FAKE NEWS DETECTION

The Link to github repository : - https://github.com/Devandra21/Detection_of_Fake_news_Revidly

SUBMITTED BY:-

DEVANDRA JAIN

3RD YEAR, B.TECH

ELECTRONICS AND COMMUNICATION ENGINEERING

NIT, TRICHY

PRE-PROCESSING

- The title column was removed from the dataset.
- Words were converted to tokens.
- Stopwords are removed by using sklearn library function.
- The dataset was divided to train and test set in 78:22 proportion with a random state of 43.
- In labels FAKE was replaced by -1 and REAL by +1.

MODELS

- Total 10 models were used.
- Count Vectorizer, Term frequency inverse document frequency(Tfidf) Vectorizer and Hashing Vectorizer these three vectorizer were used for each of the 10 models.
- The name of 10 models are Logistic Regression, Multinomial Naïve Bayes Classifier, Passive Aggressive Classifier, Stochastic Gradient Descent, Linear Support Vector Classifier, Decision Tree Classifier, Random Forest Classifier, Gradient Boosting Classifier, Ada Boost Classifier and Bagging Classifier.

ACCURACY OF DIFFERENT MODELS WITH TEXT CLASSIFIERS:-

Model Name	CountVectorizer	TfidfVectorizer	HashingVectorizer
Logistic Regression	0.920	0.918	0.917
Multinomial Naïve Bayes Classifier	0.893	0.891	0.839
Passive Aggressive Classifier	0.902	0.938	0.925
Stochastic Gradient Descent	0.912	0.938	0.928
Linear Support Vector Classifier	0.887	0.939	0.932
Decision Tree Classifier	0.813	0.823	0.812
Random Forest Classifier	0.905	0.915	0.892
Gradient Boosting Classifier	0.900	0.900	0.901
Ada Boost Classifier	0.879	0.874	0.876
Bagging Classifier	0.875	0.870	0.885

BEST PERFORMING

- Overall best performing was Linear Support Vector Classifier under text classification term frequency inverse document frequency(Tfidf).
- Under Count Vectorization Logistic Regression was best.
- Under hasing vectorization Linear Support Vector was best.

Worst performing

- Overall worst performing was Decision Tree Classifier, it was worst performing in all three types of text vectorizer.

MODELS THAT REVIDLY SHOULD USE

- For text vectorization Term frequency-Inverse document frequency(Tfidf) vectorization should be used.
- For classifiers the following models should be used:

  A) Linear Support Vector.

  B) Stochastic Gradient Descent.

  C) Passive Aggressive Classifier.

Thank You!!!