PLAGIAT MERUPAKAN TINDAKAN TIDAK TERPUJI

Abstrak

Kopi arabika adalah salah satu jenis kopi yang digemari karena cita rasa dan aroma nya yang khas. Akan tetapi, keberagaman varietas menjadikan kesulitan dalam proses identifikasi. Hal tersebut menjadi krusial karena dapat mempengaruhi kualitas dan preferensi pasar. Penelitian ini bertujuan untuk membuat sistem klasifikasi menggunakan *random forest* untuk mendeteksi varietas kopi arabika. Data yang digunakan berasal dari *Coffee Quality Database from CQI.* D*ari* data tersebut akan dilakukan beberapa penelitian yang meliputi pengujian mengenai seberapa efektif fitur yang terdapat pada data tersebut untuk melakukan klasifikasi terhadap varietas Bourbon, Caturra, Typica menggunakan *information gain*, melakukan eksperimen terhadap variasi jumlah *tree* dan jumlah cross validation untuk mencari model yang memiliki akurasi tertinggi serta melakukan pengujian model terhadap data testing. Sebelum dilakukan proses klasifikasi, dilakukan *pre-procesing* terhadap data set yang meliputi pembersihan data, transformasi data dan seleksi data. Hasil eksperimen membuahkan hasil terbaik berupa model yang memiliki akurasi sebesar 91,2 % dengan parameter *tree* sejumlah 250 dan 7 *fold cross validation*. Pengujian terhadap 11 data testing, 10 data diprediksi dengan benar dan 1 data salah.

Kata Kunci : Klasifikasi, Varietas Kopi Arabika, Random Forest

PLAGIAT MERUPAKAN TINDAKAN TIDAK TERPUJI

Abstract

Arabica coffee is one of the most favorite coffee because the taste and aroma. However, the diversity of varieties presents challenges in the identification process. It's can be important, because it's can affect both quality and market preferences. This study aims to develop a classification system using Random Forest algorithm to detect Arabica coffee varieties. The dataset used is the Coffee Quality Database from CQI. From the dataset, some experiment will be carried out like evaluating the effectiveness of each features for classifying Bourbon, Caturra, and Typica varieties using Information Gain, experimenting with different numbers of trees and folds in cross-validation to find the most accurate model, and testing the model with test data. Before the classification, preprocessing was applied, including data cleaning, data transformation, and data selection. The best experimental result achieved an accuracy of 91.2% using 250 trees and 7-fold cross-validation. Testing on 11 samples resulted in 10 correctly predicted data and 1 incorrect prediction.

Keywords : Classification, Arabika Coffee Bean, Random Forest