Bayes terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar

## REFERENSI

- Kemiskinan, T. N. P. P. (2018). Program Bantuan Pemerintah Untuk Individu, Keluarga, dan Kelompok Tidak Mampu Menuju Bantuan Sosial Terintegrasi, Jakarta: TNP2K.
- Wiyanto, W., & Atmaja, S. (2018). Penerapan Algoritma Naive Bayes Untuk Pemilihan Keluarga Yang Membutuhkan Bantuan Dalam Program Keluarga Harapan (Pkh)(Studi Kasus Di Desa Karang Asih, Cikarang Utara). *Jurnal SIGMA*, 9(1), 31-44.
- Sugianto, C. A., & Maulana, F. R. (2019). Algoritma Naïve Bayes Untuk Klasifikasi Penerima Bantuan Pangan Non Tunai (Studi Kasus Kelurahan Utama). *Techno. Com*, 18(4), 321331.
- Annur, H. (2018). Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes. *ILKOM Jurnal Ilmiah*, 10(2), 160-165.
- Jananto, A. (2013). Algoritma Naive Bayes untuk Mencari Perkiraan Waktu Studi Mahasiswa. *Dinamik*, 18(1).
- Nurmayanti, W. P., Saky, D. A. L., Malthuf, M., Gazali, M., & Hirzi, R. H. (2021). Penerapan Naïve Bayes dalam Mengklasifikasikan Masyarakat Miskin di Desa Lepak. *Geodika: Jurnal Kajian Ilmu dan Pendidikan Geografi*, 5(1), 123-132.
- Purnama, P., & Supriyanto, C. (2013). Deteksi Penyakit Diabetes Type II Dengan Naive Bayes Berbasis Particle Swarm Optimization. *jurnal teknologi informasi*, 9, 49-53.
- Bustami, B. (2013). Penerapan algoritma Naive Bayes untuk mengklasifikasi data nasabah asuransi. *TECHSI-Jurnal Teknik Informatika*, 5(2).
- Asmiati, N. (2020). Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Pengaruh Negatif Game online Bagi Remaja Milenial. *JTIM: Jurnal Teknologi Informasi Dan Multimedia*,

2(3), 141-149.

Alita, D., Sari, I., Isnain, A. R., & Styawati, S. (2021). Penerapan Naïve Bayes Classifier Untuk Pendukung Keputusan Penerima Beasiswa. *Jurnal Data Mining Dan Sistem Informasi*, 2(1), 17-23.

Ridwan, M., Suyono, H., & Sarosa, M. (2013). Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. *Jurnal EECCIS (Electrics, Electronics, Communications, Controls, Informatics, Systems)*, 7(1), 59-64.