**Machine Learning in Finance - Homework 4 – Due 10.01.2024**

Textbook reading:
Chapters 5 to page 129

- **start from 'max depth DecisionTreeExample.ipynb' in colab**
- **read in the load_wine database from https://scikit-learn.org/stable/datasets/toy_dataset.html**
- **create decision tree, random forest, and gradient boost trees of depth 3**

- **Display a pandas data frame containing the feature importances with columns the 3 tree types and rows for the features**

| Feature | Decision Tree | Random Forest | Gradient Boosting |
|---|---|---|---|
| alcohol | 0 | 0.065064 | 0.012358 |
| malic_acid | 0 | 0.030743 | 0.000018 |
| ash | 0.025425 | 0.00847 | 0.013945 |
| alcalinity_of_ash | 0 | 0.034946 | 0.002866 |
| magnesium | 0 | 0.019937 | 0.015946 |
| total_phenols | 0 | 0.080241 | 0.000005 |
| flavanoids | 0.427903 | 0.173944 | 0.103352 |
| nonflavanoid_phenols | 0 | 0.01535 | 0.000012 |
| proanthocyanins | 0 | 0.040104 | 0.000397 |
| color_intensity | 0.410289 | 0.206067 | 0.28147 |
| hue | 0 | 0.017922 | 0.030519 |
| od280/od315_of_diluted_wines | 0 | 0.084512 | 0.253884 |
| proline | 0.136383 | 0.222701 | 0.285227 |

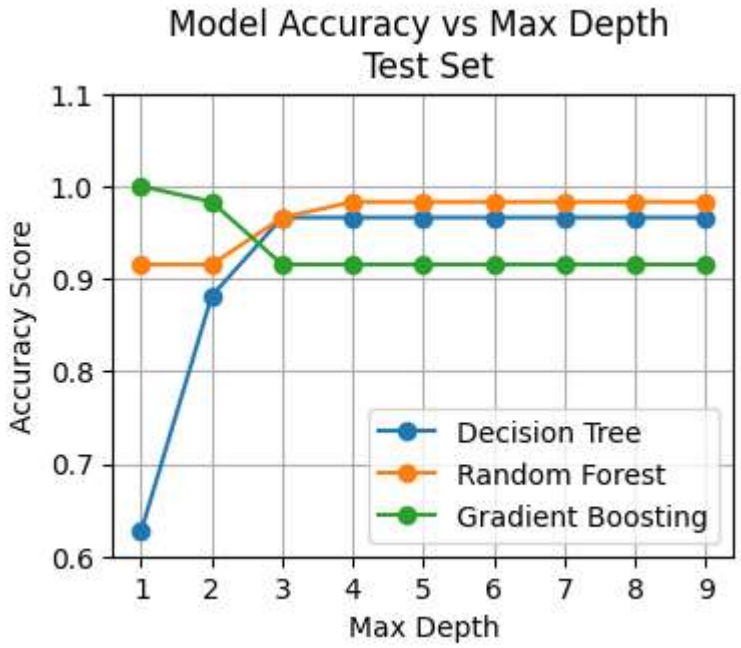
- **What are the 3 most important features for each of 3 different tree methods?**

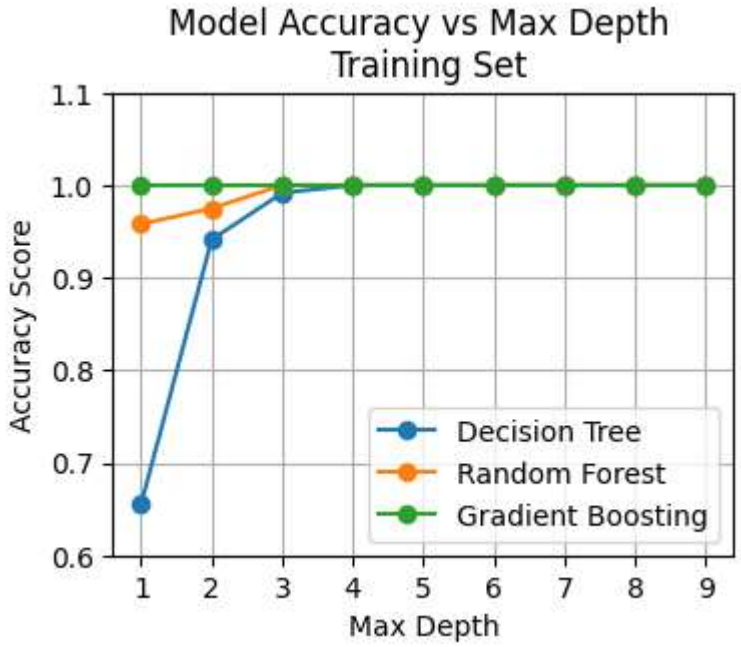
  Decision Tree: flavonoids, color_intensity, proline

  Random Forest: proline, color_intensity, flavonoids

  Grandient Boosting: proline, color_intensity, od280/od315_of_diluted_wines

- **Compute the accuracy for the 3 methods as you go from a depth of 1 up to a depth of 9 and display in a graph? (hint: save the outputs for each depth to a pickle file and write code to read the back in once you have created all the pickle files)**




- **Given this information (since you don't have a test set), argue that there is a best max depth**

We do have a test set (?). Max depth on Gradient Boosting Classifier, surprisingly, has a maximum accuracy at 1. For the other two models, a max depth of 3 seems to be the most optimal choice as accuracy score not longer improves.