

## ABSTRACT

Civet coffee is the world's most expensive and rarest coffee bean. Indonesia was the first country to be identified as the origin of civet coffee. First, it is produced spontaneously by collecting civets feces from coffee plantations near the forest. Due to limited stock, farmers began cultivating civets to obtain safe supplies of civet coffee. Based on this, civet coffee can be divided into two types: wild and fed. A combination of spectroscopy and chemometrics can be used to evaluate authenticity with high speed and precision. In this study, seven samples from different regions were analyzed using NIR Spectroscopy with various preparations: unroasted, roasted, unground, and ground. The spectroscopic data were combined with unsupervised exploratory methods (hierarchical cluster analysis (HCA), and principal component analysis (PCA)) and supervised classification methods (support vector machine (SVM) and random forest (RF)). The HCA results showed a trend between roasted and unroasted beans; meanwhile, the PCA showed a trend based on coffee bean regions. Combining the SVM with leave-one-out-cross-validation (LOOCV) successfully differentiated 57.14% in all sample groups (unground, ground, unroasted, unroasted-unground, and roasted-unground), 78.57% in roasted, 92.86% in roasted-ground, and 100% in unroasted-ground. However, using the Boruta filter, the accuracy increased to 89.29% for all samples, 85.71% for unground and unroasted-unground, and 100% for roasted, unroasted-ground, and roasted-ground. Ultimately, RF successfully differentiated 100% of all grouped samples. In general, roasting and grinding the samples before analysis improved the accuracy of differentiating between wild and feeding civet coffee using NIR Spectroscopy.

## Keywords

Boruta algorithm; civet coffee; ground coffee; hierarchical cluster analysis; principal component analysis; random forest; support vector machine.

## INTISARI

Kopi Luwak adalah biji kopi paling langka dan paling mahal di dunia. Indonesia merupakan negara pertama yang diidentifikasi sebagai asal kopi luwak. Pada awalnya, kopi luwak diproduksi secara spontan dengan mengumpulkan kotoran luwak dari perkebunan kopi di sekitar hutan. Karena stok terbatas, petani mulai memelihara luwak untuk mendapatkan pasokan kopi yang aman. Berdasarkan ini, kopi luwak dapat dibagi menjadi dua jenis: liar dan ternak. Kombinasi spektroskopi dan kemometri dapat digunakan untuk mengevaluasi keaslian dengan kecepatan dan presisi tinggi. Dalam penelitian ini, tujuh sampel dari daerah yang berbeda dianalisis dengan menggunakan NIRS dengan berbagai perlakuan: tidak disangrai, disangrai, tidak digiling, dan digiling. Data spektroskopi dikombinasikan dengan metode unsupervised (hierarchical cluster analysis (HCA) dan principal component analysis (PCA)) dan metode supervised (support vector machine (SVM) dan random forest (RF)). HCA menunjukkan tren antara biji kopi yang disangrai dan tidak; sementara itu, PCA menunjukkan tren berdasarkan daerah biji kopi. Gabungan antara SVM dengan leave-one-out-cross-validasi (LOOCV) berhasil membedakan 57,14% di semua kelompok sampel (unground, ground, unroasted, unroasted-unground, dan roasted -unground), 78.57% pada roasted, 92.86% pada roasted-ground, dan 100% pada unroasted-ground. Namun, dengan menggunakan filter Boruta, akurasi meningkat menjadi 89,29% untuk semua sampel, 85.71% untuk unground dan unroasted-unground, serta 100% untuk roasted, unroasted-ground, dan roasted-ground. Terakhir, RF berhasil membedakan 100% dari semua sampel yang dikelompokkan. Secara umum, menyangrai dan menggiling sampel sebelum analisis dapat meningkatkan akurasi dalam membedakan antara kopi luwak liar dan ternak menggunakan NIRS.

## Keywords

Boruta algorithm; ground coffee; hierarchical cluster analysis; kopi luwak; principal component analysis; random forest; support vector machine.