

PENERAPAN *DATA MINING* ALGORITMA *DECISION TREE C4.5*
UNTUK MENGLASIFIKASIKAN TINGKAT KEMISKINAN
DI KECAMATAN PAGUYANGAN

Muhammad Yasif Akbar

H1D021011

ABSTRAK

Kemiskinan merupakan permasalahan sosial yang kompleks dan memerlukan pendekatan berbasis data untuk mengidentifikasi tingkat kesejahteraan masyarakat secara lebih akurat. Penelitian ini bertujuan untuk mengklasifikasikan tingkat kemiskinan di Kecamatan Paguyangan dengan menerapkan algoritma *Decision Tree C4.5*, menggunakan data dari dua desa dalam rentang tahun 2019–2024 yang mencakup tiga variabel penting, yaitu pendapatan bulanan, jumlah anggota keluarga, kondisi tempat tinggal. Metode penelitian mencakup pengumpulan data primer dan sekunder, kemudian dilakukan *preprocessing*, transformasi label, normalisasi, serta pembentukan fitur baru berupa pendapatan per kapita sebelum data dilatih menggunakan algoritma *C4.5* dan divalidasi dengan teknik *Stratified K-Fold Cross Validation*. Hasil evaluasi menunjukkan bahwa model yang dibangun memiliki akurasi rata-rata sebesar 99,64% dengan nilai *precision*, *recall*, dan *F1-score* di atas 99%, yang menunjukkan kinerja klasifikasi sangat baik. Model ini kemudian diimplementasikan ke dalam dashboard interaktif berbasis Streamlit sebagai alat bantu bagi pemerintah desa dalam mengidentifikasi masyarakat yang layak menerima bantuan sosial. Penelitian ini membuktikan bahwa penerapan data mining dengan algoritma *C4.5* dapat menjadi solusi efektif dalam mendukung pengambilan keputusan berbasis data untuk penanggulangan kemiskinan di tingkat lokal.

Kata Kunci: Kemiskinan, *Data Mining*, *Decision Tree C4.5*, Klasifikasi, *Streamlit*

**APPLICATION OF DATA MINING C4.5 DECISION TREE ALGORITHM
TO CLASSIFY POVERTY LEVELS IN PAGUYANGAN DISTRICT**

Muhammad Yasif Akbar

H1D021011

ABSTRACT

Poverty is a complex social problem and requires a data-based approach to identify the level of community welfare more accurately. This study aims to classify the level of poverty in Paguyangan District by applying the Decision Tree C4.5 algorithm, using data from two villages in the 2019–2024 period covering three important variables, namely monthly income, number of family members, and living conditions. The research method includes collecting primary and secondary data, then preprocessing, label transformation, normalization, and the formation of new features in the form of per capita income before the data is trained using the C4.5 algorithm and validated using the Stratified K-Fold Cross Validation technique. The evaluation results show that the model built has an average accuracy of 99.64% with precision, recall, and F1-score values above 99%, indicating very good classification performance. This model is then implemented into an interactive dashboard based on Streamlit as a tool for the village government in identifying people who are eligible to receive social assistance. This study proves that the application of data mining with the C4.5 algorithm can be an effective solution in supporting data-based decision making for poverty alleviation at the local level.

Keyword: Poverty, Data Mining, Decision Tree C4.5, Classification, Streamlit