PROJECT Group 3

The problem described in our final project is the classification of chocolate bars’ rating. Chocolate has been everyone’s favorite flavor, so it is mandatory to keep the contents of chocolate in right proportion to deliver the best possible taste. To satisfy different demands, chocolate companies provide consumers with different types of chocolates varying the contents of cocoa in it. The taste of chocolate flavor varies considerably on varying the contents of its ingredients. Therefore, we want to provide chocolate companies with rating reference when they developed a new product. According to our classification model, we will rate the taste of the chocolate bar depending on its cocoa percent.

In our final project, a classification model of chocolate bars’ rating is built. The datasets we used contains around 1800 records and consist of 9 variables: company name, the specific geo-region of origin for the bar, a value linked to when the review was entered in the database., review date, cocoa percent, company location, rating, bean type, and the broad geo-region of origin for the bean. In the end, rating value from 1 to 5 is our final target.

Decision tree will be introduced in our classification model. From decision tree, Gini impurity and Information gain are metric to evaluate. As for the framework, scikit learn, plotplus, metaplotlit, numpy, pandas will be implemented in this task. Scikit learn will be used for computing model, and the rest are used to organize and display results.

In the end, according to our decision tree model, the chocolate companies can predict their ranking based on what cocoa bean and cocoa percentage is adopted in a new product. We are expecting the project to be completed before 11th August so that we will have time to make any required changes.