



Detecting Online Nonsense

April 5

Team Code – WIP 1

Authored by: Anubhav Natani

Detecting Online Nonsense

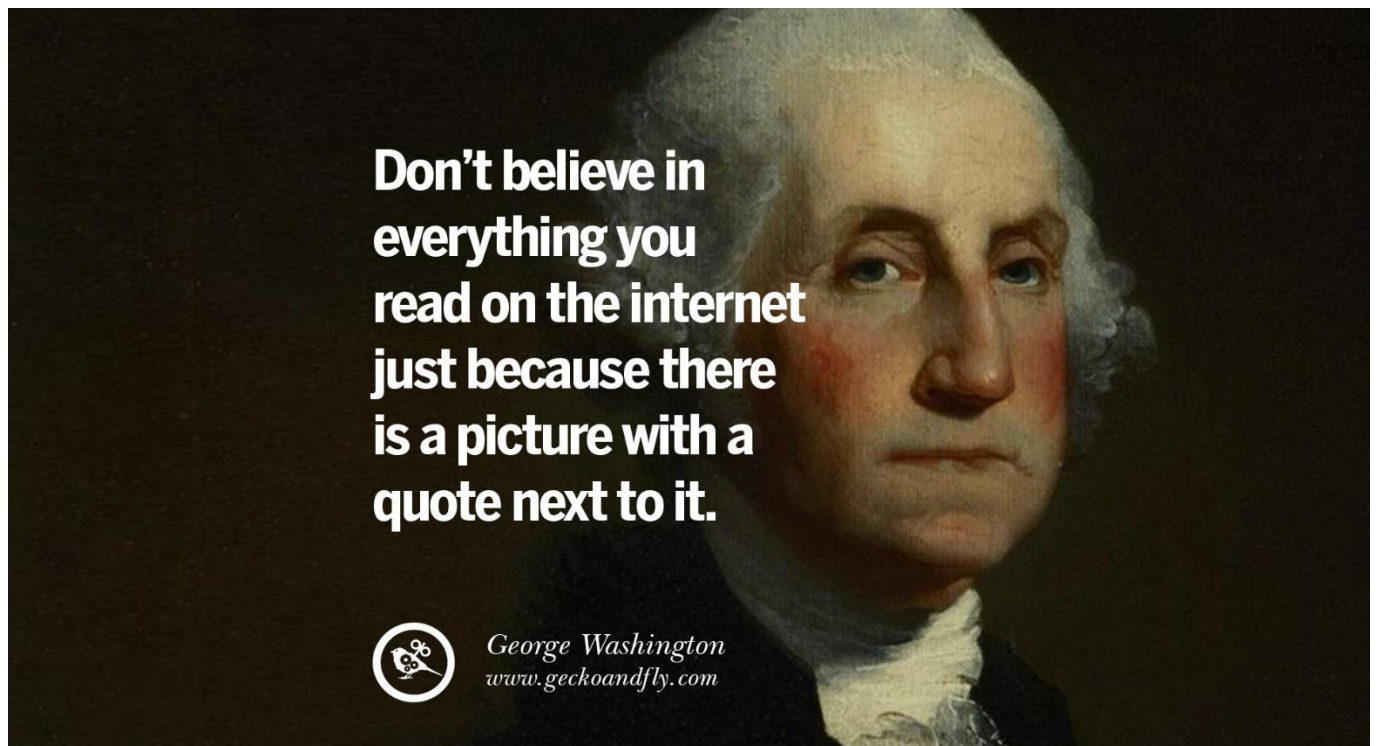
Description

The background in the first page is not just a fancy image it is the visualization generated from the fake news dataset. Shows how fake news affect our society.

Fake news is the one of the biggest problems of current internet driven world.
Fake news is there broadly for two reasons one is monetary and other in manipulation.

In this project we look at both the aspects of fake news and understand the difficulty in doing fake news prediction and How we can do it effectively using deep learning.

As the title suggest detecting online nonsense so fake news is not the only bad thing people tend to write bad comments too so in this project, we are also doing toxic comment detecting using deep learning.




Screenshots

In the project I have made an online forum for people to talk about their favorite **Hollywood celebrities** and chat. Keeping the fact that the **fake gossip is detected and toxic comment are also detected** and they do not make it into the online forum.

I name this forum as fake gossip.


I had also made a model for original news detection but not a portal for it right now only in **Jupiter notebooks**.




Fake Gossip

Gossip

Spread your gossip!!


 Add Gossip



Fake Gossip

Gossip

Ben Affleck Planning To Propose To Shauna Sexton After Rehab

 Add Gossip



Fake Gossip

Gossip

Spread your gossip!!



Add Gossip

fake



Fake Gossip

Gossip

'GAME OF THRONES' SEASON 8 PREMIERE IN NEW YORK



Add Gossip



Fake Gossip

Gossip

Spread your gossip!!



Add Gossip

real



Fake Or Real Gossip You Know?

'GAME OF THRONES' SEASON 8 PREMIERE IN NEW YORK


Checked by an Intelligent Model using tfjs(don't worry about title)



Your Comment




Add Comment




Fake Or Real Gossip You Know?

'GAME OF THRONES' SEASON 8 PREMIERE IN NEW YORK

Checked by an Intelligent Model using tfjs(don't worry about title)


 [Add Comment](#)





Fake Or Real Gossip You Know?

'GAME OF THRONES' SEASON 8 PREMIERE IN NEW YORK

Checked by an Intelligent Model using tfjs(don't worry about title)

 Your Comment


 [Add Comment](#)
toxicity




Fake Or Real Gossip You Know?

'GAME OF THRONES' SEASON 8 PREMIERE IN NEW YORK

Checked by an Intelligent Model using tfjs(don't worry about title)


 [Add Comment](#)





Fake Or Real Gossip You Know?

'GAME OF THRONES' SEASON 8 PREMIERE IN NEW YORK

Checked by an Intelligent Model using tfjs(don't worry about title)

 Anubhav Natani
Nice article by the author

 Your Comment

 [Add Comment](#)
Clean

The site is based on the data of Hollywood gossip so it accepts only Hollywood gossip not of Bollywood or any out of context news can be used to filter news website and I have also made a visualization showing which website show more fake news than real and other such statistics.

Installation Process

No installation process required as site is hosted online and model is made in TensorFlow JS so it uses in browser ml.

Site link--<https://fakegossip.herokuapp.com/>

Technologies

The main tech frameworks and languages used in project are---

- 1.**Keras** (For model making and understanding)
- 2.Pytorch (For prototyping)
- 3.**TensorFlow Js** (For deployment and in browser ml)
- 4.TensorFlow (For hub modules and some custom functions)
- 5.Pandas (Data analysis)
- 6.Nltk (For natural language analysis)
- 7.Sklearn (For Machine learning algorithms)
- 8.Numpy, Matplotlib.... (For data Analysis)
- 9.**React** (For website)
- 10.**JavaScript, Python** (Languages)

Detailed Description

As mentioned in the earlier report the first part of the project is to collect the data by collection, I mean mix of data from the dataset and websites.

I scraped some data of the internet and other data was used from the datasets.

A research paper came last week only, had a very much diverse data for the fake news from which I had trained my fake gossip model.

I cleaned the dataset and analyzed it and then I used all types of machine learning models on the dataset and tried to get high accuracy but there is some sought of trade off between accuracy and real performance until I read the research paper which google published last year and made my model from it.

It used whole sentences instead of word for embeddings different from the other bag of words methods and word2vec or other types of embeddings.

The name of model I used embedding is Universal Sentence Encoder it not only gives the accuracy but gives good results in real world situations.

Then I trained a machine learning model using the universal sentence encoder and its embeddings.

After I was satisfied with accuracy, I made a web app using react and I had to rewrite some parts of my model in java script to make use of tfjs just little code and I also converted my keras model to tfjs and then used my tfjs model in react web app and then I deployed my web app to Heroku.

I made a chat forum to make UI look real and show real world use case of model.

And for toxic comment, I did the same thing but I took dataset from Kaggle which is very much cleaned just I did the same thing which I did for the fake news and I got good results again. I also tried some complex models like attention and others.

List of Model and embeddings tested

Models Tested

- 1.Logistic Regression
- 2.SVM
- 3.Random Forest
- 4.Light GBM
- 5.XGboost
- 6.MLP
- 7.LSTM
- 8.GRU
- 9.Simple RNN
- 10.Conv1D
- 11.Attention (Combination with Conv1d and bidirectional LSTM)

Embedding and Word Vectorizers Used

- 1.TFIDF Vectorizer
- 2.Count Vectorizer
- 3.Word2Vec
- 4.Glove Embeddings
- 5.Universal Sentence Encoder