

CS 457 - Homework Assignment 7: Classification

Due Date: Monday, October 28 at 11:59 pm

Purpose:

Demonstrate understanding of Classification technique for prediction and recommendation

Points: 100

Deliverables: Submit ipynb code file along with your answer

Use the dataset `Admission_Chance_Classification.csv` and `burger-king-menu-2.csv`

- Perform analysis on the following questions. Make sure to include interpretation of each result including accuracy, confusion matrix and other necessary information to support your answer

Classification using `Admission_Chance_Classification.csv`

1. Create a train and test set. Consider **Admit** column as class/label column (Y) and use rest of the columns as inputs (X). Use 30% (`test_size=0.3`) records for test set. Use the same train and test set for all your analysis with different classifiers.
2. In your code, set the seed after you read the data. This will keep your data and calculation consistent throughout the analysis irrespective of multiple runs. See the example code for the class `random_state=99`
3. Create a classification model to predict Admit status **using Decision Tree**. Visualize the decision tree. Interpret the decision tree. Discuss which attributes are important and which are not important.
4. Perform the Tree Pruning Analysis and evaluate the results. Visualize the tree after pruning. Discuss the tree and overall results before and after pruning.
5. Create a classification model to predict Admit status **using RandomForest**. Include all the analysis steps including variable importance plot. Try at least 5 different values of `n_estimators` (number of trees) such as 10, 50, 100, 150, 200 and compare the classification accuracy for each of them. Pick the best model (based on `n_estimators`) for your final comparison.
6. Report the comparison between Decision Tree and RandomForest in terms of classification performance.

Recommendation using `burger-king-menu-2.csv`

1. Build a decision tree to recommend similar food based on their attributes. Interpret your decision tree and tell some interesting insights from it. You do not (necessarily) need to pick all the food (rows) and attributes (columns). Pick 1 category (Burgers, Chicken or Breakfast) of your choice and explain/interpret your recommendations. *Note: After you filter your data based on Category, you need to remove Category column from the data and then create a decision tree. Use Item column as your class/label (Y)*