9. Outcome

From the Exploratory analysis, the Aftertaste of coffee is largely dependent upon these attributes Flavor, Acidity, Cupper Points, Balance, Aroma, Body, Uniformity, Clean Cup, Sweetness, Moisture, and Category Two Defects. When these attributes are visually analyzed, Moisture Category Two Defects, Uniformity, Clean Cup, and Sweetness have minimal impact in Aftertaste. Hence, Flavor, Acidity, Cupper Points, Balance, Body, and Aroma affect the Aftertaste of coffee largely. The Flavor feature has a correlation of 0.9 with Aftertaste. While visualizing the relationship between Aftertaste and Flavor, the data points densely lie along a positively skewed line. Thus, Flavor is the most important feature which impacts significantly on the Aftertaste of Coffee. This level of importance is followed by Balance and Cupper Points.

A dataframe with these final features is selected along with the target variable. This frame is divided into training and testing sets. The training set is used in training six different machine learning algorithms. Each model is hyper-tuned manually. These trained models are evaluated by both training set and testing set using the R-square approach. Most the model gets overfitted and the Decision Tree model have the largest gap between training and testing accuracies whose respective values are 99.9% and 55.5%. This problem is minimal in Linear Regression with a training accuracy of 79.3% and testing accuracy of 78.5%.

In the second case, we take only four features dropping Aroma and Body. Thus, the features that are taken are Flavor, Balance, Cupper Points, and Acidity. Again, the models are trained and evaluated similarly. The problem of overfitting is minimal in most of the models except the Decision Tree. The training and testing accuracies in the Decision Tree are 98.3% and 64.3% respectively. Though, the testing accuracy increases in the Decision Tree model, it is still enormously underperforming in unseen data. The Support Vector Machine performs better than others. Its training and testing accuracies are 82.7% and 81.7% respectively and this model is saved for future purposes.

10. Future Work

Future enhancements of this project are listed below:

i. Though we develop the model. But it is not deployed in a real-world scenario. It can be deployed through a proper website/app and many people like research scientists, coffee sellers, farmers, and mills can use this deployed model easily for predicting and analyzing the overall coffee Aftertaste.

ii. Since we have tested the model performance using training and testing data set which were made by splitting the initially given dataset. We do not know how our analysis and model are relevant in a real-world case. Detail research can be done on this analysis and model performance by making a new dataset of the latest coffee data.

iii. The neural network can be used in training the new model. Since the neural network can deal with large data and can find the most complex pattern within data. Thus, we can increase the feature space and train the model.

Furthermore, the dataset has 44 columns of different data. In this project, only the factors affecting the Aftertaste of coffee are analyzed. In the future, the target variable can be changed to any other like Category Two Defects, Category One Defects, etc. and the features impacting them can be studied. There are many other features of Farm metadata which include Country of Origin, Region, Producer, etc. These sorts of data can be used in making advanced analyses of coffee production scenarios based upon these scenarios. These future findings might be beneficial for coffee exporters, sellers.

11. References

[1] National Coffee Association USA, "What is Coffee?," NCAUSA, [Online]. Available: https://www.ncausa.org/About-Coffee/What-is-Coffee. [Accessed 23 July 2021].

[2] International Coffee Organization, "Coffee Market Report," October 2020. [Online]. Available: https://www.ico.org/news/cmr-1020-e.pdf. [Accessed 21 July 2021].

[3] K. Bernard, "WorldAtlas," 6 August 2020. [Online]. Available: https://www.worldatlas.com/articles/top-10-coffee-consuming-nations.html. [Accessed 23 July 2021].

[4] Espresso and Coffee Guide, "What is Coffee Quality?," Espresso and Coffee Guide, 2006. [Online]. Available: https://espressocoffeeguide.com/quality-coffee/. [Accessed 23 July 2021].

[5] G. Oden, "Java Presse," [Online]. Available: https://www.javapresse.com/blogs/enjoying-coffee/how-to-taste-coffee-aftertaste. [Accessed 23 July 2021].