Data Partitioning and Modeling

Coursera Big Data Specialization Capstone Project, Week 2

Peer Graded Assignment: Classifying in KNIME to identify big spenders in Catch the Pink Flamingo

The data was partitioned into train and test datasets.
The **training** data set was used to create the decision tree model.
The trained model was then applied to the **test** dataset.
This is important because **partitioning the data set into training and test data allows us to verify the accuracy of the trained model.**

When partitioning the data using sampling, it is important to set the random seed because **it allows you to obtain reproducible results each time you run the partition.**

A screenshot of the resulting decision tree can be seen below: