



Klasifikasi Rumah Tangga Miskin Menggunakan Metode Naive Bayes (Studi Kasus: Provinsi Papua Tahun 2017)

¹Aisyah Chairani P, ²Fauzi Edi Hariyanto, ³Ni Luh Evindia Andini, ⁴Zullyana Cempaka S

^{1,2,3,4}Prodi Statistika, Politeknik Statistika STIS

Jl. Otto Iskandardinata No 64C Jatinegara, Jakarta Timur, 13330

Email: 1211709516@stis.ac.id, 211709693@stis.ac.id, 3211709891@stis.ac.id,

4211710090@stis.ac.id

Abstrak

Kemiskinan merupakan salah satu permasalahan kependudukan yang banyak terjadi pada negara berkembang, salah satunya Indonesia. Badan Pusat Statistik (BPS) mencatat bahwa persentase penduduk miskin di Indonesia tahun 2017 mencapai 10,12 persen. Salah satu provinsi dengan persentase penduduk miskin tertinggi di Indonesia adalah Provinsi Papua, yakni sebesar 27,76 persen. Penelitian ini bertujuan untuk melakukan klasifikasi rumah tangga berdasarkan status ekonominya di Provinsi Papua yang diperoleh dari Survei Sosial Ekonomi Nasional (Susenas) 2017 menggunakan teknik data mining. Metode analisis yang digunakan adalah metode pengklasifikasian Naive Bayes. Variabel yang digunakan antara lain jenis kelamin, pendidikan KRT, lapangan usaha KRT, jenis atap terluas, jenis dinding terluas, jenis lantai terluas, sumber air minum, sumber penerangan, dan bahan bakar untuk memasak. Berdasarkan hasil analisis, disimpulkan bahwa pengklasifikasian status ekonomi rumah tangga di Provinsi Papua memiliki tingkat akurasi sebesar 80 persen dan termasuk dalam kategori "good" dengan nilai presisi sebesar 52 persen dan spesifisitas sebesar 91 persen.

Kata Kunci: Kemiskinan, naive bayes, klasifikasi.

Abstract

Poverty is one of the demographic problems that often occurs in developing countries, one of which is Indonesia. The Central Bureau of Statistics (BPS) noted that the percentage of poor people in Indonesia in 2017 reached 10.12 percent. One of the provinces with the highest percentage of poor people in Indonesia is Papua, which is 27.76 percent. This study will classify household data based on their economic status in Papua which was obtained from the 2017 National Socio-Economic Survey (Susenas) using the technique of data mining. The analytical method used is the Naive Bayes classification method. The variables used were gender, education of household head, business field of household head, type of widest roof, type of widest wall, type of widest floor, source of drinking water, source of lighting, and cooking fuel. Based on the analysis, it is concluded that the classification of household economic status in Papua has an accuracy rate of 80 percent and included in the "good" category with precision values of 52 percent and specificity of 91 percent.

Keywords: Poverty, naive bayes, classification

1. Pendahuluan

Penilaian pembangunan manusia menurut *United Nations Development Programme* (UNDP) adalah sebuah cara untuk melakukan pengukuran terhadap kualitas manusia di suatu wilayah. Masalah dalam proses pembangunan salah satunya adalah kemiskinan. Kemiskinan seringkali ditandai dengan tingginya tingkat pengangguran dan keterbelakangan. Masyarakat miskin umumnya lemah dalam kemampuan berusaha dan terbatasnya akses terhadap kegiatan ekonomi sehingga akan tertinggal jauh dari masyarakat lainnya yang mempunyai potensi lebih tinggi [1]. Bagi mereka yang tergolong miskin, kemiskinan bersifat nyata yang terjadi dalam kehidupan mereka sehari-hari karena mereka merasakan hidup dalam kemiskinan. Kemiskinan memiliki arti yang lebih luas, bukan hanya mengenai pendapatan ataupun garis kemiskinan akan tetapi kemiskinan memiliki arti yang lebih dalam karena berkaitan dengan ketidakmampuan mencapai aspek diluar pendapatan seperti akses kebutuhan minum, pendidikan, air bersih, dan sanitasi [2]. Ketidakmampuan tersebut dapat menimbulkan berbagai permasalahan.

Permasalahan kemiskinan dihadapi oleh seluruh negara di dunia, salah satunya di Indonesia. Berkaitan dengan hal tersebut terdapat target nasional maupun global untuk mengatasi permasalahan kemiskinan. Salah satu target global adalah *Sustainable Development Goals* (SDGs). Pada SDGs tujuan pertama ditargetkan pada tahun 2030 mengakhiri kemiskinan di manapun dan dalam semua bentuk [3]. Target tersebut masih sangat jauh jika dibandingkan dengan publikasi *BPS yaitu Statistik Indonesia 2018, persentase penduduk miskin tahun 2017 di Indonesia yaitu sebesar 10,12 persen*. Berkaitan dengan hal tersebut, masalah kemiskinan di Indonesia merupakan salah satu masalah yang perlu diatasi. Hal ini tentunya menjadi suatu tantangan yang cukup berat mengingat angka persentase penduduk miskin di seluruh wilayah Indonesia masih jauh dari target global dan terjadi disparitas regional dalam menurunkan kemiskinan menjadi salah satu penyebab masih terjadinya ketimpangan pembangunan antarwilayah.

Disparitas kemiskinan antarwilayah di Indonesia tercermin dari persentase penduduk miskin yang berbeda nyata pada level provinsi. Berdasarkan publikasi BPS yaitu *Statistik Indonesia 2018, persentase penduduk miskin tahun 2017 di Kawasan Timur Indonesia (KTI) cenderung lebih tinggi dibandingkan Kawasan Barat Indonesia (KBI)*. Jika ditinjau menurut provinsi, persentase tertinggi penduduk miskin bulan Maret dan September berada di Provinsi Papua yaitu berturut-turut sebesar 27,62 dan 27,76. Persentase penduduk miskin kembali menempatkan KTI khususnya Provinsi Papua sebagai wilayah paling timpang dengan persentase penduduk miskin tertinggi. Hal ini menunjukkan ketimpangan persentase penduduk miskin pada setiap wilayah dan ketidakmerataan pembangunan sehingga pemerintah perlu menyusun upaya strategis dalam rangka meningkatkan pembangunan nasional khususnya di Provinsi Papua.

Berdasarkan kondisi tersebut maka melalui penelitian ini dilakukan pengidentifikasian kemiskinan rumah tangga di Papua pada tahun 2017 berdasarkan atribut-atribut umum yang diduga memengaruhi. Metode yang akan digunakan adalah metode *Naïve Bayes Classifier*, yang merupakan salah satu teknik pengklasifikasian dalam *data mining* dengan berbagai atribut [4]. Atribut umum yang akan digunakan dalam melakukan klasifikasi penduduk miskin tahun 2017 adalah jenis kelamin, pendidikan KRT, lapangan

usaha KRT, jenis atap terluas, jenis dinding terluas, jenis lantai terluas, sumber air minum, sumber penerangan, klasifikasi wilayah, dan bahan bakar untuk memasak. Sementara itu untuk kelas/label adalah rumah tangga miskin/ tidak miskin. Rumusan Masalah penelitian ini adalah menentukan cara merekayasa sistem untuk klasifikasi masyarakat miskin menggunakan metode *naive bayes* dan tujuan dari penelitian ini adalah klasifikasi masyarakat miskin di Provinsi Papua menggunakan metode *naive bayes*. Berdasarkan hal tersebut diatas, peneliti tertarik untuk meneliti “Klasifikasi Rumah Tangga Miskin di Provinsi Papua Tahun 2017 dengan Naive Bayes”.

2. Metode Penelitian

Penelitian ini menganalisis data menggunakan teknik data mining metode Naive Bayes yang dilakukan menggunakan *Software R*. Data dalam penelitian ini merupakan data sekunder yang diperoleh dari Badan Pusat Statistik berupa raw data Survey Sosial Ekonomi Nasional (Susenas) tahun 2017. Penelitian ini mencakup seluruh rumah tangga di Provinsi Papua yang terpilih menjadi sampel dalam Susenas. Adapun variabel yang digunakan dalam penelitian ini yaitu diberikan pada Tabel 1.

Tabel 1. Definisi operasional variabel

Nama variabel	Jenis	Keterangan
Status ekonomi	Kategorik	Label/Kelas
Jenis kelamin	Kategorik	Variabel <i>independent</i>
Pendidikan KRT	Kategorik	Variabel <i>independent</i>
Lapangan usaha KRT	Kategorik	Variabel <i>independent</i>
Jenias atap terluas	Kategorik	Variabel <i>independent</i>
Jenis lantai terluas	Kategorik	Variabel <i>independent</i>
Jenis dinding terluas	Kategorik	Variabel <i>independent</i>
Sumber air minum	Kategorik	Variabel <i>independent</i>
Sumber penerangan	Kategorik	Variabel <i>independent</i>
Bahan bakar untuk memasak	Kategorik	Variabel <i>independent</i>

Data mining merupakan gabungan dari beberapa disiplin ilmu yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik database dan visualisasi untuk penanganan permasalahan pengambilan informasi dari database yang besar [5]. Teori lain menyebutkan *data mining* adalah analisa terhadap data yang berukuran besar untuk menemukan hubungan yang jelas serta menyimpulkan sesuatu yang sebelumnya tidak diketahui dengan cara terkini dan berguna bagi pemilik data tersebut [6]. Beberapa tugas yang dapat dilakukan oleh data mining antara lain deskripsi, estimasi, prediksi, klasifikasi dan clustering [7].

Naive Bayes merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan [8]. Pengklasifikasian Naive Bayes diasumsikan bahwa ada atau tidaknya ciri tertentu dari suatu kelas tidak berkaitan dengan ciri kelas lainnya [9]. Dalam teori lain dijelaskan bahwa Naive Bayes merupakan pengklasifikasian dengan metode

probabilitas dan statistik yang memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya [10]. Salah satu keuntungan menggunakan metode Naive Bayes adalah metode ini hanya membutuhkan jumlah *training data* yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian [11]. Persamaan dari teorema Bayes [10] dapat digambarkan sebagai berikut:

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

dengan:

- X : Data dengan kelas yang belum diketahui,
- H : Hipotesis data merupakan suatu kelas 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.

Adapun langkah-langkah dalam melakukan pengolahan data menggunakan teknik *data mining* metode Naive Bayes antara lain:

1. Penentuan label kelas (variabel y) untuk pengklasifikasian yang akan dilakukan pada setiap observasi, yaitu miskin dan tidak miskin. Hal ini dilakukan karena pada raw data Susenas belum ada variabel yang secara langsung menggambarkan status ekonomi rumah tangga (miskin/tidak miskin). Penentuan label ini dilakukan dengan pendekatan:
 - Melakukan perbandingan antara pengeluaran setiap rumah tangga dengan garis kemiskinan rata-rata rumah tangga di Provinsi Papua.
 - Garis kemiskinan di Provinsi Papua adalah sebesar 464.056 dan rata-rata rumah tangganya adalah sebesar 4,3. Sehingga garis kemiskinan rumah tangga di Provinsi Papua adalah sebesar 1.994.440.
 - Rumah tangga dengan pengeluaran per kapita dibawah 1.994.440 akan masuk kedalam kategori "Miskin" sementara rumah tangga dengan pengeluaran lebih dari 1.994.440 akan masuk kedalam kategori "Tidak Miskin".

Label yang terbentuk nantinya akan dijadikan *actual class* yang diikutsertakan dalam pengolahan untuk akhirnya menghasilkan *predicted class*.

2. *Data Preprocessing*, yaitu proses pembersihan data yang meliputi imputasi nilai yang kosong (*missing value*), pengkategorian ulang pada beberapa variabel, dan pengecekan outlier serta noise pada data.
3. *Data Transformation*, yaitu data diubah menjadi bentuk yang sesuai untuk diproses dalam pengolahan.
4. Penggunaan metode Naive Bayes yang dilakukan dengan mempartisi data menjadi *set data training* dan *set data testing* menggunakan proporsi 70:30, melakukan pemodelan Bayesian menggunakan *training set* yang sebelumnya telah dibentuk, dan

juga mengevaluasi hasil pemodelan dengan melihat nilai-nilai yang tertera pada *confusion matrix* yang dihasilkan.

3. Hasil dan Pembahasan

Pengujian dilakukan menggunakan metode pengklasifikasian Naive Bayes dengan teknik *split validation*, dimana dataset akan dibagi kedalam dua bagian yaitu sebanyak 70 persen dijadikan sebagai *data training* dan sisanya 30 persen dijadikan sebagai *data testing*.

3.1 Confusion Matrix

Hasil proses klasifikasi yang telah dikonversi ke dalam *confusion matrix* ditampilkan pada tabel berikut.

Tabel 2. Hasil *confusion matrix*.

Prediction	Reference	
	Tidak Miskin	Miskin
Tidak Miskin	242	217
Miskin	420	2369

Berdasarkan tabel tersebut dapat dilihat bahwa terdapat 242 rumah tangga dengan kategori tidak miskin diprediksi dengan tepat sementara sisanya sebanyak 420 rumah tangga diklasifikasikan kedalam kategori miskin. Kemudian, hasil pada *confusion matrix* juga menunjukkan ada sebanyak 2369 rumah tangga miskin yang dapat diklasifikasikan dengan tepat dan sebanyak 217 sisanya secara tidak tepat diklasifikasikan ke dalam kategori tidak miskin.

3.2 Klasifikasi Naive Bayes

Kinerja dari penggunaan metode klasifikasi Naive Bayes dapat diukur dengan beberapa nilai statistik pada Tabel 3 berikut.

Tabel 3. Hasil klasifikasi Naive Bayes

Statistik	Nilai
Accuracy	0,8039
95% CI	(0,7898; 0,8174)
Sensitivity	0,36556
Specificity	0,91609
Pros Ped Value	0,52723
Neg Pred Value	0,84941
Prevalence	0,20382
Detection Rate	0,07451
Detection Prevalence	0,14132
Balanced Accuracy	0,64082

Hasilnya dapat diketahui dari Tabel 3 di atas bahwa metode Naive Bayes cukup baik digunakan untuk mengklasifikasikan rumah tangga dengan tingkat akurasi yang cukup tinggi yaitu sebesar 0,8039 atau 80,39% yang dikategorikan “Good” sehingga dapat dinyatakan bahwa pengklasifikasian dapat dipercaya. Kemudian, dengan tingkat kepercayaan 95% juga dapat disimpulkan bahwa tingkat akurasi dari model klasifikasi dengan metode Naive Bayes berada pada selang akurasi 0,7898 hingga 0,8174.

Secara keseluruhan, klasifikasi yang dihasilkan memiliki tingkat sensitivitas dan spesifisitas sebesar 0,36556 dan 0,91609. Sensitivitas sendiri merupakan rasio antara prediksi benar positif dengan keseluruhan data yang benar positif. Sensitivitas ini menjawab pertanyaan tentang persentase rumah tangga yang diprediksi tidak miskin dibandingkan keseluruhan rumah tangga yang sebenarnya tidak miskin. Sementara tingkat spesifisitas menggambarkan ketepatan memprediksi negatif dibandingkan keseluruhan data negatif yaitu persentase rumah tangga yang benar diprediksi miskin dibandingkan dengan keseluruhan rumah tangga yang sebenarnya miskin.

Selanjutnya, nilai presisi dari klasifikasi data yang dilakukan adalah sebesar 0,5273. Nilai ini menggambarkan rasio antara prediksi benar positif dengan keseluruhan hasil yang diprediksi positif atau bisa juga dikatakan sebagai nilai persentase rumah tangga yang benar tidak miskin dari keseluruhan rumah tangga yang diprediksi tidak miskin.

4. Kesimpulan (terlalu sederhana,

Penggunaan metode Naive Bayes dalam klasifikasi rumah tangga miskin di Provinsi Papua menghasilkan akurasi dengan nilai sebesar 80,39% dan termasuk dalam kategori “good”. Sementara nilai presisinya yang dihasilkan adalah 52% dan spesifisitas sebesar 91%. Berdasarkan hal tersebut dapat dinyatakan bahwa sistem klasifikasi dengan metode Naive Bayes yang dibentuk memiliki keakuratan yang cukup tinggi dan dapat digunakan sebagai bahan masukan bagi pengambil keputusan.

Daftar Pustaka

- [1] G. Kartasasmita, *Pemberdayaan Masyarakat: Konsep Pembangunan yang Berakar pada Masyarakat*, Jakarta: Bappenas, 1996.
- [2] N. Nurwati, “Kemiskinan: Model Pengukuran, Permasalahan dan Alternatif Kebijakan,” *Jurnal Kependudukan Padjadjaran*, vol. 10, no. 01, 2008.
- [3] Ishartono, S.T. Raharjo, “Sustainable Development Goals (SDGs) dan Pengentasan Kemiskinan,” *Jurnal Universitas Padjajaran*, vol. 6, no. 2, 2016.
- [4] Mustafa, M. S., & Simpen, I. W, “Perancangan Aplikasi Prediksi Kelulusan Tepat Waktu Bagi Mahasiswa Baru Dengan Teknik Data Mining Studi Kasus: Data Akademik Mahasiswa STMIK Dipanegara Makassar),” *Creative Information Technology Journal*, vol. 1, no. 4, hal. 270. 2015. <https://doi.org/10.24076/citec.2014v1i4.27>
- [5] D. T. Larose, *Discovering Knowledge in Data Mining An Introduction to Data Mining*. Wiley Interscience. 2005.

- [6] Y. Mardi, *Analisa Data Rekam Medis untuk Menentukan Penyakit Terbanyak Berdasarkan ICD Menggunakan Decision Tree*. UPI YPTK Padang, 2018.
- [7] Y. Mardi, "Data Mining: Klasifikasi menggunakan Algoritma C4.5," *Jurnal Edik Informatika*, vol. 2, 2016.
- [8] A. Saleh, "Implementasi Metode Klasifikasi Naive Bayes dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga," *Citee Journal*, vol. 2, no.3, 2015.
- [9] V. Agarwal, S. Thakare, and A. Jaiswal, "Survey on Classification Techniques for Data Mining," *Int. J. Comput. Appl.*, vol. 132, no. 4, pp. 13–16, 2015.
- [10] Bustami, "Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi," *TECHSI : Jurnal Penelitian Teknik Informatika*, vol. 3, no.2, hal. 127-146. 2013.
- [11] Pattekari, S. A., Parveen, A., "Prediction System for Heart Disease Using Naive Bayes," *International Journal of Advanced Computer and Mathematical Sciences*, vol. 3, no. 3, hal 290-294. 2012.
- [12] Fleiss, J. L., *Statistical methods for rates and proportion*, 2nd ed. (New York: John Wiley) pp. 38-46. 1981.