**Homework 4**
Problem 2
Ayad Masud
733009045

I certify that I have <u>personally</u> done the coding, generated the figures and written the report without aid from anybody else, and that I have not plagiarized, self-plagiarized, or used AI-generated text. I certify that I have acknowledged any sources I used to complete this assignment. ARM.

## 1 Part 1: Train Bagged Tree Models

Figure 1 shows the performance of individual trees in the ensemble on both training and validation sets. The box plots illustrate the distribution of accuracies for individual trees in the ensemble. The left plot shows the training set, and the right plots shows the validation set. We can see how the number of trees in the ensemble affects the performance of each tree. For both training and validation sets as the number of trees goes up the accuracy stays around the same, however, looking between the accuracy of both sets we can see higher accuracy in the validation set. This shows that the model is generalizing well which is a good sign for the model.
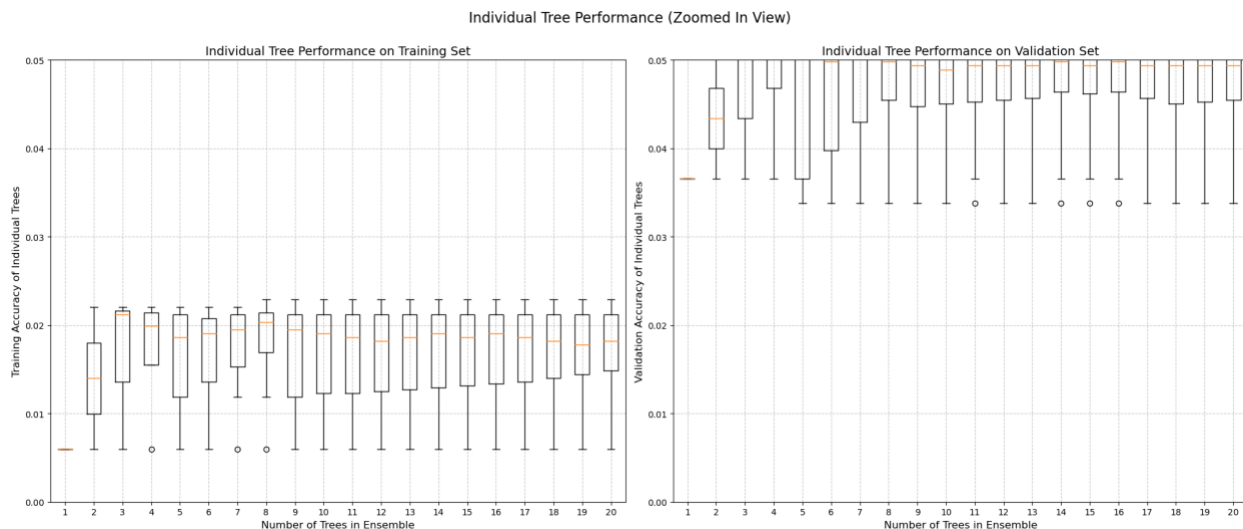
Figure 1. Tree performance on training and validation sets with increasing number of trees in ensemble from 1 to 20.

2 Part 2: Variable Importance Plot

Figure 2 is a plot of the top 15 most important features for food category classification. The most important feature is feature 15 which is ash. Ash is used to see the mineral content in a food. Higher amounts of ash indicate high mineral content and indicates more processing, while less ash is less mineral content has indicated more natural and whole foods. Another very important feature is feature is feature 13 which is potassium. Potassium is used to tell whether a food is part of a healthy die as with high potassium, or low potassium, which may be better for people with certain dietary needs or medical conditions. Another important feature that was found is feature 4 which is total lipids or fats. Fats are extremely useful in classifying foods because natural foods with high oil or fat content can be classified as some type of nut, seeds, or avocados, while foods with low fat content can be classified as vegetables or grains. Overall, this graph shows the most important features that are used to classify food categories in this dataset.
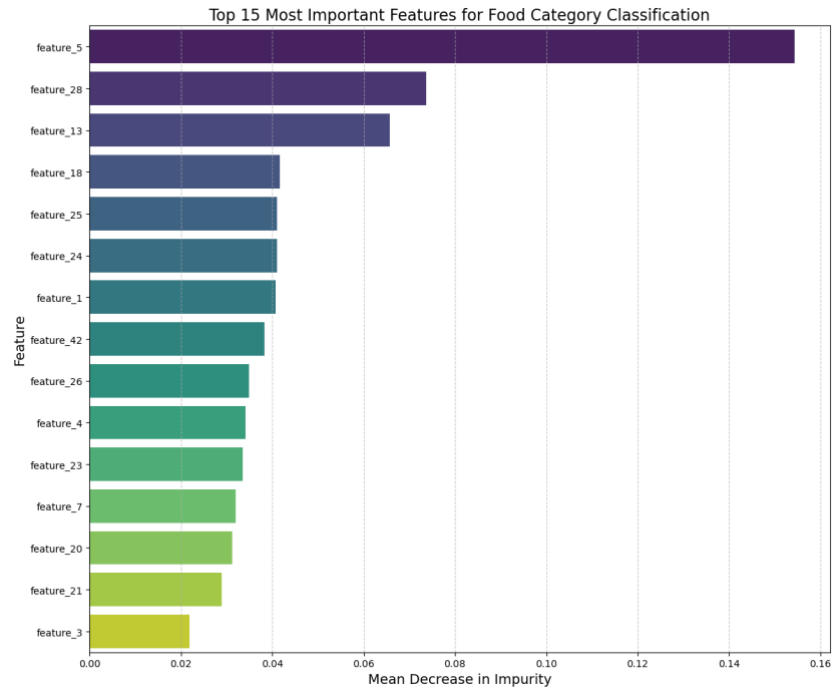
Figure 2. Top 15 most important features in food category classification.

3   Resources used to achieve this goal

Canvas: Homework template

Python Libraries: NumPy, pandas, seaborn, matplotlib, sci-kit learn

4   References

scikit-learn. "1.11. Ensemble Methods — Scikit-Learn 0.22.1 Documentation." *Scikit-Learn.org*, 2012,

scikit-learn.org/stable/modules/ensemble.html.