FAKE NEWS DETECTION ON SOCIAL MEDIA

Presented To :
Prof. Jebez Christopher

Presented By :
Abhijeet Upadhyay (2019H1030121)
Ravi Bhatia (2019H1030508)
Shapath Mehta (2019H1030506)

Introduction

- The pace at which the misinformation goes viral is one of the major vulnerability of Information technology.
- As the scope of social media is large, hence a tool or medium is needed which will help people from getting negatively influenced with something.
- We have used decision tree and naive bayes classifier for classification of social media content into two predefined classes of fake or real.

Methodologies

Rtweet

- It is an R library which we have used for having an interaction with twitter API for fetching the latest tweets for the dataset.
- With this library, we can do the following things :
  - Search of upto 18,000 tweets containing the hashtag mentioned in the query string.
  - Plotting the frequency of tweets.
  - Search tweets with the geographical coordinates.
  - Search friends.
  - Get timelines, favorites and trends.

C50

- Library in R.
- Generally used for implementing decision tree and rule based models.
- We have used this for the implementation of decision trees as part of our algorithm.

hrbrmstr/nominatim

- This is the library we have used for the geolocation processing as per the preprocessing step.
- In the data, we are getting a location field with multiple variations like some tweets have location as its city name while others have location as its country name.
- With the help of this library, we have converted all the locations to a unified value.

Gmodels

- R Tool for model fitting
- We have used this for calculating the accuracy of our model using confusion matrix.

Parameter Selection

- Tweet Source
- User Activity
- Verified
- Country(Location)
- Followers
- Isfake

Algorithm

Step1. Data collection from twitter

Step 2. Preprocessing of data

- Selection of required parameters
- Handling Missing Entries
- Handling Ambiguous Entries
- Numeric to nominal conversion

Step3. Manual labeling of Fake data

Algorithm

Step4. Classification

- Dividing dataset in training and testing datasets
- Classification using Decision Tree classifier
- Classification using Naive Bayes Classifier

Step5. Performance measurement

- Confusion Matrix
- Accuracy

Performance measurements

- Confusion matrix

Literature survey

1. **Kai shu(2017)** paper

2. **Sejeong Kwon's** paper

3. **P W Liang's** paper

4. **Sheryl Mathias and Namratha Jagadee's** paper

Future Enhancements and conclusion

- Sentiment analysis
- Image and video analysis
- Quotes analysis
- Username analysis