

PENERAPAN DATA MINING UNTUK MENGLASIFIKASI PENERIMA BANTUAN PKH DESA WAE JARE MENGGUNAKAN METODE NAÏVE BAYES

Apolinaria Ifon Purnama¹⁾, Abdul Aziz²⁾, Anggri Sartika Wiguna³⁾

Teknik Informatika Universitas Kanjuruhan Malang, Jl. S. Supriadi, Malang, Indonesia
email: ifonpurnama@gmail.com¹⁾, abdul.aziz@unikama.ac.id²⁾, 4n66121@gmail.com³⁾

Abstrak

Program Keluarga Harapan (PKH) merupakan sebuah bantuan dari pemerintah untuk menanggulangi kemiskinan. Wae Jare merupakan sebuah desa yang juga mendapatkan bantuan PKH tersebut. Dalam melakukan pendataan untuk penerimaan bantuan PKH tersebut masih mengalami kesulitan dalam mengklasifikasi antara penerima dan bukan penerima bantuan PKH. Penelitian ini bertujuan untuk mengatasi permasalahan kesulitan dalam mengklasifikasi penerima dan bukan penerima bantuan PKH. Metode yang digunakan adalah Naïve Bayes dengan menggunakan fungsi Klasifikasi. Jumlah data yang digunakan adalah sebanyak 210 data dengan 9 kriteria. Penerapan data mining menggunakan metode Naïve Bayes dalam penelitian ini memperoleh tingkat akurasi sebanyak 82,14%. Untuk menghitung akurasi dalam penelitian ini menggunakan teknik Confusion Matrix. Tingkat akurasi yang diperoleh termasuk dalam kategori Sedang. Perhitungan akurasi pada Confusion Matrix dilakukan secara manual serta menggunakan software RapidMiner untuk membandingkan hasil yang didapatkan.

Kata Kunci :

Data Mining, Klasifikasi, Naïve Bayes, PKH

Abstract

PKH is an assistance program from government to overcome poverty in Indonesia. Wae Jare is one of the villages which had got the assistance from PKH program. In categorizing the data receiver of PKH program, there are some difficulties in classifying between the receiver and non-receiver of PHK's assistance. This study is aimed to solve the problem of difficulties in classifying the receiver and non-receiver of PHK's assistance program. The method that used in this study is Naïve Bayes by using the function of classification. The amount of data in this study is about 210 data with 9 different criteria. The application of data mining by using Naïve Bayes method in this study got 82,14% level of accuracy. To calculate the accuracy in this research, the researcher used Confusion Matrix Technique. The accuracy level of this study is included in Medium Level. The calculation of the accuracy is done manually by using the software of RapidMiner in order to compare the result that obtained in this study.

Keywords :

Data Mining, Classification, Naïve Bayes, PKH

1. PENDAHULUAN

Bantuan dari pemerintah sangatlah diharapkan oleh banyak masyarakat terutama masyarakat yang kurang mampu atau berekonomi lemah (Sari & Prabowo, 2017). Di Desa Wae Jare Kecamatan Mbeliling Kabupaten Manggarai Barat yang juga merupakan salah satu desa yang masyarakatnya kurang mampu. Untuk mengatasi permasalahan tersebut pemerintah berusaha membantu masyarakat yang kurang mampu dengan berbagai macam jenis bantuan salah satunya adalah Program Keluarga Harapan (PKH). PKH merupakan suatu program penanggulangan kemiskinan (Magfiroh, 2016). Dengan adanya bantuan PKH tersebut sangatlah membantu masyarakat yang kurang mampu dan juga tentu berkurangnya angka kemiskinan di desa tersebut.

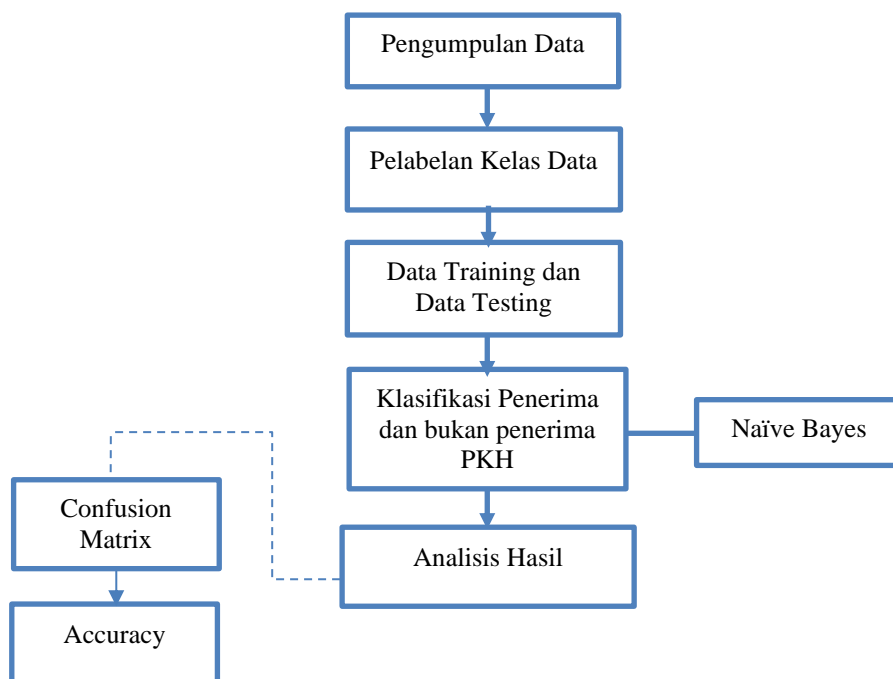
Pada penelitian ini menggunakan data mining dengan teknik klasifikasi. Beberapa peneliti terdahulu dijadikan sebagai bahan referensi penelitian diantaranya, penelitian yang dilakukan oleh Magfiroh (2016) dalam penelitiannya yang berjudul “Rekomendasi Penentuan Penerima Bantuan Iuran (PBI) Menggunakan Algoritma Naïve Bayes” dalam penelitiannya menggunakan algoritma Naïve Bayes dan memperoleh hasil akurasi sebesar 81,67% sedangkan Sari dan Prabowo (2017) “Penentuan Kelayakan Penerima Bantuan Renovasi Rumah Warga Miskin Menggunakan Naïve Bayes” memperoleh tingkat akurasi sebesar 90%.

Sejak tahun 2009 desa Wae Jare telah mendapat bantuan PKH dari pemerintah. Setelah mendapatkan bantuan PKH Masyarakat sangat terbantu dan ada perubahan peningkatan dalam perekonomiannya. Namun pada saat melakukan pendataan untuk penerimaan bantuan PKH tersebut masih mengalami kesulitan dalam mengklasifikasi antara penerima dan bukan penerima bantuan PKH (Karyadiputra dkk, 2016). Untuk mengatasi permasalahan kesulitan dalam mengklasifikasi penerima dan bukan penerima bantuan maka dalam penelitian ini menerapkan sebuah metode pengklasifikasian untuk mengatasi permasalahan tersebut. Metode yang digunakan adalah Naïve Bayes dengan menggunakan fungsi Klasifikasi (Annur 2018).

2. METODE PENELITIAN

Penelitian kuantitatif menggunakan metode Naïve Bayes untuk mengklasifikasi penerima dan bukan penerima bantuan PKH di desa Wae Jare. Klasifikasi merupakan sebuah fungsi untuk menentukan sebuah data berdasarkan kategorinya (Lestari, Akmaludin, dan Badrul, 2020). Naïve Bayes digunakan untuk menjalankan fungsi klasifikasi pada penerima dan bukan penerima bantuan PKH di desa Wae Jare tahun 2019 (Herlina, 2016). Rumus metode Naïve Bayes yaitu :

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



Gambar 1. Tahapan Penelitian

Pengumpulan data dengan cara meminta secara langsung kepada aparat desa yang berwenang yaitu Sekertaris desa (Karyadiputra & Hijriana, 2018). Setelah mendapatkan data kemudian pemberian label pada data untuk dilakukan pengklasifikasian. Label yang digunakan yaitu Terima dan Tidak. Data training dan data testing ditentukan menggunakan teknik purposive sampling (Fauzi & Tukiyyat, 2019). Klasifikasi menggunakan Naïve Bayes dengan melakukan perhitungan probabilitas pada data training (Rifqo & Wijaya, 2017). Setelah melakukan klasifikasi selanjutnya menganalisis data dengan menggunakan teknik confusion matrix untuk mendapatkan tingkat keakuratan hasil klasifikasi (Salmu & Solichin, 2017) dan (Sulistiani, 2018).

3. HASIL DAN PEMBAHASAN

Hasil yang didapatkan sesuai tahapan penelitian yang dilakukan adalah:

3.1 Pengumpulan Data

Jumlah data penerima dan bukan penerima bantuan PKH di desa Wae Jare tahun 2019 yang diperoleh pada penelitian ini adalah sebanyak 210 data. Kriteria yang digunakan sebanyak 9 diantaranya adalah: Jumlah Tanggungan, Kondisi Rumah, Pekerjaan, Penghasilan, Status Kepemilikan Rumah, Sumber Penerangan, Sumber Air Minum, Bahan Bakar, Ketersediaan WC (Karyadiputra, 2016).

3.2 Pelabelan Kelas Data

Pelabelan kelas data dilakukan pada data yang telah dikumpulkan kelas Tidak dan Terima. Jumlah data pada kelas Tidak adalah sebanyak 111 data sedangkan data pada kelas Terima tsebanyak 99 data.

3.3 Data Training dan Data Testing

Teknik purposive sampling mendapatkan jumlah data penerima dan bukan penerima bantuan PKH desa Wae Jare tahun 2019 sebagai data training sebanyak 154 data sedangkan data penerima dan bukan penerima bantuan PKH baru yang ditambahkan pada tahun 2019 sebagai data testing sebanyak 56 data.

3.4 Perhitungan Naïve Bayes

Pengklasifikasian dilakukan dengan menghitung probabilitas prior kateogori Tidak dan Terima serta probabilitas prior semua kriteria berdasarkan kategori masing-masing pada data training. Hasil Probabilitas pada data training selanjutnya digunakan untuk perhitungan probabilitas pada data testing.

3.5 Perhitungan Probabilitas Prior

Jumlah data training sebanyak 154 data dengan jumlah data bukan penerima (C0) sebanyak 74 KK dan jumlah data penerima (C1) sebanyak 80 KK. Hasil perhitungan probabilitas prior dilakukan dengan cara seperti dibawah ini.

$$P(C0) = \frac{74}{154} = 0.480$$

Sedangkan untuk hasil perhitungan probabilitas penerima yaitu:

$$P(C1) = \frac{80}{154} = 0.519$$

Nilai probabilitas kelas Tidak pada data training yang didapatkan dari perhitungan probabilitas prior adalah 0.480 sedangkan nilai probabilitas pada kelas Terima adalah 0.519. perbandingan antara nilai probabilitas kelas Tidak dan Terima adalah $0.480 < 0.519$ yang artinya peluang menerima pada data training lebih banyak darai pada yang Tidak menerima.

3.6 Perhitungan Probabilitas Prior Kriteria

Perhitungan probabilitas prior dilakukan pada data training sebanyak 154 data. Pada setiap masing-masing kriteria akan dilakukan perhitungan sesuai dengan kategorinya masing-masing. Hasil probabilitas prior kriteria yang diperoleh akan digunakan untuk menghitung probabilitas prosterior untuk mendapatkan hasil pengklasifikasian. Penjabaran perhitungan probabilitas prior kriteria yaitu : $(P(XJumlah Tanggungan) * P(XKondisi Rumah) * P(XPekerjaan) * P(XStatus Kepemilikan Rumah) * P(XSumber Penerangan) * P(XSumber Air Minum) * P(XBahan Bakar) * P(XKetersediaan WC|Ci)$

3.7 Perhitungan Probabilitas Posterior

Perhitungan probabilitas posterior pada data testing dilakukan dengan cara menghitung nilai probabilitas prior dan probabilitas prior kriteria. Nilai probabailitas prior kriteria pada data training akan digunakan pada perhitungan probabilitas posterior atau data testing sesuai dengan kriteria dan kategorinya masing-masing. Penjabaran dari perhitungan probabilitas posterior adalah seperti dibawah ini.

$$P(X|Tidak) = P(K1|Tidak) * P(K2|Tidak) * P(K3|Tidak) * P(K4|Tidak) * P(K5|Tidak) * P(K6|Tidak) * P(K7|Tidak) * P(K8|Tidak) * P(K9|Tidak)$$

$$P(X|Terima) = P(K1|Tidak) * P(K2|Tidak) * P(K3|Tidak) * P(K4|Tidak) * P(K5|Tidak) * P(K6|Tidak) * P(K7|Tidak) * P(K8|Tidak) * P(K9|Tidak)$$

Berdasarkan data testing pada tabel 1perhitungan probabilitas posterior dari data testing adalah :

Tabel 1 Data Testing

Nama	K	K2	K3	K4	K5	K6	K7	K8	K9
	1								
		Darura	Swast	Sedan	Milik	Listri	PDA	Kay	Ad
	7	t	a	g	Sendir	k	M	u	a

$$P(X|\text{Tidak}) = 0,4865 * 0,5676 * 0,9054 * 0,4054 * 0,8514 * 0,2432 * 0,4595 * 0,8919 * 0,7973 = 0,0069.$$

$$P(X|\text{Terima}) = 0,8500 * 0,6875 * 1 * 0,2625 * 0,9500 * 0,2375 * 0,4500 * 0,9750 * 0,8625 = 0,0131.$$

Nilai probabilitas yang digunakan pada perhitungan probabilitas data testing menggunakan nilai probabilitas yang diperoleh pada data training dengan menyesuaikan kriteria dari data dan kategorinya. Setelah mendapatkan nilai probabilitas posterior, selanjutnya yaitu perhitungan pemaksimal untuk mengklasifikasi dengan cara melakukan perkalian pada Nilai probabilitas prior dan nilai probabilitas prior kriteria

$$P(X|C_i) * P(C_i)$$

$$P(X|\text{Tidak}) = 0,0069 * 0,4805 = 0,0033$$

$$P(X|\text{Terima}) = 0,0131 * 0,5194 = 0,0068$$

Hasil probabilitas yang didapatkan dari perhitungan $P(X|C_i) * P(C_i)$ selanjutnya dilakukan perbandingan nilai. Perbandingan nilai probabilitas $P(X|\text{Tidak})$ dan $P(X|\text{Terima})$ adalah $0,0033 < 0,0068$ sehingga data tersebut diklasifikasikan kedalam class Terima.

Hasil klasifikasi pada 56 data testing yang dilakukan pada penelitian ini adalah terdapat 46 data yang diklasifikasi dengan benar sedangkan 10 data diklasifikasikan dengan salah.

3.8 Analisis Data

Hasil klasifikasi data penerima dan bukan penerima bantuan PKH tahun 2019 dengan jumlah data testing 56 data terdapat 46 data yang diklasifikasi dengan benar sedangkan 10 data diklasifikasi dengan salah. Untuk menguji tingkat keakuratan data dari hasil klasifikasi yang pada penelitian ini menggunakan teknik confusion matrix dan RappidMiner. Perhitungan akurasi dilakukan dengan menggunakan teknik confusion matrix. RappidMiner digunakan untuk membandingkan hasil akurasi yang diperoleh pada confusion matrix. Tingkat akurasi yang diperoleh adalah sebanyak 82,14%. Dengan hasil perhitungan klasifikasi adalah terdapat 46 data yang diklasifikasi dengan benar dan 10 data yang diklasifikasi salah.

Untuk mendapatkan nilai akurasi dilakukan dengan cara dimana jumlah data yang diklasifikasi dengan benar pada data testing dibagi dengan keseluruhan data testing. Rumus untuk mendapatkan nilai *accuracy* adalah sebagai berikut:

$$\text{Accuracy} : \frac{TP+TN}{TP+TN+FP+FN} * 100$$

$$\frac{34 + 12}{34 + 12 + 7 + 3} * 100$$

$$\frac{46}{56} * 100 = 82.14\%$$

Tabel 2. Confusion Matrix

Kelas	True Tidak	True Terima
Tidak	34	7
Terima	3	12

accuracy: 82.14%

	true TIDAK	true TERIMA
pred. TIDAK	34	7
pred. TERIMA	3	12

Gambar 2 Akurasi pada RappidMiner

Tabel 3. kategorisasi

Kategori	Akurasi
Rendah	<76%
Sedang	>=76% - <93%
Tinggi	>=93%

Penelitian terdahulu yang dilakukan oleh (Devita dkk, 2017), (Tempola dkk, 2017), digunakan sebagai perbandingan akurasi yang didapatkan dari metode Naïve bayes dan metode lainnya. Hasil akurasi yang diperoleh dari penelitian ini dengan penelitian sebelumnya dilakukan pembagian kategori menggunakan skala interval. Tingkat akurasi untuk kategori rendah adalah lebih kecil dari 76%, akurasi untuk kategori sedang adalah lebih besar sama dengan 76% sampai dengan lebih kecil dari 93% sedangkan untuk akurasi kategori tinggi adalah lebih besar dari 93%. Hasil akurasi dalam penelitian ini adalah 82,14% dan termasuk dalam kategori Sedang.

4 KESIMPULAN

Berikut terdapat beberapa hal yang dapat disimpulkan dari hasil penelitian ini serta beberapa saran dari peneliti untuk peneliti selanjutnya.

5.1 Kesimpulan

1. Berdasarkan hasil dari penelitian, maka dapat disimpulkan bahwa: Naïve Bayes merupakan sebuah metode yang baik digunakan mengklasifikasi.

2. Hasil pengujian akurasi yang diperoleh dalam penelitian ini menggunakan model Confusion Matrix adalah 82,14% dengan jumlah data training sebanyak 154 sedangkan data testing sebanyak 56 data.
3. Rappidminer digunakan untuk membandingkan hasil analisis secara manual.
4. Pada penelitian ini terdapat 46 data yang diklasifikasikan dengan benar sedangkan terdapat 10 data yang diklasifikasikan salah pada Naïve Bayes.
5. Tingkat akurasi yang diperoleh dalam penelitian ini termasuk dalam golongan kategori Sedang

5.2 Saran

Adapun saran untuk penelitian selanjutnya guna mengembangkan metode Naive Bayes agar lebih baik dalam mengklasifikasi.

1. Jumlah data yang akan digunakan lebih banyak agar mendapatkan hasil yang baik, karena semakin banyak data maka hasil akurasi yang diperoleh juga akan semakin baik.
2. Penulis mengharapkan pengklasifikasian seperti pada penelitian ini dapat dilakukan oleh peneliti selanjutnya dengan menggunakan metode lainnya.
3. Untuk peneliti selanjutnya, penulis mengharapkan agar penelitian seperti ini dapat menggunakan dua metode atau lebih untuk membandingkan hasil akurasi yang didapatkan.

6. REFERENSI

- [1.] Annur, H. (2018). Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes. *ILKOM Jurnal Ilmiah Volume 10 Nomor 2 Agustus 2018*, 160-165.
- Devita, N, R., Herwanto, W, H., Wibawa, P, A (2018). Perbandingan Kinerja Metode Naive Bayes Dan K-Nearest Neighbor Untuk Klasifikasi Artikel Berbahasa Indonesia. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, Vol. 5(No. 4), 427-434.
- [2.] Fauzi, A. & Tukiyat, T (2019). Analisis Potensi Dana Retail pada Nasabah PT. Bank Tabungan Negara (Persero), Tbk. Dengan Metode Decision Tree Dan Naive Bayes Berbasis Optimize Selection (Evolutionary). *Jurnal Administrasi Dan Manajemen*, Vol.9 (No.1)
- [3.] Herlina, A. (2016). Optimasi Klasifikasi Sel Tunggal PAP SMEAR Menggunakan Correlation Based Features (CFS) Berbasis C4.5 Dan Naive Bayes. *Jurnal INFORMATIKA*, Vol.3, 148- 155.
- [4.] Karyadiputra, E. (2016). Analisis Algoritma Naive Bayes Untuk Klasifikasi Status Kesejahteraan Rumah Tangga Keluarga Binaan Sosial. *Technologia” Vol 7, No.4, Oktober – Desember 2016*, 199-208.
- [5.] Karyadiputra, E & Hijriana, N (2018). Analisis Penerima Algoritma Naive Bayes Untuk Klasifikasi Prioritas Pengembangan Jalan Di Provinsi Kalimantan Selatan. *Technologia” Vol 9, No.2, April – Juni 2018*, 105-108.
- [6.] Karyadiputra, E., Noersasongko, E., Marjuni, A (2016). Klasifikasi Status Kesejahteraan Rumah Tangga Keluarga Binaan Sosial Menggunakan Naive Bayes Berbasis Seleksi Atribut CHI Squared. *Jurnal Teknologi Informasi, Volume 12 Nomor 2, Oktober 2016, ISSN 1907-3380*, 116-122.
- [7.] Lestari, S. Akmaludin, A., Badrul, M (2020). Implementasi Klasifikasi Naive Bayes Untuk Prediksi Kelayakan Pemberian Plinjaman Pada Koperasi Anugerah Bintang Cemerlang. *Jurnal PROSISKO Vol. 7 No. 1 Maret 2020*, 8-16.

- [8.] Magfiroh, Y. (2016). Rekomendasi Penentuan Penerima Bantuan Iuran (PBI). *Artikel Skripsi*, 4-9. Maricar, A. M & Pramana, D (2019). Perbandingan Akurasi Naïve Bayes dan K-Nearest Neighbor pada Klasifikasi untuk Meramalkan Status Pekerjaan Alumni ITB STIKOM Bali. *Jurnal Sistem Dan Informatika (JSI)*, 16-22.
- [9.] Rifqo, H. M & Wijaya, A (2017). Implementasi Algoritma Naive Bayes Dalam Penentuan Pemberian Kredit. *Jurnal Pseudocode, Volume IV Nomor 2, September 2017, ISSN 2355-5920*, 120-128.
- [10.] Salmu, S. & Solichin, A. (2017). Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu Menggunakan Naïve Bayes. *Prosiding Seminar Nasional Multidisiplin Ilmu*, 701-709.
- [11.] Sari, W. B & Prabowo, D (2017). Penentuan Kelayakan Penerima Bantuan Renovasi Rumah Warga Miskin Menggunakan Naive Bayes. *Jurnal Ilmiah DASI Vol. 18. No. 4 Desember 2017, hlm. 34-38*, 34-38.
- [12.] Sulistiani, H. (2018). Penerapan Algoritma Klasifikasi Sebagai Pendukung Pemberian Beasiswa Mahasiswa. *SNTI VI-2018 Universitas Trisakti*, 300-305.
- [13.] Tempola, F. Muhammad, M., Khairan, A. (2018). Perbandingan Klasifikasi Antara KNN Dan Naive Bayes Pada Penentuan Status Gunung Berapi Dengan K-Fold Cross Validaton. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, Vol. 5(No. 5), 557-584.