

ABSTRAK

Di era yang serba modern ini semua hal dapat diidentifikasi berdasarkan kriteria atau atribut yang di miliki, mulai dari benda, hewan, tanaman, dan semua hal yang memiliki materi dan dapat di definisikan. Penelitian ini bertujuan untuk mengklasifikasikan jenis kacang, menggunakan metode *decision tree C4.5*. Metode ini digunakan untuk membuat model klasifikasi yang dapat mengelompokkan kacang ke dalam kelas yang berbeda berdasarkan atribut-atributnya. Data yang digunakan pada penelitian ini data yang berasal dari peneliti bernama Murat Koklu, dengan data yang terbagi menjadi 7 *class* yaitu *Seker 2027 data*, *Barbunya 1322 data*, *Bombay 522 data*, *Cali 1630 data*, *Dermason 3546 data*, *Horoz 1928 data* dan *Sira 2636 data* dengan total data 13611 dan memiliki atribut data berjumlah 16 jenis. Data yang diperoleh akan di normalisasi lalu dilakukan penyeimbangan data karena data tidak seimbang tiap *Class* dengan teknik SMOTE (Synthetic Minority Over-sampling Technique), berikutnya dilakukan pembagian data menggunakan K-Fold Cross Validation. Klasifikasi yang digunakan pada penelitian ini adalah *decision tree C4.5*. Metode ini dapat membangun pohon keputusan yang berisi aturan-aturan klasifikasi, sehingga dari penelitian ini diharapkan menunjukkan bahwa metode *decision tree C4.5* dapat mengklasifikasikan jenis kacang dengan akurasi yang baik.

Kata Kunci : klasifikasi, *Decision tree C4.5*, Kacang, SMOTE

ABSTACT

In this modern era, everything can be identified based on its criteria or attributes, ranging from objects, animals, plants, and all tangible matters that can be defined. This study aims to classify bean types using the Decision Tree C4.5 method. This method is employed to develop a classification model capable of grouping beans into different classes based on their attributes. The data used in this study comes from research conducted by Murat Koklu, consisting of seven classes: Seker with 2,027 data points, Barbunya with 1,322 data points, Bombay with 522 data points, Cali with 1,630 data points, Dermason with 3,546 data points, Horoz with 1,928 data points, and Sira with 2,636 data points, totalling 13,611 data points with 16 attributes.

The collected data is normalized and then balanced due to class imbalance using the SMOTE (Synthetic Minority Over-sampling Technique). Subsequently, the data is split using the K-Fold Cross Validation method. The classification method applied in this study is Decision Tree C4.5, which constructs a decision tree containing classification rules. This research is expected to demonstrate that the Decision Tree C4.5 method can effectively classify bean types with high accuracy.

Keywords: classification, Decision Tree C4.5, beans, SMOTE