Hackathon Report

AI20BTECH11010-haritha.R

Procedure:

1)loading the files into dataframe

2)eyeball the data

3)dropping the columns containing more null values

4)Processing the features

5)training the model

6)predicting the output

Classifier used: LightGBM classifier

Train accuracy LightGBM classifier.: 0.9263413898544045

Score in Kaggle : 0.85962

Advantages of this classifier:

Produced good training accuracy without overfitting or underfitting. The execution time was very low.

Reason for choosing this model:

LightGBM is a gradient boosting framework that uses tree-based learning algorithms. It is designed to be distributed and efficient as compared to other boosting algorithms.

LightGBM can handle a large amount of data, less memory usage, has parallel and GPU learning, good accuracy, faster training speed and efficiency. So what makes LightGBM a better model, well for one it grows the tree Leaf Wise while other algorithms grow level wise.

LightGBM aims to reduce complexity of histogram building (O(data * feature)) by down sampling data and feature using *GOSS* and *EFB.* This will bring down the complexity to (O(data2 * bundles)) where data2 < data and bundles << feature.