

ABSTRAK

Klasifikasi kualitas *red wine* dilakukan dengan harapan dapat mempermudah dalam melakukan penilaian kualitas *red wine*. Data yang digunakan untuk penelitian ini adalah data fisikokimia *red wine* yang berjumlah 1599 *records*, dan didapatkan dari *website* Kaggle. Klasifikasi kualitas *red wine* dilakukan dengan menggunakan metode *random forest*, karena metode ini memiliki akurasi yang baik, dan relatif kuat terhadap *noise* data. *Preprocessing* dilakukan untuk mempersiapkan data agar dapat diolah ke dalam algoritma. *Preprocessing* yang dilakukan adalah pengecekan terhadap *noise* dan *missing value*, dan transformasi data menggunakan metode normalisasi *min-max*. Setelah *preprocessing* dilakukan, maka selanjutnya dilakukan *balancing* kelas data minoritas dengan menggunakan metode SMOTE. Proses pembagian data *testing* dan data *training* dilakukan dengan metode *K-Fold Cross Validation*. Pada tahap pengujian, digunakan *3-fold*, *5-fold*, *7-fold*, *9-fold*, dan *11-fold* dan menggunakan kombinasi pohon 2^n , di mana $n = 1, 2, \dots, 11$. Berdasarkan pada pengujian yang dilakukan, didapatkan hasil akurasi 82.3478% dengan menggunakan *11-fold* dan 1024 pohon.

Kata kunci: *Confusion Matrix*, klasifikasi kualitas *red wine*, *K-Fold Cross Validation*, *Random Forest*,

ABSTRACT

Classification of red wine quality is done in the hope of making it easier in performing red wine quality assessments. The data used for this research was red wine physicochemical data totaling 1599 records, and was obtained from the *Kaggle* website. Classification of red wine quality is done using the random forest method, as this method has good accuracy, and is relatively strong to data noise. The preprocessing process is done to prepare the data in order to be processed into the algorithm. Preprocessing performed is checking against noise and missing value, data transformation using min-max normalization. Once the preprocessing is done, then it is further performed balancing the class of minority data using the SMOTE method. The process of data sharing testing and data training is done by the K-Fold Cross Validation method. At the testing stage, it is used 3-fold, 5-fold, 7-fold, 9-fold, and 11-fold and uses a combination of $2n$ trees, where $n = 1, 2, \dots, 11$. Based on the test carried out, there was an accuracy result of 82,3478% using 11-folds and 1024 trees.

Keywords: *Confusion Matrix, K-Fold Cross Validation, red wine quality classification, Random Forest.*

