AIML Online

Frequently Asked Questions in Problem Statement

Course: Ensemble Techniques

PART - A [30 Marks]

** Direct or Self-explanatory questions are not covered in this FAQ.*

1. Data Understanding & Exploration:

1 D. Verify if all the columns are incorporated in the merged DataFrame by using a simple comparison Operator in Python. [1 Marks]

→ *Compare the columns of the original dataframes and the merged ones along with the dimensions of the dataframes (no. of columns & no. of rows).*

2. Data Cleaning & Analysis: [5 Marks]

2 A. Impute missing/unexpected values in the DataFrame. [2 Marks]

→ *Unexpected values are nothing but irrelevant values of those columns or any other special characters or empty spaces.*

2 B. Make sure all the variables with continuous values are of 'Float' type. [2 Marks]

→ *Convert all numeric variables to float and verify.*

2 C. Create a function that will accept a DataFrame as input and return pie-charts for all the appropriate Categorical features. Clearly show percentage distribution in the pie-chart. [4 Marks]

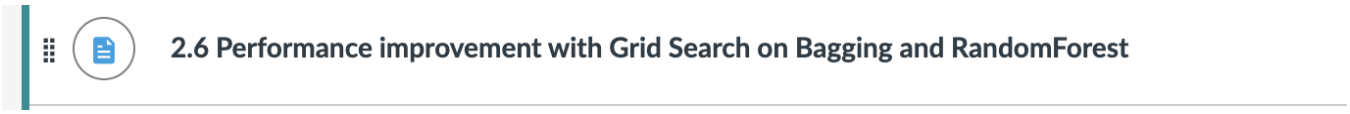
→ *You have to define a function and within that create a loop to select columns that are object type. Exclude unnecessary columns. Calculate the % within the loop and plot a pie chart within the loop.*

Hint: use value_counts to calculate the % of distribution for pie chart

3. Model building and Performance improvement: [10 Marks]

Model building to be performed on Decision tree , Random forest, Adaboost and Gradient boost

Performance improvement on these models to be done using Grid search , which is explained in the video " Performance improvement with Grid search on Bagging and Random forest " , Grid search technique can be applied to all the four models in the question



Please continue to explore the GridSearchCV

Reference link -

https://www.mygreatlearning.com/blog/gridsearchcv/