

ABSTRAK

Penelitian ini bertujuan untuk mengembangkan sistem klasifikasi kebutuhan beras di Kabupaten Tapanuli Selatan menggunakan algoritma C4.5. Kebutuhan beras dikategorikan menjadi rendah dan tinggi berdasarkan jumlah penduduk, kebutuhan beras, tahun, kecamatan, dan desa/kelurahan. Pertumbuhan penduduk yang terus meningkat berdampak pada naiknya kebutuhan beras, yang dapat memicu kelangkaan di masa depan. Data yang digunakan berjumlah 744 data sekunder dari Badan Pusat Statistik (BPS) Kabupaten Tapanuli Selatan tahun 2021–2023. Sistem dikembangkan menggunakan *Python* dan *framework Flask* untuk membangun aplikasi web interaktif. Algoritma C4.5 dipilih karena dapat menangani data numerik dan kategorikal serta menghasilkan pohon keputusan yang mudah dipahami. Hasil pengujian menunjukkan bahwa akurasi model meningkat tajam hingga kedalaman pohon ke-3, dengan split 80:20 mencapai 97,99% dan split 70:30 mencapai 97,32%. Pada kedalaman 1 dan 2, akurasi masih rendah, yaitu sekitar 83,22% (80:20) dan 84,38% (70:30). Setelah *depth* 3, akurasi stabil hingga *depth* 15, menandakan tidak ada peningkatan signifikan dan tidak terjadi *overfitting*. Kedalaman optimal adalah 3, dengan split 80:20 memberikan akurasi lebih tinggi. Sistem ini dapat menjadi alat bantu dalam perencanaan dan pengambilan keputusan terkait penyediaan beras.

Kata Kunci: Kebutuhan Beras, Klasifikasi, Algoritma C4.5, Akurasi, Model

ABSTRACT

This study aims to develop a rice demand classification system in South Tapanuli Regency using the C4.5 algorithm. Rice demand is categorized into low and high based on population size, rice needs, year, sub-district, and village/urban area. The continuous population growth leads to an increase in rice demand, which may trigger shortages in the future. The dataset used consists of 744 secondary data records obtained from the Central Bureau of Statistics (BPS) of South Tapanuli Regency from 2021 to 2023. The system was developed using Python and the Flask framework to build an interactive web application. The C4.5 algorithm was chosen for its ability to handle both numerical and categorical data and for producing decision trees that are easy to interpret. Testing results show that model accuracy increases sharply up to a tree depth of 3, with the 80:20 split reaching 98% and the 70:30 split reaching 97%. At depths 1 and 2, accuracy is still low, around 83% (80:20) and 84% (70:30). After depth 3, accuracy remains stable up to depth 15, indicating no significant improvement and no overfitting. The optimal depth is 3, with the 80:20 split yielding higher accuracy. This system can serve as a decision-support tool for rice supply planning and management.

Keywords: *Rice Demand, Classification, C4.5 Algorithm, Accuracy, Model*