ML Homework Report

Dataset 1: ID: 971

Description:
This is a binarized version of the original data set. The multi-class target feature is converted to a two-class nominal target feature by re-labeling the majority class as positive ('P') and all others as negative ('N'). Originally converted by Quan Sun.

Features: `['att1','att2','att3','att4','att5','att6''att7','att8','att9','att10','att11','att12','att13','att14','att15','att16','att17'….]`

Target: This is a binary dataset with two Categories (2, object): ['P', 'N']

Dataset 2: ID: 1056

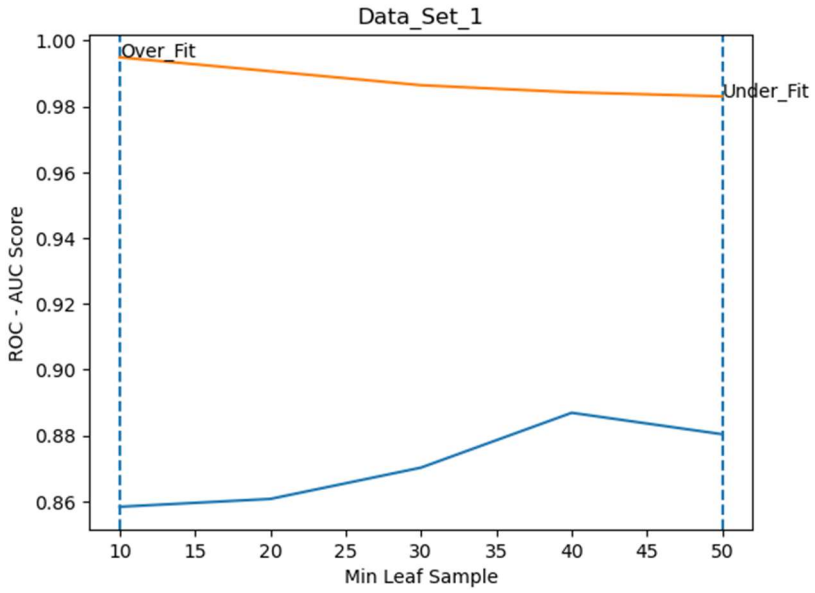
Description:
The report describes a software defect prediction dataset called MC1, which is part of the NASA Metrics Data Program. The author, Mike Chapman, is affiliated with NASA. The source of the report is tera-PROMISE, a database of software engineering datasets.

Features: `['LOC_BLANK','BRANCH_COUNT','CALL_PAIRS','LOC_CODE_AND_COMMENT''LOC_COMMENTS','CONDITION_COUNT','CYCLOMATIC_COMPLEXITY','CYCLOMATIC_DENSITY','DECISION_COUNT','DESIGN_COMPLEXITY','DESIGN_DENSITY','EDGE_COUNT…]`

Target: This is also a binarized Categories (2, object): ['FALSE', 'TRUE']

Subtask-1

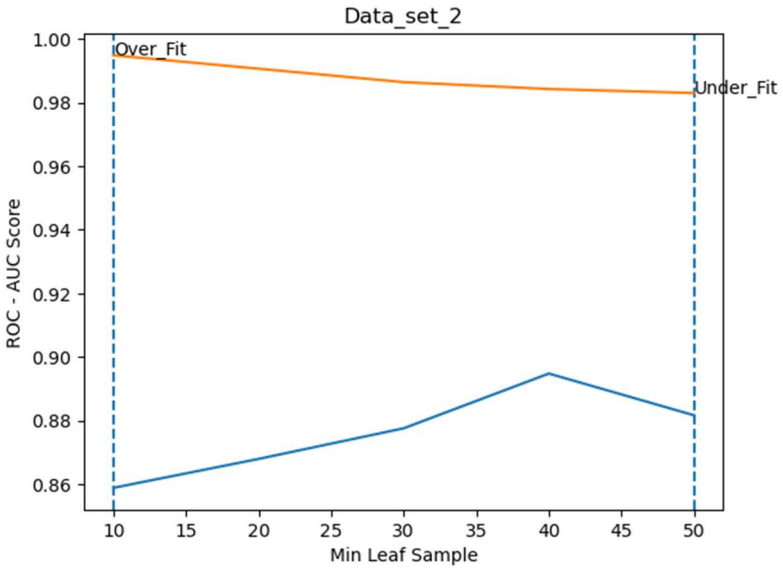
Dataset: 1
Mean score: 0.9177248196199151
Plot values: [10,20,30,40,50]



Here the graph has been plotted against the ROC-AUC results and the Sample leaf values which are depicting the over_fit and under_fit regions. And we have found the best value also at 50.

Dataset: 2

Mean Score: 0.9418745433773725
Plot Values: [10,20,30,40,50]



Here the graph has been plotted against the same categories as the first dataset, and we have found the best results occur at 50.

Subtask: 2

Dataset 1

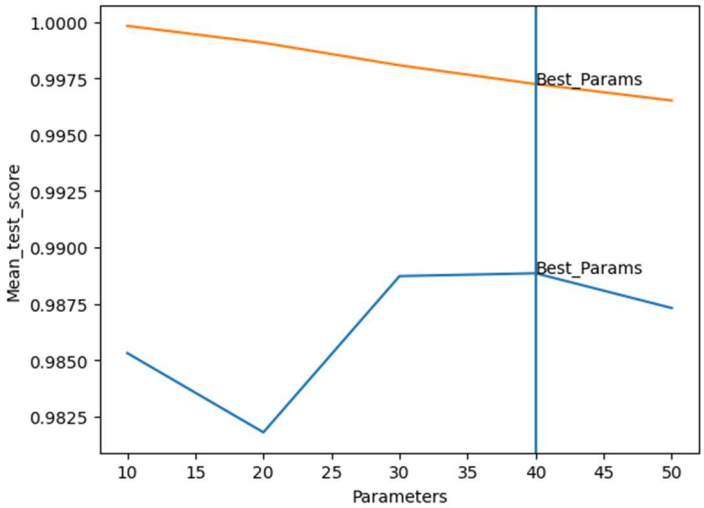
We can see the min_sample leaf and parameter values which are reported by the program are `{'min_samples_leaf': 40}`

The mean score value are as : 0.9863944444444446

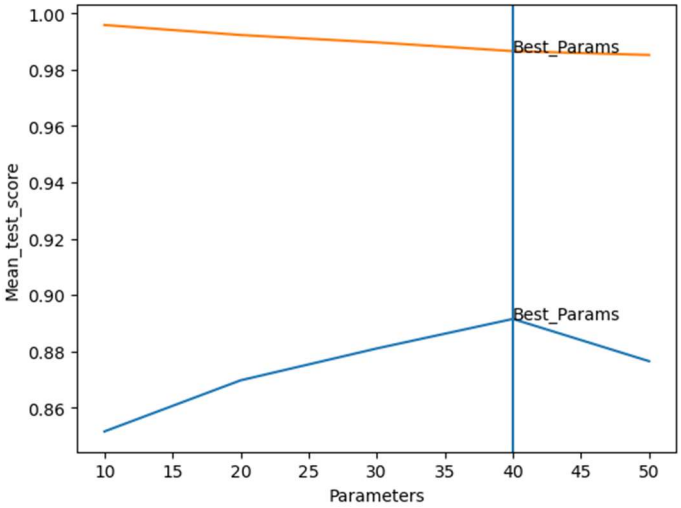
Dataset: 2

For dataset 2 also we have the nearly same parameter value as: `{'min_samples_leaf': 40}`

The mean score value for this graph is as: 0.8740081320747999



For Dataset: 1



For Dataset: 2