

IEEE WIE Recruitments: Technical and Projects

Machine Learning Task (Report)

Name: Rohini Das

Reg.No.: 20BCB0098

Question Number: 2

Date: 06.02.2022

AIM: Using an appropriate machine learning algorithm, to make a model which takes the person's input as its age, gender and fare and predicts whether the person survived or died when the RMS Titanic sank in the North Atlantic ocean.

ALGORITHM: Decision Tree Algorithm has been used to build an ML model to accomplish the given task. I used the CART Decision Tree Algorithm. It uses the “gini” as its criterion instead of entropy.

CART (Classification and Regression Trees) can be used for both classification and regression problems. With classification, we attempt to predict a class label, e.g., whether the passengers survived or not.

- In the CART decision tree, the nodes are split into sub-nodes on the basis of a threshold value of Gini Index.
- The root node is taken as the training set and is split into two by considering the best attribute and threshold value.
- Further, the subsets are also split using the same logic. This continues till the last pure sub-set is found in the tree or the maximum number of leaves possible in that growing tree.

DATASET USED: Titanic Dataset from Kaggle (<https://www.kaggle.com/c/titanic-dataset/data>)

- The training dataset is used to build our machine learning models. For the training set, the outcome is given for each passenger i.e. “survived”. Our model is based on “features” like passengers’ age, gender, and fare.
- The test dataset is used to see how well our model performs on unseen data. For the test dataset, the final survival status for each passenger is not given. Our model would predict these outcomes for the unseen data.

PROGRAM & OUTPUT: I have written well-documented code, explaining each and every code block and also the general flow of execution of my solution for the task. For reference I have attached a copy of the Colab notebook consisting of the documented code and explanations, below.