

IMPLEMENTASI METODE *NAIVE BAYES* UNTUK KLASIFIKASI PENERIMA PROGRAM KELUARGA HARAPAN

Andhini Asri Awaliyah Arifin¹, Wiwin Handoko^{2*}, Zulfan Efendi

¹Mahasiswa Prodi Sistem Informasi, STMIK Royal

²Prodi Sistem Informasi, STMIK Royal

**email*: win.van.handoko@gmail.com

Abstract: The Family Hope Program (PKH) is a government program in the form of cash for Very Poor Households (RTSM) whose qualifications are met related to efforts to improve human quality, not in the field of education but also health. In the short term, the PKH program is calculated to reduce the expenditure of poor families and reduce poverty in the long term. To receive the PKH Program) the government has set several criteria, including income, house ownership status, house size, floor type, roof, wall, and type of water source. As for the way to do the settlement of the criteria that have been set, namely by utilizing the Data Mining technique through the Naïve Bayes method. The dataset in this study is the data of recipients of the 2020 Family Hope Program as many as 82 samples. The results of the classification modeling with the Naïve Bayes Algorithm produce precision values for the positive class 100%, for the negative class 77%, the recall value for the positive class 80%, for the negative class 100%, the f1-score value for the positive class 89%, for the negative class 87%, and 88% accuracy value. The purpose of this research is to help the Social Service in classifying the recipients of the Family Hope Program (PKH).

Keywords: The Family Hope Program; Data Mining; Naïve Bayes

Abstrak: Program Keluarga Harapan (PKH) merupakan program pemerintah dalam bentuk tunai untuk Rumah Tangga Sangat Miskin (RTSM) yang kualifikasinya terpenuhi terkait dengan upaya peningkatan kualitas manusia tersebut bukan dalam bidang pendidikan tetapi juga kesehatan. dalam jangka pendek Program PKH diperhitungkan bisa mengurangi biaya pengeluaran keluarga miskin serta mengurangi kemiskinan dalam jangka panjang. Untuk menerima Program PKH) pemerintah sudah menetapkan beberapa kriteria, diantaranya Penghasilan, Status Kepemilikan Rumah, Ukuran Rumah, tipe Lantai, Atap, Dinding, serta jenis sumber air. Adapun cara untuk melakukan penyelesaian terhadap kriteria yang sudah ditetapkan yaitu dengan memanfaatkan teknik *Data Mining* melalui Metode *Naïve Bayes*. Dataset dalam penelitian ini adalah data penerima Program Keluarga Harapan tahun 2020 sebanyak 82 sampel. Hasil pemodelan klasifikasi dengan Algoritma *Naïve Bayes* menghasilkan besaran nilai *precision* untuk kelas positif 100%, untuk kelas negatif 77%, nilai *recall* untuk kelas positif 80%, untuk kelas negatif 100%, nilai *f1-score* untuk kelas positif 89%, untuk kelas negatif 87%, dan nilai akurasi 88%. Tujuan dilaksanakannya penelitian ini adalah untuk membantu Dinas Sosial mengklasifikasikan penerima Program Keluarga Harapan (PKH).

Kata kunci: Program Keluarga Harapan; *Data Mining*; *Naïve Bayes*

PENDAHULUAN

PKH merupakan program dari pemerintah melalui bantuan kepada Rumah Tangga Sangat Miskin (RTSM) dalam bentuk tunai. Dalam hal ini RTSM diharuskan untuk memenuhi kualifikasi yang ditentukan dalam upaya peningkatan kualitas SDM khususnya bidang kesehatan serta kesehatan. Sebenarnya, secara mendasar PKH bertujuan untuk menaikkan jangkauan terhadap pelayanan kesejahteraan sosial, kesehatan & pendidikan agar kualitas keluarga yang baik dapat tercapai. Dalam jangka pendek Program PKH diperhitungkan bisa mengurangi biaya pengeluaran keluarga miskin serta mengurangi kemiskinan dalam jangka panjang [1].

Program PKH di Desa Sarang Helang Kecamatan Sei Kepayang Timur belum dapat dirasakan warga yang berhak, sehingga kecemburuan sosial dalam lingkungan masyarakat. belum adanya sistem yang baik adalah salah satu penyebabnya sehingga tidak dapat mempermudah dalam mengklasifikasi penerima Program PKH. Proses olah data masih dengan cara sederhana. Hal tersebut tentunya memakan waktu yang signifikan dalam memproses data tersebut, resiko lainnya adalah terjadinya dokumen sering hilang atau rusak. Sehingga, penyaluran Program PKH dinilai kurang tepat.

Berdasarkan masalah diatas, Desa Sarang Helang Kecamatan Sei Kepayang Timur membutuhkan suatu sistem yang dapat menentukan kelayakan penerima Program Keluarga Harapan (PKH) secara efisien. Masalah penentuan kelayakan tersebut dapat diatasi dengan penerapan data mining dengan menggunakan metode Algoritma *Naïve Bayes*. Berikut rangkuman riset terdahulu yang sudah dicoba antara lain: riset yang dicoba oleh [2] dengan judul “Implementasi Tata cara *Naïve Bayes* dalam dengan klasifikasi Penerima Program Keluarga Harapan (PKH) Desa Minggiran Kediri”. Riset tersebut dicoba untuk mengklasifikasikan calon penerima dorongan PKH berbasis web dengan tingkatan akurasi sebesar 93,33%

Riset berikutnya yang dicoba oleh [3] yang bertajuk “Penerapan Metode *Naïve Bayes* dan *Simple Additive Weighting* Dalam Penentuan Penerima Program Keluarga Harapan”. Riset tersebut memakai 2 model, ialah *Naïve Bayes* serta SAW untuk memastikan penerima PKH dengan kriteria luas bangunan, tipe lantai, tipe bilik, energi listrik, jumlah makan daging per pekan, kemampuan berobat, jumlah makan per hari, pemasukan, kepemilikan kendaraan, tipe pekerjaan, serta mempunyai anak sekolah hingga jenjang SMP serta memiliki tingkat akurasi sebesar 93,8% dengan model *Naïve Bayes* serta tingkat akurasi mencapai 100% dengan model SAW. Pada penelitian yang dilakukan oleh [4] dengan judul “Klasifikasi Penerima Bantuan Sosial Program Keluarga Harapan Menggunakan Metode *Naïve Bayes*” menyimpulkan bahwasannya *data mining* model *Naïve Bayes* akan mempermudah pemerintah dalam penentuan penerima bantuan PKH dengan tingkat akurasi menggunakan *Confusion Matrix* sebesar 84,2411% yang termasuk ke dalam kategori *Good Classification*.

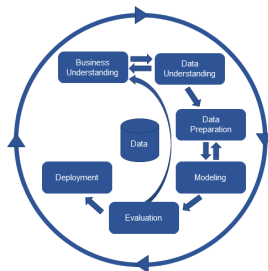
Selain itu, dalam penelitian yang dilakukan oleh [5] dengan judul “Penerapan Algoritma *Naïve Bayes* untuk Pemilihan Keluarga Yang Membutuhkan Bantuan Dalam Program Keluarga Harapan (PKH)” menggunakan 6 dasar utama kriteria, yakni: Status PKH, , Kepala Rumah Tangga, Jumlah Tanggungan, Jumlah Penghasilan, Kondisi Rumah, dan Status Pemilik Rumah. Hasil akhir dari penerapan metode *Naïve Bayes* adalah dari 70 sampel yang digunakan terdapat 3,5% sampel tidak tepat untuk menerima bantuan PKH.

Penelitian yang dilakukan di Desa Sarang Helang Kecamatan Sei Kepayang Timur bertujuan untuk menerapkan Algoritma *Naive Bayes* untuk mengklasifikasi warga tidak mampu berdasarkan pengalaman di masa sebelumnya yang berhak memperoleh bantuan pemerintah dalam bentuk PKH. output dari penelitian ini diharapkan menjadi solusi alternatif khususnya dinas sosial dalam menanggulangi permasalahan penerima bantuan PKH di Desa Sarang Helang Kecamatan Sei Kepayang Timur.

METODE

Data mining diartikan sebagai proses menemukan informasi dan *pattern* yang lebih berguna dari data yang cukup besar. *Data mining* mencakup pengumpulan, ekstraksi, analisis, dan statistik data [6]. *Data mining* yaitu suatu proses bantuan matematika, metode statistik, kecerdasan artifisial, serta *machine learning* untuk mengekstraksi serta inialisasi data berguna menciptakan pengetahuan yang terpaut berbagai sumber data besar [7]. *Output* dalam *data mining* dipergunakan sebagai alternatif solusi menentukan keputusan untuk memperbaiki keputusan dimasa akan datang.

Alur penelitian ini mengacu pada *Cross-Industry Standard Process (CRISP-DM)* untuk *Data Mining*. CRISP-DM merupakan standart proses data mining untuk strategi pemecahan permasalahan secara universal dari bisnis ataupun unit riset [8].



Gambar 1. Tahapan CRISP-DM

Berdasarkan gambar 1 dapat dilihat terdapat 6 dari CRISP-DM yaitu:

Bussines Understanding

PKH adalah program pemerintah dalam bentuk bantuan sosial bersyarat yang diberikan kepada warga kategori miskin. salah satunya adalah warga miskin yang berada di Desa Sarang Helang Kecamatan Sei Kepayang Timur. Fokus penelitian ini adalah analisis bagaimana membuat klasifikasi untuk menentukan warga kategori miskin menerima bantuan PKH atau tidak sehingga bantuan disalurkan tepat sasaran.

Data Understanding

Sumber data yang didapat berasal dari Dinas Sosial Kabupaten Asahan, khususnya Desa Sarang Helang Kecamatan Sei Kepayang Timur. Data tersebut berisikan atribut antara lain: Nama, Penghasilan, Kepemilikan Rumah, Ukuran Rumah, tipe Lantai, tipe Atap, tipe Dinding, jenis sumber air, dan Status PKH.

```
df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 82 entries, 0 to 81
Data columns (total 10 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   No                     82 non-null     int64
1   Nama                   82 non-null     object
2   Penghasilan            82 non-null     int64
3   Status Kepemilikan Rumah 82 non-null     object
4   Ukuran Rumah          82 non-null     object
5   Jenis Lantai           82 non-null     object
6   Jenis Dinding          82 non-null     object
7   Jenis Atap             82 non-null     object
8   Sumber Air Minum       82 non-null     object
9   Keputusan              82 non-null     object
dtypes: int64(2), object(8)
memory usage: 6.5+ KB
```

Gambar 2. Deskripsi Data

Data Preparation

Beberapa hal yang dilakukan dalam persiapan data diantaranya yaitu memilih data, membangun data, mengintegrasikan data dan pembersihan data.

```
df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 82 entries, 0 to 81
Data columns (total 8 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   Penghasilan            82 non-null     int64
1   Status Kepemilikan Rumah 82 non-null     object
2   Ukuran Rumah          82 non-null     object
3   Jenis Lantai           82 non-null     object
4   Jenis Dinding          82 non-null     object
5   Jenis Atap             82 non-null     object
6   Sumber Air Minum       82 non-null     object
7   Keputusan              82 non-null     object
dtypes: int64(1), object(7)
memory usage: 5.2+ KB
```

Gambar 3. Deskripsi Data yang dipakai

Modelling

Metode naive bayes ialah pengklasifikasi peluang sederhana melalui perhitungan sekumpulan probabilitas menambahkan kombinasi nilai dari kumpulan data tertentu frekuensinya. Model ini didasari pada asumsi yang disederhanakan nilai atribut independen bersyarat diberikan nilai luaran. sehingga nilai luaran (*output*), probabilitas mengamati secara bersama adalah produk dari probabilitas individu [9].

Persamaan dari *teorema bayes* adalah:

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

Keterangan:

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 82 data yang dimiliki, dibagi menjadi 2 bagian yaitu *data training* dan *data testing* dengan porsi 70:30.

```
x_train.shape
(57, 7)

x_test.shape
(25, 7)
```

Gambar 4. Split Data

Evaluation

Model ini alat yang umum digunakan untuk analisis kualitas pengklasifikasi

yang dapat mengenal kelas tupel yang berbeda. Evaluasi dengan *Confusion Matrix* biasanya menampilkan akurasi, presisi, dan recall. *Confusion matrix* adalah tabel matriks yang terbagi menjadi dua kelas yaitu kelas negatif dan kelas positif [10].

Deployment

Tampilan hasil data mining divisualisasikan dalam bahasa pemrograman agar masyarakat umum dapat memahami dan membacanya.

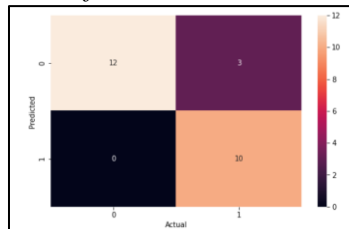
HASIL DAN PEMBAHASAN

Tahap awal untuk analisis yaitu meng-*import* modul yang diperlukan serta dataset Penerima Program Keluarga Harapan (PKH).

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

Gambar 5. Dataset Penerima PKH

Setelah membangun model dengan data training, gunakan data testing untuk melakukan pengujian. perolehan *confusion matrix* untuk metode *Naïve Bayes* adalah:



Gambar 6. Visualisasi Confusion Matrix

	precision	recall	f1-score	support
0	1.00	0.80	0.89	15
1	0.77	1.00	0.87	10
accuracy			0.88	25
macro avg	0.88	0.90	0.88	25
weighted avg	0.91	0.88	0.88	25

Gambar 7. Tingkat Akurasi

Berdasarkan Gambar 8 maka didapat hasil Algoritma *Naïve Bayes* menghasilkan nilai *precision* untuk kelas positif 100%, untuk kelas negatif 77%, nilai *recall* untuk kelas positif 80%, untuk kelas negatif 100%, nilai *f1-score* untuk kelas positif 89%, untuk kelas negatif sebesar 87%, dan nilai akurasi 88% untuk metode *Naïve Bayes*.

SIMPULAN

Berdasarkan pembahasan menggunakan Algoritma *Naïve Bayes* dan Model *Confusion Matrix*, didapatkan nilai *precision* 91%, nilai *recall* 88%, nilai *f1-score* 87%,

dan nilai akurasi sebesar 88% yang masuk dalam kategori *Good Classification*. Oleh karena itu, metode *Naïve Bayes* adalah model yang dikategorikan baik dan diimplementasikan memprediksi peluang berdasarkan pengalaman histori sebelumnya untuk mempermudah dalam proses klasifikasi calon penerima program PKH.

DAFTAR PUSTAKA

- [1] E. Fitriani, “Perbandingan Algoritma C4.5 Dan Naïve Bayes Untuk Menentukan Kelayakan Penerima Bantuan Program Keluarga Harapan,” *Sistemasi*, vol. 9, no. 1, p. 103, 2020, doi: 10.32520/stmsi.v9i1.596.
- [2] F. K. Pratama, D. W. Widodo, and N. Shofia, “Implementasi Metode Naïve Bayes dalam Mengklasifikasi Penerima Program Keluarga Harapan (PKH) Desa Minggiran Kediri,” 2021.
- [3] Y. Susanti, Y. D. Rosita, and D. Hanum, “Penerapan Metode Naive Bayes Dan Simple Additive Weighting Dalam Penentuan Penerima Program Keluarga Harapan,” *Semin. Nas. Sains dan Teknol. Terap.*, pp. 205–210, 2018.
- [4] N. Alfiah, “Klasifikasi Penerima Bantuan Sosial Program Keluarga Harapan Menggunakan Metode Naive Bayes,” *J. Teknol. Inf.*, vol. XVI, pp. 32–40, 2021.
- [5] Wiyanto and S. Atmaja, “Penerapan Algoritma Naive Bayes Untuk Pemilihan Keluarga Yang Membutuhkan Bantuan Dalam Program Keluarga Harapan (PKH) (Studi Kasus di Desa Karang Asih, Cikarang Utara),” *J. Teknol. Pelita Bangsa*, vol. 46, no. 4, pp. 31–44, 2018.
- [6] N. E. Putria, “Data Mining Penjualan Tiket Pesawat Menggunakan Algoritma Apriori Pada Terminal Tiket Batam Tour & Travel,” *Comput. Based Inf. Syst. J.*, vol. 6, no. 1, pp. 29–39, 2018, [Online]. Available: <http://ejournal.upbatam.ac.id/index.php/cbis/article/download/643/410>.
- [7] K. Handoko and L. S. Lesmana, “Data Mining Pada Jumlah Penumpang Menggunakan Metode Clustering,” *Snistek*, no. 1, pp. 97–102, 2018.
- [8] M. Fitriani and G. F. Nama, “Implementasi Association Rule Dengan Algoritma Apriori Pada Data Peminjaman Buku Upt Perpustakaan Universitas Lampung Menggunakan Metodologi CRISP-DM Jurnal Informatika dan Teknik Elektro Terapan (JITET) Vol 10 , No . 1 , Januari 2022,” vol. 10, no. 1, pp. 310–318, 2022.
- [9] E. Manalu, F. A. Sianturi, and M. R. Manalu, “Penerapan Algoritma Naive Bayes Untuk Memprediksi Jumlah Produksi Barang Berdasarkan Data Persediaan dan Jumlah Pemesanan Pada CV. Papadan Mama Pastries,” *J. Mantik Penusa*, vol. 1, no. 2, pp. 16–21, 2017.
- [10] L. A. Utami, “Melalui Komparasi Algoritma Support Vector Machine Dan K-Nearest Neighbor Berbasis Particle Swarm Optimization,” vol. 13, no. 1, pp. 103–112, 2017.