DS 110 Presentation Script v2

Thank you Kayla for the wonderful introduction. Now lets talk about our methods.

We used both decision tree and random forest regression.

And then split the data into training and testing data with random.seed(), which ensure that the data is split in the same way each time we run the model.

Then we used the dtmtoCode to turn our trees into functions, which allows us to easily understand and interpret the results.

For parameter tuning, we used a cross-validation method called GridSearchCV, to systematically working through multiple combinations of parameter values to find the best suited one for the model.

Lastly, we calculated the accuracy based on the Mean Absolute Percentage Error (MAPE), with further insights provided by the Root Mean Squared Error (RMSE) and R2 score to evaluate the model's performance.

So, we have found that cocoa percentage has the greatest impact on chocolate bar ratings, with higher cocoa percentages leading to higher ratings. This was

consistently observed with both models, with accuracy of 86.7% and 87%.

Our main takeaway from this research is that by checking the data for unexpected or incorrect values and confirming its validity before using it to train the model, we can make sure that it is trained on reliable and accurate data, which again will enhance the accuracy of the predictions it provides.

Here is our reference, and thank you for being here with us today.