

INTISARI

KLASIFIKASI KUALITAS TEH BERDASARKAN AROMA AMPAS DAUN TEH HITAM DENGAN ALGORITMA *RANDOM FOREST* DAN *EXTREME GRADIENT BOOSTING*

Oleh

Ratih Suci Lestari

20/456556/PA/19743

Teh merupakan minuman olahan yang berasal dari daun *Camellia sinensis* dan merupakan komoditas perkebunan unggulan di Indonesia. Sehingga menjaga kualitas dan mutu teh menjadi faktor utama dalam industri. Penilaian kualitas teh masih mengandalkan indra manusia, yang bersifat subjektif dan bervariasi, terutama dalam menilai kualitas berdasarkan aroma. Masalah tersebut dapat diatasi dengan menerapkan sistem *electronic nose (e-nose)* dan algoritma pembelajaran mesin untuk mengidentifikasi serta membedakan kualitas teh secara lebih objektif. Namun, pembacaan aroma oleh *e-nose* sering kali terganggu oleh kondisi lingkungan maupun ketidakstabilan sensor, yang menghasilkan data menyimpang (*outlier*). *Outlier* ini berdampak pada penurunan performa sistem dalam mengklasifikasikan kualitas teh. Oleh karena itu, diperlukan pengembangan sebuah model klasifikasi yang mampu mengatasi *outlier* pada kualitas teh ke supaya konsisten dengan akurasi yang baik.

Perancangan sistem dilakukan dengan mempertimbangkan kebutuhan penelitian dengan memastikan bahwa setiap langkah menghasilkan sistem sesuai yang diharapkan. Tahapan perancangan sistem dimulai dengan *pre-processing*, yang berupa *data cleaning*, ekstraksi fitur, seleksi fitur, kemudian pemodelan algoritma *Random Forest* dan *XGBoost*, serta pengujian sistem. Model dibandingkan dengan model tanpa proses seleksi fitur untuk mengetahui kinerja algoritma.

Hasil akurasi model *Random Forest* dengan seleksi fitur memiliki akurasi terbaik yaitu 85,83%, tanpa seleksi fitur dengan akurasi 84,83%, sedangkan akurasi model *XGBoost* dengan seleksi fitur yaitu 84,67%, tanpa seleksi fitur 83,50%. Berdasarkan penelitian ini, selain mengatasi *outlier*, tahapan seleksi fitur juga berpengaruh terhadap peningkatan akurasi model.

Kata Kunci : Teh Hitam, *Random Forest*, *XGBoost*

ABSTRACT

CLASSIFICATION OF TEA QUALITY BASED ON THE AROMA OF BLACK TEA DREGS USING RANDOM FOREST AND EXTREME GRADIENT BOOSTING ALGORITHMS

By

Ratih Suci Lestari

20/456556/PA/19743

*Tea is a processed beverage derived from the leaves of *Camellia sinensis* and is a leading plantation commodity in Indonesia. Therefore, maintaining the quality and standards of tea is a key factor in the industry. The assessment of tea quality still relies on human senses, which are subjective and vary, especially when evaluating quality based on aroma. This issue can be addressed by implementing an electronic nose (e-nose) system and machine learning algorithms to identify and differentiate tea quality more objectively. However, aroma readings by the e-nose are often disrupted by environmental conditions or sensor instability, resulting in deviating data (outliers). These outliers negatively impact the system's performance in classifying tea quality. Therefore, it is necessary to develop a classification model capable of handling outliers in tea quality data to ensure consistent and accurate results.*

The system design is carried out by considering the research needs and ensuring that each step produces the expected system outcome. The system design stages begin with pre-processing, which includes data cleaning, feature extraction, feature selection, followed by modeling using the Random Forest and XGBoost algorithms, and finally, system evaluation. The models are compared with those without the feature selection process to assess algorithm performance.

The accuracy results show that the Random Forest model with feature selection achieves the highest accuracy of 85.83%, while without feature selection it achieves 84.83%. The XGBoost model achieves 84.67% with feature selection and 83.50% without. Based on this study, in addition to handling outliers, the feature selection stage also contributes to improving model accuracy.

Keyword : *Black Tea, Random Forest, XGBoost*