

ABSTRAK

Keju merupakan salah satu produk olahan susu yang paling populer. Kadar lemak pada keju dapat bervariasi tergantung pada berbagai faktor selama proses pembuatannya. Penelitian ini bertujuan untuk mengklasifikasikan kadar lemak keju ke dalam dua kategori, yakni *lower fat* dan *higher fat*, dengan membandingkan performa dua algoritma *machine learning*: AdaBoost dan Gradient Boosting. Dataset yang digunakan berasal dari *Canadian Cheese Directory*. Untuk memaksimalkan hapus data dan kinerja model, diterapkan serangkaian tahapan pra-pemrosesan data yang komprehensif. Tahapan ini meliputi penanganan nilai yang hilang, *feature engineering* untuk menciptakan variabel prediktif baru, penyeimbangan data menggunakan *Synthetic Minority Over-sampling Technique (SMOTE)*, dan diakhiri dengan seleksi fitur untuk memilih atribut yang paling relevan. Data yang telah diolah kemudian dibagi menjadi data latih dan sepuluh sampel data uji—lima dari kategori *lower fat* dan lima dari *higher fat*. Hasil pengujian menunjukkan algoritma Gradient Boosting mencapai akurasi sempurna 93%, mengungguli AdaBoost yang memperoleh akurasi 86%. Temuan ini mengindikasikan bahwa setelah melalui tahapan pra-pemrosesan data yang cermat, algoritma Gradient Boosting menunjukkan performa yang superior dalam mengklasifikasikan kadar lemak keju pada dataset ini.

Kata kunci: Keju, Klasifikasi, Pra-pemrosesan Data, *Feature Engineering*, AdaBoost, Gradient Boosting

ABSTRACT

Cheese is one of the most popular processed dairy products. The fat content in cheese can vary depending on various factors during its manufacturing process. This study aims to classify the fat content of cheese into two categories, lower fat and higher fat, by comparing the performance of two machine learning algorithms: AdaBoost and Gradient Boosting. The dataset used was sourced from the Canadian Cheese Directory. To maximize data quality and model performance, a comprehensive series of data pre-processing stages were implemented. These stages included handling missing values, feature engineering to create new predictive variables, data balancing using the Synthetic Minority Over-sampling Technique (SMOTE), and concluded with feature selection to choose the most relevant attributes. The processed data was then divided into training data and ten test data samples—five from the lower fat category and five from the higher fat category. The test results showed that the Gradient Boosting algorithm achieved a superior accuracy of 93%, outperforming AdaBoost, which obtained an accuracy of 86%. These findings indicate that after undergoing meticulous data pre-processing stages, the Gradient Boosting algorithm demonstrates superior performance in classifying cheese fat content on this dataset.

Keywords: Cheese, Classification, Data Pre-processing, Feature Engineering, AdaBoost, Gradient Boosting

