

Klasifikasi Sosial Ekonomi Menggunakan *Naïve Bayes Classifier*

Irmayani^{a*}, Budyanita Asrun^b

^aTeknik Sipil, Akademi Teknologi Industri Dewantara Palopo, Jalan KH. Ahmad Razak 2 No. 7, Wara Selatan, Kota Palopo, Sulawesi Selatan, Indonesia

^bBioinformatika, Universitas Megarezky, Jalan Antang Raya No. 45, Kota Makassar, Sulawesi Selatan, Indonesia

*Email : irnia2104@gmail.com

Abstrak

Data mining meliputi beberapa metode untuk membantu pengambilan keputusan salah satunya adalah metode klasifikasi. Metode klasifikasi meliputi beberapa cara salah satunya adalah *Naive Bayes Classifier*. Model *Naive Bayes* didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi serupa dengan *decision tree*. Pemanfaatan metode klasifikasi *Naive Bayes Classifier* digunakan pada penelitian ini dengan menggunakan data sosial ekonomi kelurahan Amessangeng Kota Palopo. Dengan mengambil sampel pada masyarakat dan menggunakan variabel-variabel predictor yang dapat digunakan menghasilkan suatu kesimpulan demi menghasilkan informasi yang akurat yang dapat membantu pengambilan keputusan terhadap kebijakan sosial ekonomi di Kelurahan Amessangeng

Kata Kunci : *Klasifikasi, Naive Bayes Classifier, Kelurahan, Amessangeng*

1. Latar Belakang

Teknologi internet telah mengubah dunia, dengan banyaknya data yang tersedia sehingga lebih memudahkan untuk mengaksesnya. Data digunakan sebagai acuan untuk melihat kondisi yang pernah terjadi bahkan dapat memprediksi kejadian yang akan terjadi melalui histori data tersebut. Dengan bantuan data mining kejadian-kejadian dapat diprediksi jauh sebelum waktu kejadian. Selain itu juga dapat membantu pengambilan keputusan untuk mengantisipasi sebelum terjadinya suatu kejadian[1]

Data mining meliputi beberapa metode untuk membantu pengambilan keputusan salah satunya adalah metode klasifikasi. Klasifikasi adalah Proses penemuan model (atau fungsi) yang menggambarkan dan membedakan kelas

data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui [2].

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) [2]

Naïve Bayes adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. *Naive Bayes* didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi serupa

dengan decision tree dan neural network. *Naive Bayes* terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar [3].

Prediksi *Bayes* didasarkan pada formula teorema *Bayes* dengan formula umum sebagai berikut:

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

Dimana

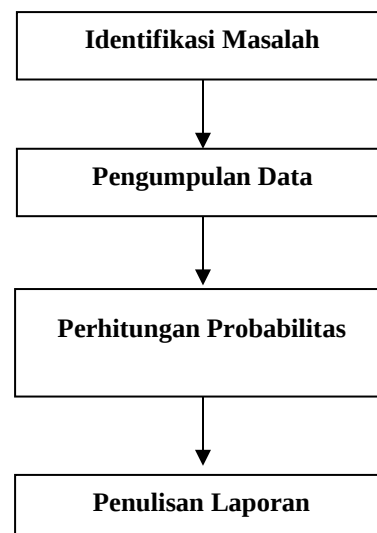
- X : Data dengan *class* yang belum diketahui
 H : Hipotesis data merupakan suatu *class* spesifik
 $P(H|X)$: Probabilitas hipotesis H berdasar kondisi X (posteriori probabilitas)
 $P(H)$: Probabilitas hipotesis H (prior probabilitas)
 $P(X|H)$: Probabilitas X berdasarkan kondisi pada hipotesis H
 $P(X)$: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

Dalam metode *naive bayes* diperlukan data latih dan data uji yang ingin diklasifikasikan, dalam *naïve bayes*, semakin banyak data latih yang yang dilibatkan, semakin baik hasil yang prediksi yang diberikan [4].

Pemanfaatan metode klasifikasi *Naive Bayes Classifier* digunakan pada penelitian ini dengan menggunakan data sosial ekonomi kelurahan Amessangeng Kota Palopo. Dengan mengambil sampel pada masyarakat dan menggunakan variabel-variabel predictor yang dapat digunakan menghasilkan suatu kesimpulan demi menghasilkan informasi yang akurat yang dapat digunakan oleh pihak yang membutuhkan terkhusus lembaga-lembaga pemerintahan seperti BAPPEDA.[1]

2. Metodologi

Pada Tahapan Penelitian yang digunakan dalam penelitian ini dapat dibagi ke dalam 3 (tiga) tahap yaitu: (1) Identifikasi Masalah, (2) Pengumpulan data, (3) Perhitungan Probabilitas, (4) Penulisan laporan.



Gambar 1. Tahapan Penelitian

Tahapan penelitian pada Gambar 1, dapat dijelaskan sebagai berikut:

1. Tahap Identifikasi Masalah : Pada tahapan ini dilakukan analisis terhadap permasalahan yang ada.
2. Tahap Pengumpulan Data : metode pengumpulan data yang dilakukan pada penelitian ini berdasarkan wawancara dengan masyarakat kelurahan Amessangeng Kota Palopo. Data yang digunakan dalam penelitian ini adalah Status rumah tinggal, Aset, Usaha/pekerjaan dan pendapatan tiap bulan;
3. Tahap perhitungan probabilitas : pada tahap ini akan dilakukan pengolahan data untuk mengklasifikasikan data sosial ekonomi keluarga kelurahan Amessangeng melalui perhitungan Probabilitas .
4. Penulisan Laporan : pada tahap ini yang dilakukan adalah menyusun

laporan dari hasil pengujian yang telah dilakukan

3. Hasil dan Pembahasan

Pada tahapan kali ini akan dijelaskan secara detail mengenai tahapan pada metode Naïve Bayes Classifier. Tahapan tersebut:

1. Baca Data Training

Pada penentuan data yang akan diolah dan dianalisis dengan menggunakan Naïve Bayes data tersebut harus sesuai dengan variabel yang dibutuhkan seperti Status rumah tinggal, Aset, Usaha, dan pendapatan perbulan. Variabel yang digunakan untuk klasifikasi social ekonomi seperti:

- a. Nama kepala keluarga
Variable ini terdiri dari dua pengelompokan yakni layak atau tidak layak.
- b. Status Rumah Tinggal
Pada variable ini digunakan tiga kategori antara lain Milik sendiri, Sewa, dan Numpang.
- c. Aset
Variable asset dibagi menjadi tiga pengelompokan asset tinggi, Sedang, dan Rendah.
- d. Usaha/Pekerjaan.
Usaha yang dijadikan variable dalam pengklasifikasian antara lain Wiraswasta, Buruh, Pedagang, Karyawan, Petani dan Pensiunan.
- e. Penghasilan per bulan
Pada variable penghasilan per bulan dibagi menjadi tiga kategori antara $< 1.500.000$, $1.500.000 \leq x < 3.500.000$, dan $\geq 3.500.000$

2. Perhitungan Probabilitas Prior (P(Ci))

Pada penelitian ini termuat 98 data latih yang digunakan. Terdapat 31 data pada kelas C_0 kehidupan tidak layak (menengah ke bawah) dan 67 data pada kelas C_1 (kehidupan layak). Probabilitas

prior layak dan tidak layak menerima bantuan menggunakan persamaan

$$P(C_0) = \frac{31}{98} = 0.31$$

Untuk probabilitas kehidupan layak (menengah ke atas):

$$P(C_1) = \frac{68}{98} = 0.69$$

3. Perhitungan Probabilitas Posterior X bersyarat C (P(X|Ci))

Penentuan probabilitas posterior yang menggunakan 98 data dengan X sebagai vector pemilihan kriteria keluarga dengan status “kehidupan social ekonomi layak” dan status “kehidupan social ekonomi tidak layak” yaitu $X_{\text{status rumah tinggal}}$, X_{Aset} , X_{Usaha} , dan $X_{\text{pendapatan}}$ yang dapat dijabarkan oleh $P(X|C_i)$ yang dinotasikan menjadi $P(X|C_i)$ untuk setiap X dihitung kemungkinannya pada setiap C_i .

Untuk menghitung setiap peluang hasil kriteria $P(X|C_i)$ dapat dilihat pada table 2.

Tabel 2. Probabilitas Status Rumah Tinggal

Status Rumah Tinggal	Jumlah kejadian		Probabilitas	
	Layak	Tidak Layak	Layak (C_1)	Tidak Layak (C_0)
Milik Sendiri	31	23	0,4697	0,7188
Sewa	20	4	0,3030	0,1250
Numpang	15	5	0,2273	0,1563
Jumlah	66	32	1	1

Untuk menghitung setiap peluang hasil kriteria $P(X|C_i)$ dapat dilihat pada table 3.

Tabel 3. Probabilitas Aset

Aset	Jumlah kejadian		Probabilitas	
	Layak	Tidak Layak	Layak (C_1)	Tidak Layak (C_0)
Tinggi	20	1	0,2941	0,0333
Sedang	45	19	0,6618	0,6333
Rendah	3	7	0,0441	0,2333
Tidak Ada	0	3	0,0000	0,1000
Jumlah	68	30	1	1

Peluang hasil kriteria $P(XX \text{ Pekerjaan} | C_i)$ dapat dihitung seperti pada table 4.

Tabel 4. Probabilitas Pekerjaan

Pekerjaan	Jumlah kejadian		Probabilitas	
	Layak	Tidak Layak	Layak (C_1)	Tidak Layak (C_0)
Wiraswasta	5	6	0,0725	0,2069
Buruh	6	5	0,0870	0,1724
Pedagang	7	4	0,1014	0,1379
Karyawan	29	12	0,4203	0,4138
Petani	19	2	0,2754	0,0690
Pensiunan	3	0	0,0435	0,0000
Jumlah	68	30	1	1

Untuk menghitung setiap peluang hasil kriteria $P(XX \text{ Pendapatan} | C_i)$ dapat dilihat pada table 5.

Tabel 5. Probabilitas Pendapatan

Pendapatan	Jumlah kejadian		Probabilitas	
	Layak	Tidak Layak	Layak (C_1)	Tidak Layak (C_0)
<1.500.000	8	19	0,2000	0,3276
<=1.500.000 dan <3.500.000	20	39	0,5000	0,6724
>3.500.000	12	0	0,3000	0,0000
Jumlah	40	58	1	1

4. Perhitungan Manual

Pada perhitungan manual yang menggunakan data uji dapat dilihat pada table 6 yang menggunakan metode Naïve Bayes Classifier.

Tabel 6. Data Uji

No	Nama KK	Status Rumah Tinggal	Aset	Pekerjaan	Pendapatan	Status
62	Lewa	Sewa	sedang	karyawan	5.000.000	?

a. Pendefinisian Variabel

Dari tabel 6 dapat didefinisikan data uji X berdasarkan aturan metode Naive Bayes sebagai berikut:

$X = \text{status Rumah Tinggal} = \text{Sewa}$,
 $X_{\text{aset}} = \text{Sedang}$, $X_{\text{pekerjaan}} = \text{Karyawan}$,
 $X_{\text{pendapatan}} = 5.000.000\}$

b. Pendefinisian Probabilitas Prior $P(C_i)$

Pada perhitungan hasil Probabilitas Priornya berdasarkan persamaan perhitungan peluang prior untuk menentukan kehidupan layak dan kehidupan tidak layak diperoleh kehidupan layak $P(C_1) = 0.69$ dan kehidupan tidak layak $P(C_0) = 0.31$.

c. Perhitungan Probabilitas Data Uji

Pada data uji pada tabel 6 dapat dilakukan klasifikasi ke dalam kelas kehidupan tidak layak $P(C_0)$ dengan nilai peluang pada masing-masing variable sebagai berikut: 0,0125; 0,6333; 0,4138; 0. Sehingga nilai $P(X|C_0) = P(X_{\text{status rumah tinggal}}|C_0) \times P(X_{\text{aset}}|C_0) \times P(X_{\text{pekerjaan}}|C_0) \times P(X_{\text{pendapatan}}|C_0) = 0.0125 \times 0.6333 \times 0.4138 \times 0 = 0$

Untuk perhitungan $P(X|C_i)$ ke dalam kelas Kehidupan layak maka nilai peluang dari masing-masing variabel adalah 0,3030; 0,6618; 0,4203; 0,3000

Sehingga nilai $P(X|C_1) = P(X_{\text{status rumah tinggal}}|C_1) \times P(X_{\text{aset}}|C_1) \times P(X_{\text{pekerjaan}}|C_1) \times P(X_{\text{pendapatan}}|C_1) = 0,3030 \times 0,6618 \times 0,4203 \times 0,3000 = 0,0253$

d. Maksimisasi $P(X|C_i) P(C_i)$

Untuk memaksimalkan nilai klasifikasi pada tiap kelas maka yang harus dilakukan dengan mengalikan nilai probabiliti data uji dan probabiliti priornya.

Pada kelas kehidupan tidak layak

$$\begin{aligned} P(C_0|X) &= P(X|C_0) \times P(C_0) \\ &= 0 \times 0.31 \\ &= 0 \end{aligned}$$

Untuk menghitung kelas kehidupan layak

$$\begin{aligned} P(C_1|X) &= P(X|C_1) \times P(C_1) \\ &= 0,0253 \times 0.69 \\ &= 0,0174 \end{aligned}$$

Dari perhitungan di atas dapat dihasilkan bahwa nilai $P(C_0|X) = 0$ sedangkan nilai $P(C_1|X) = 0,0174$. Berdasarkan nilai tersebut dapat disimpulkan bahwa $P(C_1|X) > P(C_0|X)$ maka data uji tersebut diklasifikasikan dalam kelas kehidupan layak.

4. Kesimpulan

Dari penelitian yang dilakukan dapat ditarik kesimpulan sebagai berikut :

1. Penerapan data mining dalam menentukan klasifikasi status sosial ekonomi masyarakat dapat digunakan untuk memprediksi kehidupan layak atau tidak layak sebuah keluarga dalam status sosial ekonominya dengan menggunakan algoritma *Naive Bayes*
2. Dari data yang diperoleh disimpulkan bahwa secara umum masyarakat kelurahan Amessangeng Kota Palopo memiliki status sosial ekonomi kehidupan yang layak, dengan persentasi 53.06 % layak dan 46.94 % tidak layak

Daftar Pustaka

- [1] Irmayani., Penerapan Algoritma Cart Klasifikasi Sosial Ekonomi Masyarakat Kelurahan Amessangeng., Jurnal Ilmiah Information Technology d'Computare, 10 (2), 17-22,2020
- [2] Annur, Haditsah., Klasifikasi Masyarakat Miskin Menggunakan

Metode Naïve Bayes., ILKOM Jurnal Ilmiah Volume, 10(2),160-165, 2018

- [3] Nugroho, Y.S., Data Mining Menggunakan Algoritma *Naive Bayes* Untuk Klasifikasi Kelulusan Mahasiswa Universitas Dian Nuswantoro., Academia.
- [4] Fadlan, Chaerul, dkk Penerapan Metode Naïve Bayes Dalam Klasifikasi Kelayakan Keluarga Penerima Beras Rastra.,JUTIM, 3(1),1-8, 2018