ML-MAJOR-JULY-ML07B17

News Category Prediction

20th September 2020

Objective

The Project is of solving the Multiclass classification problem by using different Machine Learning Algorithms. Then use the ensemble learning approach to develop a classifier with maximum accuracy by stacking multiple classifiers.

Scope of The Project

The Project easily achieved accuracy up to 90% with a single classifier and by stacking multiple classifiers, Maximum approachable accuracy is 96% but having few drawbacks. As in Project, accuracy is maintained to 94%, is a success of their own.

Project Overview

The project dataset consists of labels and articles of News. There are 5 different types of News categories such as Technology, Business, Politics, Sports, Entertainment, and having an average of 250 words in each Article.

Project Detail :

As this is a Multiclass Classification Project:
Dependent Variable : Labels of News (Technology, Sports, Business...)
Independent Variable : News Article
Models/Classifiers : Naive Bayes, Random Forest, KNN, Decision Tree Classifier, Mlultilayer Perceptron
Ensemble Technique : Voting and Bagging

Four Question is asked in Problem statement and these are :

Ques 1: Can You Tell the Words which are Close in Meaning (can Say Synonyms but not exactly) used in Data provided?

Example: What a magnificent catch.
What a Super Catch.
What a tremendous Catch.
What a spectacular Catch.
What an impressive Catch.

(Remember All the Words Must Be used in Data)

input: Magnificient
output: Super, tremendous, Spectacular, impressive…

Ans 1: For this Answer tensorflow Projector is Used as it is quite simple to view different words at a time.
For this go to https://projector.tensorflow.org/
And on load section upload vecs.tsv on Vector Section (Step 1)

and on metadata section upload meta.tsv (step 2)
By this You will see the Projection of Words as Asked in Question

Ques 2: In This Type of Dataset (Multiclass Classification) Which Machine Learning Algorithms will Perform Good and Which Algo Will Perform Worse ?

Give a Well Define Table having Name of Algorithm ,
(Accuracy , Precision Score, Recall Score, F1 Score) on both Training And Validating Dataset.

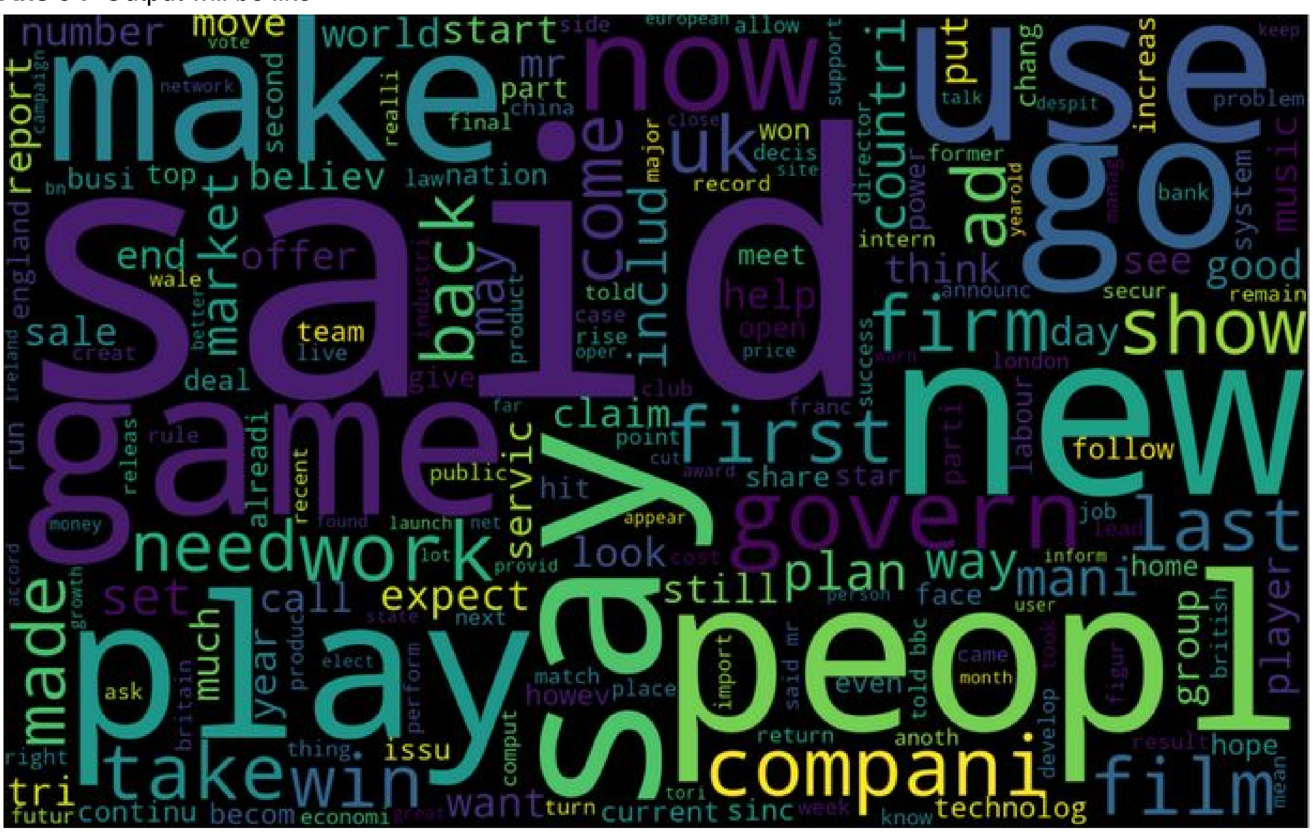
Then conclude the Algo according to above Observation

Ans 2: After Implementing all algorithms We will see this as answer.

	Title	Train_Accuracy	Train_Precision	Train_Recall	Train_F1-Score	Valid_Accuracy	Valid_Precision	Valid_Recall	Valid_F1-Score
0	Multi Layer Perceptron	87.490636	100.000000	100.000000	100.000000	96.179777	96.260196	96.068294	96.141763
1	Naive Bayes	99.475655	99.406046	99.475655	99.476719	96.404494	96.417421	96.304927	96.316876
2	KNeighborsClassifier	95.430712	95.240987	95.430712	95.448296	92.359551	93.036958	92.059787	92.281535
3	DecisionTreeClassifier	96.928839	96.734929	96.928839	96.936021	80.898876	81.083989	80.351368	80.433516
4	RandomForestClassifier	100.000000	100.000000	100.000000	100.000000	93.932584	94.048707	93.864609	93.892370

Ques 3 : What are the Most Common Words used in The Dataset Provided for Evaluation of Model.

Ans 3 : Output will be like



Ques 4 : For a given Sentence predict the Class among classes with Maximum Probability of belonging the Sentence to That Class.

Ans 4: For this Question We have used a part of Our Data set that is X_test As this part of Dataset is kept untouched for this Question Purpose specially .

```
Results are out
Accuracy of Ensamble Modelling is 93.48%
Recall of Ensamble Modelling is 93.02%
Precision Score of Ensamble Modelling is 94.56%
F1-Score of Ensamble Modelling is 93.53%
```

Project Stage :

Stage 1 : Data Cleaning
Stage 2 : Data Preprocessing (Stopwords ,Punctuations,Stemming,Lemmatization .. etc)
Stage 3 : Data Visualiation (Plots/Graphs, Wordcloud, pie, donuts, Heatmap)
Stage 4 : Feature Selection and feature Engineering
Stage 5 : Multi Layer Perceptron

Stage 6 : Model 1, Model 2, Model3, Model4, Model 5
Stage 7 : Accuracy Prediction and Visualization
Stage 7 : Ensemble Modeling (Voting Method)
Stage 8 : Combined Accuracy Prediction and Visualization
Stage 9 : Final Testing of Model
Stage 10 : Final Prediction on Test dataset and Visualisation

Reference :

1- https://www.deeplearning.ai/
2- https://machinelearningmastery.com/
3- https://www.towardsdatascience.com/
4- LMS verzeo

Note:This is not a Group Project. This is an Solo Project , Group Project will be Uploaded later.

Submitted By:

Gaurav Verma
gverma0549@gmail.com
+91 8182000549
Batch : **ML07B17**