|  |  |
| --- | --- |
| NAME: RAHUL THAMBI | ROLL NUMBER: A059 |
| SUBJECT: DWM | PROJECT-3 |

**INCREASING THE ACCURACY:-**

[**https://www.kaggle.com/code/vinothan/titanic-model-with-90-accuracy**](https://www.kaggle.com/code/vinothan/titanic-model-with-90-accuracy)

In the above link the method shown was implemented using python but I have taken useful steps from the tutorial and increased the accuracy of the titanic dataset.

**Step – 1 :** Initially i have preprocessed the data by removing the name attribute since it is redundant in finding whether the person survived or not and it is out of scope with our problem statement.

**Step – 2:** Identified missing values ,missing values were in the Age column, and replaced the missing values with the mean Age of the column

**Step - 3:** Converted the numerical value in the class attribute (Survived) into nominal values of Yes and No (below is the updated dataset)

**Table

Description automatically generated**

**Step – 4:** Used Ensemble technique of AbsoluteSelectedClassifier which is a meta-classifier in Weka that selcts a specific classifier based on a per-defined criterion. This meta classifier allows us to select a classifier based on the performance of the base classifier on a particular data set. Then the performance of each classifier in the list is evaluated using cross-validation, and the best performing classifier is selecred based on the the threshold value. If none of the classifiers in the list meet the threshold value then the classifier with the highest performance is selected.