

COFFEE BEAN CLASSIFICATION

NASSER ALQAHTANI
MOHAMMED ALZABYEDI

ABSTRACT

Proceeding from the vision of the Kingdom of Saudi Arabia 2030 and its announcement of the Quality of Life Program, which aims to raise the level of citizen's well-being by creating the appropriate environment for this, one of the most important goals is to raise the number of cafes from 171 per million people to 1032 per million people by 2030.

Data was taken for a number of coffee crops in a number of famous countries in the field and the coffee quality was predicted. In the future, entrepreneurs in the field of coffee will be contacted, and the project will be presented and developed according to customer needs.

DESIGN

By applying the dataset on machine learning models such as LogisticRegression, KNN, Decision Tree, Random Forest, Voting, and Bagging.

DATA

The dataset contains 2650 crops with 44 features for each.. A few feature highlights include Aroma, Flavor, Aftertaste, Acidity, Body, Balance, Uniformity, Clean.Cup, Sweetness, Cupper.Points, Total_Cup_Points, and Moisture.

ALGORITHMS

- Preparing the data and feature selection:
 1. Exploration of the data and visualization.
 2. Feature Selection by calculating the correlation of the features.
- Methods:
LogisticRegression, KNN, Decision Tree, Random Forest, Voting, and Bagging. Splitting the dataset by using cross validation to measure each model score.

TOOLS

- Python and Jupyter Notebook.
- Numpy and Pandas for data manipulation.
- Matplotlib and Seaborn for plotting visualization.
- Sklearn for ML algorithms.
- Imblearn to solve imbalance problem.