

KLASIFIKASI TINGKAT KEMISKINAN DI INDONESIA MENGUNAKAN METODE NAIVE BAYES

Erwin Ferry Syahputra

(180411200161)

Dosen Pembimbing

Dr. Fika Hastarita Rachman, S.T., M.Eng.

Dr. Eka Mala Sari Rochman, S.Kom., M.Kom.

ABSTRAK

Kemiskinan merupakan permasalahan yang masih menjadi tantangan besar di Indonesia. Untuk mendukung upaya penanggulangan kemiskinan, dibutuhkan sistem yang mampu mengklasifikasikan tingkat kemiskinan masyarakat secara akurat dan efisien. Penelitian ini bertujuan untuk mengklasifikasikan tingkat kemiskinan di Indonesia berdasarkan data sosial ekonomi menggunakan metode Naive Bayes yang diimplementasikan dalam software Orange Data Mining. Data yang digunakan mencakup variabel seperti pengeluaran per kapita, rata rata lama sekolah penduduk, indeks pembangunan manusia, umur harapan hidup, presentase rumah tangga yang memiliki akses terhadap sanitasi layak, presentase rumah tangga yang memiliki akses terhadap air minum layak, tingkat pengangguran terbuka, tingkat partisipasi angkatan kerja. Metode Naive Bayes dipilih karena kemampuannya dalam menangani data kategorikal dan memberikan hasil yang baik meskipun asumsi independensi antar fitur tidak sepenuhnya terpenuhi. Hasil penelitian menunjukkan bahwa model Naive Bayes mampu mengklasifikasikan tingkat kemiskinan dengan kinerja yang cukup baik. Pengujian pada data training sebesar 60% menghasilkan akurasi sebesar 86,1%, F1 Score 87,8%, Precision 91,9%, dan Recall 86,1%. Meskipun terjadi sedikit penurunan akurasi pada pengujian dengan 70% dan 80% data training (masing-masing 85,9% dan 85,1%), model tetap mempertahankan tingkat presisi di atas 91%.

Dengan memanfaatkan 514 data kabupaten/kota dan 10 indikator sosial ekonomi, metode ini terbukti efektif dalam mendukung klasifikasi wilayah miskin dan tidak miskin secara efisien. Temuan ini menunjukkan potensi penerapan metode machine learning dalam mendukung kebijakan berbasis data untuk penanggulangan kemiskinan di Indonesia.

Kata Kunci : Kemiskinan, Klasifikasi, *Naive Bayes*, Orange Data Mining, *Machine Learning*.

ABSTRACT

Poverty remains a major challenge in Indonesia. To support efforts in poverty alleviation, a system capable of accurately and efficiently classifying the level of poverty in society is needed. This study aims to classify poverty levels in Indonesia based on socio-economic data using the Naive Bayes method, implemented through the Orange Data Mining software. The data used includes variables such as per capita expenditure, average years of schooling, human development index, life expectancy, percentage of households with access to proper sanitation, percentage of households with access to safe drinking water, open unemployment rate, and labor force participation rate. The Naive Bayes method was chosen due to its ability to handle categorical data and produce good results even when the assumption of feature independence is not fully met. The results of the study indicate that the Naive Bayes model is capable of classifying poverty levels with satisfactory performance. Testing on 60% of the training data yielded an accuracy of 86.1%, an F1 Score of 87.8%, Precision of 91.9%, and Recall of 86.1%. Although there was a slight decrease in accuracy when tested with 70% and 80% of the training data (85.9% and 85.1%, respectively), the model maintained a precision rate above 91%. By utilizing 514 district/city data and 10 socioeconomic indicators, this method has proven effective in supporting the efficient classification of poor and non-poor areas. These findings highlight the potential for applying machine learning methods to support data-driven policies for poverty alleviation in Indonesia.

Keywords : Poverty, Classification, Naïve Bayes, Orange Data Mining, Machine Learning.