BVRIT HYDERABAD COLLEGE OF ENGINEERING FOR WOMEN

TEAM 7 - THE SCRABBLE GAME

DEPARTMENT OF CSE(AIML)2020-2024

20WH1A6632 - K.Nikitha

20WH1A6633 - Y.Yashaswini

20WH1A6634 - Tasneem Firdous

20WH1A6635 - V.Namitha

20WH1A6636 - G.Amulya

01/04/2023

Problem Statement

- The dataset includes information about Scrabble games that are played between the bots and their opponents who are regular registered users.

- There are three bots:
  BetterBot (beginner),
  STEEBot (intermediate) and
  HastyBot (advanced).

- The goal is to predict the rating of the human opponents in the test set(test.csv).

PYTHON PACKAGES AND LIBRARIES

- Pandas

- Numpy

- Matplotlib

- Seaborn

- Sklearn

Algorithms

- Models used to predict the rating are given below:
- Random Forest Regressor
- Gradient Boosting Regressor
- Decision Tree Regressor
- Lgb
- Linear Regression
- Ridge
- Lasso
- MLP Regressor
  Among all the above algorithms "RANDOM FOREST REGRESSOR" is the best fit.

Random Forest Regressor

- Random forest is a type of supervised learning algorithm that uses ensemble methods (bagging) to solve both regression and classification problems.

- The algorithm operates by constructing a multitude of decision trees at training time and outputting the mean/mode of prediction of the individual trees.

Gradient Boosting Regressor

- Gradient boosting Regression calculates the difference between the current prediction and the known correct target value.
- This difference is called residual.
- After that Gradient boosting Regression trains a weak model that maps features to that residual.

Decision Tree Regressor

- Decision tree builds regression or classification models in the form of a tree structure.
- It breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed.
- The final result is a tree with decision nodes and leaf nodes.

Lgb

- LightGBM is a gradient boosting framework based on decision trees to increases the efficiency of the model and reduces memory usage.
- It uses two novel techniques:
  Gradient-based One Side Sampling and
  Exclusive Feature Bundling (EFB)
- These fulfills the limitations of histogram-based algorithm that is primarily used in all GBDT (Gradient Boosting Decision Tree) frameworks.

Linear Regressor

- Linear regression analysis is used to predict the value of a variable based on the value of another variable.
- The variable you want to predict is called the dependent variable.
- The variable you are using to predict the other variable's value is called the independent variable.

Ridge

- Ridge regression is the method used for the analysis of multicollinearity in multiple regression data.
- It is most suitable when a data set contains a higher number of predictor variables than the number of observations.
- The second-best scenario is when multicollinearity is experienced in a set.

Lasso

- Lasso regression is a regularization technique.
- It is used over regression methods for a more accurate prediction.
- This model uses shrinkage.

- Shrinkage is where data values are shrunk towards a central point as the mean.
- The lasso procedure encourages simple, sparse models (i.e. models with fewer parameters).

MLP Regressor

- MLPRegressor is an artificial neural network model that uses backpropagation to adjust the weights between neurons in order to improve prediction accuracy.

- MLPRegressor implements a Multi-Layer Perceptron (MLP) algorithm for training and testing data sets using backpropagation and stochastic gradient descent methods.

- It is an efficient method for solving regression problems as it can learn complex non-linear relationships between input and output variables.

Output

- Sample game id's and their ratings

	game_id	rating
0	2	2074.73
1	7	1926.90
2	11	1745.69
3	14	1929.79
4	27	2072.90

Comparison Table

- Comparisons between the different algorithms based on the values of R2 ,RMSE AND MAE.

	model	(R2)	(RMSE)	(MAE)
4	random_forest	0.918125	62.958010	37.382130
7	lgb	0.899922	69.612189	46.834566
3	decision_tree	0.831624	90.286223	47.919838
5	gradient_boosting	0.797119	99.119893	70.481361
0	linear	0.599163	139.323362	105.111821
1	ridge	0.599163	139.323363	105.111824
2	lasso	0.599116	139.331411	105.122168
6	neural_network	-0.535252	251.560803	164.593033

- Random Forest Regressor fits best among all the algorithms used.

Code Execution

- Notebook Link

- https://colab.research.google.com/drive/1T0loFGcfPtPKRQRpqjIX-SOJfTpLxsrUscrollTo=YypuSereoPA1

THANK YOU