Assignment 1 - Set 7

Start Assignment

Due Jan 1, 2022 by 23:59 **Points** 10 **Submitting** a file upload **Attempts** 0
Allowed Attempts 1 **Available** Dec 12 at 0:00 - Jan 1, 2022 at 23:59 21 days

Assignment Naïve Bayes

1. Follow the instructions in each question carefully.
2. A Jupyter notebook along with the output for each cell is expected.
3. Any assignment submitted using other python IDEs is not considered for grading.
4. Use appropriate labels for all visualizations.
5. Upload the output.csv file along with the notebook when required.
6. **If the dataset link is expired, search for the same dataset online from any repository and use it.**

Considering the Census Income data with response variable as Income(>50k and <=50k).

1. Import the CSV dataset from https://www.kaggle.com/uciml/adult-census-income (Links to an external site.) .
2. Identify the presence of missing values, fill the missing values with mean for numerical attributes and mode value for categorical attributes.
3. Extract X as all columns except the Income column and Y as the Income column.
4. Split the data into a training set and testing set.
5. Model the classifier using GaussianNB, BernoulliNB, and MultinomialNB
6. Compute the accuracy and confusion matrix for each model.
7. Plot the decision boundary, visualize training and test results of all the models

2. Create an output .csv file consisting of actual Test set values of Y (column name: Actual) and Predictions of Y(column name: Predicted)