Detecting Real Or Fake Tweets

INDRAPRASTHA INSTITUTE of INFORMATION TECHNOLOGY DELHI

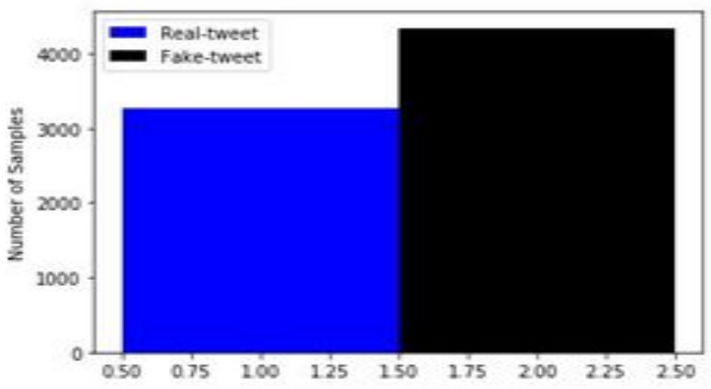
Era Sharma, MT19121
M-Tech CSE
IIIT DELHI

Megha Mathur, MT19104
M-Tech CSE
IIIT DELHI

Saumya Jain, MT19098
M-Tech CSE
IIIT DELHI

1 Problem Statement

The Problem is about predicting which Tweets are talking about real disasters and which are talking about fake disaster's. Task is to predict Tweets behavior.



2 Dataset Used

Dataset is taken from Kaggle named "Real or Not? NLP with Disaster Tweets. It comprises of 10,000 tweets which are classified into fake or real one. Each data point in test and training dataset has information such as tweets text, keyword and the location from where the text has been sent . Target column comprises of 2 values-0 or 1. 1 is for real disaster tweet and 0 for fake tweets.

3 Proposed Strategies

To tackle such a classification problem, we have extracted some features from given data such as count of #,@,!,,etc and implemented various models which are as follows:-

- **Tf-idf**:-. Vector for each tweet and for terms of each class is made. Cosine similarity between vector of that tweet with vector of both the classes is calculated and tweet is assigned to class having higher cosine similarity.
- **Naïve Bayes**:- Dictionary for unigrams and their term frequency which are appearing in both the classes is made. Score is calculated for each term using formula log((1+no. of times term t appears)/(total terms in class c + total terms in vocab)).
- **Decision Tree**:- We have preprocessed and normalized tweet text from training data.After preprocessing a document term tfidf matrix is created and model is trained using it.
- **KNN**:-The data is preprocessed and a dictionary is made, which contains all the words in the text of the data with its frequency. Vector is made using CountVectorization.

Similarity Measures

- **Resnik Similarity** :- It measures the similarity for each of possible synonyms of the words present in the text and maximum value of similarity is considered as final.
- **LSA**:It extracts hidden topics in the corpus and semantically similar words to that topic.we extracted disastrous and non-disastrous words and classified documents with majority of matched words among 2 topic
- **N-Gram**:-It is used to estimate the probability of a word given conditional probability of the previous words. We have used bigram (the value of n is 2).
- **Linear Regression:-** Linear Regression model is used to find the relationship between the dependent and independent variables. The distance from the line representing data sets and the data points are calculated.

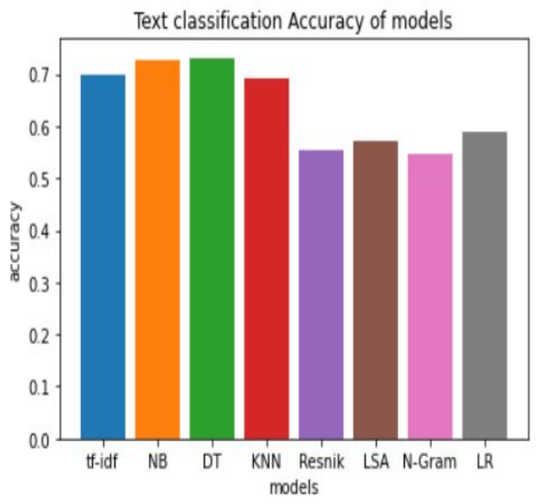
4 Literature Survey

Various work is done on this problem starting early with classifying tweets on the basis of user account information, tweet information, images, tweet text, replies to that tweet. Feature Extraction, preprocessing and classification are mainly used.A survey of text similarity approaches[5] helped in analyzing the semantic and lexical properties of text.

5 Experimental Results

It was observed and concluded that similarity measures like resnik, LSA and n-Gram are less reliable for the dataset provided,whereas Naive-Bayes and decision-tree is providing the best accuracy among all the models applied. High-dimensional model like linear regression is not giving accurate result due to less availability of good features.

S.NO	MODEL APPLIED	ACCURACY
1	TF-IDF MODEL	69.9%
2	NAIVE-BAYES	72.81%
3	DECISION-TREE	73.14%
4	K-NEAREST NEIGHBOUR	69.0%
5	RESNIK SIMILARITY	55.38%
6	LSA SIMILARITY	57.2%
7	N-GRAM LEXICAL SIMILARITY	54.56%
8	LINEAR-REGRESSION	58.76%



6 Conclusion

The problem statement was to predict the nature of tweets about disaster, whether it's real or fake. We applied different models including tfidf based cosine similarity model, Naive Bayes, Decision Tree, K-Nearest Neighbour, Resnik Similarity, LSA, N-gram and Linear regression. The accuracy of Decision Tree is the highest among all the 6 models applied. Therefore, we are able to predict the labels of test data with an accuracy of nearly 73%.

7 References

1. Castillo C, Mendoza M, Poblete B (2011) Information credibility on twitter. In: Proceedings of the 20th international conference on World wide web, ACM, pp 675-684
2. Gupta A, Lamba H, Kumaraguru P, Joshi A (2013) Faking sandy: characterizing and identifying fake images on twitter during hurricane sandy. In: Proceedings of the 22nd international conference on World Wide Web companion, pp 729-736
3. Vosoughi S, Mohsenvand M, Roy D (2017) Rumor gauge: Predicting the veracity of rumors on twitter. ACM Transactions on Knowledge Discovery from Data 11:1-36
4. Volkova S, Shaer K, Jang JY, Hodas N (2017) Separating facts from Linguistic models to classify suspicious and trusted news posts on twitter. In: Proceedings of the 55th Annual Meeting of the Association for Compu-ational Linguistics, vol 2, pp 647-653
5. Wael H. Gomaa,Aly A. Fahmy,"A Survey of Text Similarity Approaches",International Journal of Computer Applications (0975 – 8887),Volume 68– No.13, April 2013