Early Fake News Detection

Abdelrahman Tolba, Radwa Mostafa, Mariam Khaled, John Gerges

November 6, 2019

Abstract

In our modern society the Internet has been taking a huge part in people's life and it's so easy to publish any rumors that can mislead people away from the truth as it's a fast way to get news and doesn't consume time. The spread of fake news has different aspects and can affect lots of different entities; politically and industrially. The term "Fake News" became mostly common in 2016 the fame of disinformation within American political discourse was the subject of attention, particularly following the election of President Trump. This term was particularly used to describe incorrect and misleading articles published mostly for the purpose of making money or to change people's point of view. In This project we aim to detect the published fake news using Machine Learning. We will combine two of the most efficient classifiers that will lead us to generate more effective results by adding browser extension which works automatically while surfing the internet by this people will be updated to every real event.

1 Introduction

1.1 Background

Nowadays Social media is the most common place for fake news; such as Facebook and Twitter. As it has no restrictions on people's posts, as people can edit their posts anytime they want with no limit to it opposite to Newspaper which has more restrictions.

However, most people now use Social Media as their source despite its misleading of not showing whether the news are true or not. when a certain event happens fake news starts to increase to mislead people.

Fake news detection is one of the most common and critical types of deception, Fake news is a type of yellow journalism or propaganda that consists of intended disinformation spread via news, Media or Social Media.
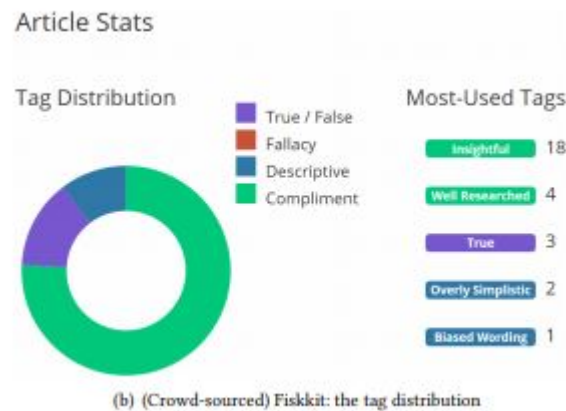
Studies have shown that human ability to differentiate between fake news and real one are only 54 percent. Machine learning is the best way to enhance the accuracy of this detection.

1.2   Motivation

1.2.1   Market Motivation

Our project is motivated by previous work and domain.

1. The fake news can be easily created and published as it's cheaper than real news

2. It play a great role as 67% of the Americans get their new from social media

3. The effect of fake news on the economic and political benefits is huge and as result it encourage fake news posting rate

4. Teenagers as reported by the NBC earned at least $60,000 due to spreading fake news and their expected wage is $4,800 which motivates them to increase spreading.

5. Donald Trump's file was mostly false which motivated many developers to check to the reality of his file and as stated this was the result:

(b) (Crowd-sourced) Fiskkit: the tag distribution

1.3   Problem Definitions

News play a great role in industry , as they can control how people to see specific product or company. fake new can cause a whole company to lose it's shares in the market or cause a great loss because fake new can affect how people see's or feel about this company and act upon it.

also fake news gives misleading information about the country which will affect our tourism badly and make them think it's not safe and maybe effect the children way of thinking about their country and what it offers for them.

detection fake news will prevent this incidents to happen and will help people figure out what they can trust and know the credible sources to know the news

2 Project Description

Fake news has became an issue for the public and government as it takes advantage of multimedia content to mislead readers and get published, which can lead to negative effects or even manipulation of the citizens.
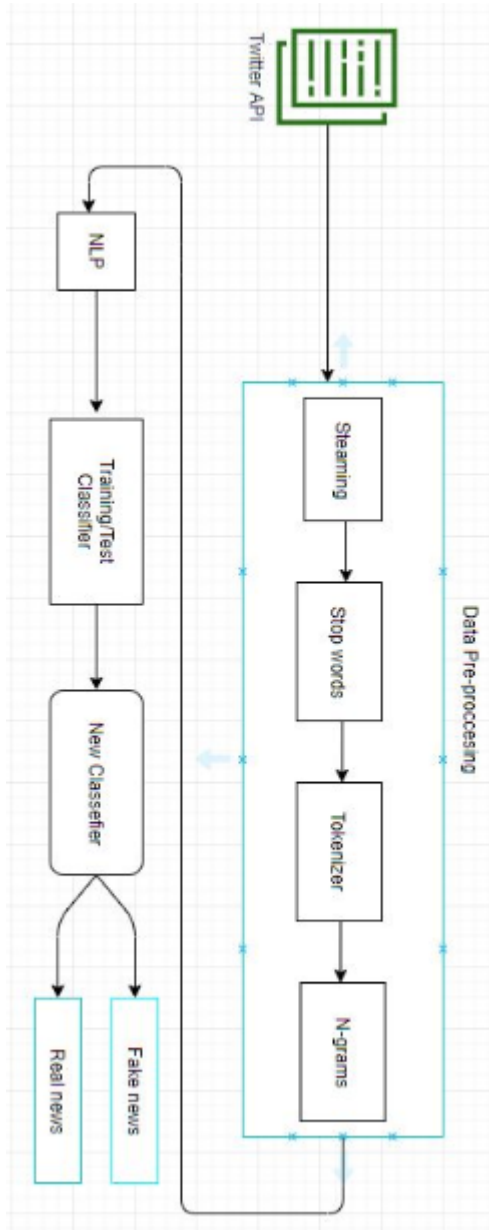
2.1 Objective

In order to minimize the disadvantages of spreading fake news all over the social media (Twitter), Machine Learning is recently used in order to detect the fake news and it showed good results. However, our project aims to increase the percentage of accuracy to achieve the best results by using hybrid classifier and NLP that combines two different classifier together to achieve the optimum accuracy results.

2.2 Scope

In our project we aim to improve the accuracy of detecting the fake news and trying to reach the maximum results that can be reached. To do that we will use a hybrid classifier that combine two different classifiers together, also we want to make this project useful from aspects by providing an additional feature which is allowing the user to enter keywords for the news they want to track for a certain duration.

2.3 Project Overview

2.3.1 Data set

In our project we have two types of the dataset that we can use:

1. Article news :this type of data-set consist of whole separated articles, as

the fake article in a folder and the legit articles in separated folder.

009fake - Notepad

File Edit Format View Help

Miley Cyrus Wedding Rumors: Will Miley Force Liam Hemsworth To Sign A Prenu
 to know if either party plans to protect their assets before tying the kno
 of that, there are reports that she really wants Liam to sign a prenup bef
caused stress in her relationship with Liam, and that the couple is current
 $160 million while Liam boasts just $16 million himself. Life & Style clai
toward marriage. Apparently, it doesn't matter to Miley if everything about
she knows things can go south very quickly, just like they did in 2013. Mil
 stand the test of time."They've had heated arguments over this," a source
that if they are truly in love, it's not needed. He's assured Miley he isn'
ensures that her money is protected with a prenup in the event she does mar
 happy. It was Miley's wild ways that pushed Liam away the first time aroun
longer and Miley isn't acting out nearly as much as she was in past years.O
that he loves. While Miley has toned it down some, Liam is definitely meeti
.After reports that Miley moved in with Liam, it looks like she also bought
 4-bedroom home in Malibu. "The plan is for Liam to stay there while they r
 told the gossip site. It's unclear why Miley bought the home since it was
 homes. Could this be their way to get to know each other again without bei
reconciliation. Then again, they aren't hiding their love either. Ever sinc
seen spending time together. There have even been reports that mutual frien
 Miley. Not to mention that Cyrus was seen sporting her old engagement ring
Cyrus and Liam Hemsworth? Fans hope not and many are excited that Miley has
doesn't tame the entertainer so much that it hurts her career, if they can

009legit - Notepad

File Edit Format View Help

Miley And Liam Fighting? False Rumors Swirl That They're In A Feud Over A S
and Hemsworth had a huge fight over their supposed prenup while they were m
planning their wedding in the Himalayan Mountains this summer. But there ha
OR ON? UPDATE ...Just last month, Cyrus and Hemsworth faced rumors they wer
whether to live in L.A. or Australia. And sources said they might not get m
 (Photo: XactpiX / Splash News).But those rumors, like all the other split,
 their relationship is doing better than ever. The cute couple even recentl
 married soon?Sources say that Cyrus and Hemsworth don't want to rush thing
pressure to get married soon. Instead, they're taking time to build a solid
THE LATEST ..."They want to figure out their life together before they get
They just realized that there is no rush with the wedding and that it's bet
Let us know your take in the comments section below!

2. statement news :This type of data-set consist of the news statement and the other column it's classification

	A
1	Statement
2	Says the Annies List political group supports third-trimester abortions on demand.
3	When did the decline of coal start? It started when natural gas took off that started to begin in
4	Hillary Clinton agrees with John McCain "by voting to give George Bush the benefit of the dou
5	Health care reform legislation is likely to mandate free sex change surgeries.
6	The economic turnaround started at the end of my term.
7	The Chicago Bears have had more starting quarterbacks in the last 10 years than the total num
8	Jim Dunnam has not lived in the district he represents for years now.
9	I'm the only person on this stage who has worked actively just last year passing, along with Ru
10	However, it took $19.5 million in Oregon Lottery funds for the Port of Newport to eventually l
11	Says GOP primary opponents Glenn Grothman and Joe Leibham cast a compromise vote that c
12	For the first time in history, the share of the national popular vote margin is smaller than the L
13	Since 2000, nearly 12 million Americans have slipped out of the middle class and into poverty.
14	When Mitt Romney was governor of Massachusetts, we didnt just slow the rate of growth of o
15	The economy bled $24 billion due to the government shutdown.
16	Most of the (Affordable Care Act) has already in some sense been waived or otherwise suspe
17	In this last election in November, ... 63 percent of the American people chose not to vote, ... 8
18	McCain opposed a requirement that the government buy American-made motorcycles. And h
19	U.S. Rep. Ron Kind, D-Wis., and his fellow Democrats went on a spending spree and now their
20	Water rates in Manila, Philippines, were raised up to 845 percent when a subsidiary of the Wo
21	Almost 100,000 people left Puerto Rico last year.
22	Women and men both are making less when you adjust for inflation than when John Kitzhabe

3   Similar System Information

[1] Fake News Detection Using A Deep Neural Network The social media is a double edges weapon; although it's an easy way to be accessed and doesn't consume alot of time it can also be manipulated by different entities. so in this paper it identifies fake news using various models and classifiers; They proposed a Dataset for Kaggle which combine various technologies as machine learning, natural language processing and deep learning. Using these Dataset and techniques they investigate the analyses to identify linguistic properties that are present in the content.

[2] Fake News Detection As Facebook is considered one of the most popular platforms for news and information this paper applied it's methodology on it by using Naive Bayes classifier which classify whether the post is real or fake. To increase the accuracy a huge Dataset of posts must me used to obtain better results.

[3] Sentiment Analysis Based on Support Vector Machine and Big Data: In this paper Lukas Povoda, Radim Burget and Malay Kishore Dutta dicussed that Support Vector Machine[SVM] is mainly used in text analysis as it can be trained by huge amount of data and can deal with complex combined words as well as dealing with different languages. While working with big data [SVM] improved the accuracy while working with big data with percentage of 11 and the best accuracy obtained was 95.31 percent. For recognition on the rate of positivity and negativity.

[4] In this paper Dr Eslam Amer have talked in the first about the social media and how it's a good place for the fake news, he also mentioned that Facebook have a percentage of 50% of fake news traffic and he mentioned that 62% of US adults get there news from social media. He also discussed the impact of the fake news and how useful this project will be and how it will help people to over come this problem. The paper also mentioned that NLP is used to reach the optimum results. The used algorithms is RNN and LSTM and he used the Liar dataset

[5] In This paper Gilda, Shlok used two different datasets which are signal media which gets news from local news, blogs and articles as well as outgoing list of fake and real news from OpenSource.co. They used SVM, Bounded decision tree, Random forests and Gradient Boosting as classifiers to increase accuracy and to get the diference between them and SVM happened to get the highest accuracy level with 77.2 percent.

3.0.1 Comparison

Paper name	classifier	Dataset	Accuracy	Reference
Fake News Detection Using A Deep Neural Net-work.	CNN Naive Bayes model Decision Tree	various data-sets	91% 89% 73%	Saliyar, Rohit Kumar. "Fake News Detection Using A Deep Neural Net-work." 2018 4th International Conference on Computing Communication and Automation (ICCCA), 2018. https://doi.org/10.1109/ccaa.2018.8777343
Sentimet Analysis Based on Support Vector Machine and Big Data.	SVM	manually collected From websites	95.31%	Povoda, Lukas, Radim Burget, and Malay Kishore Dutta. "Sentiment Analysis Based on Support Vector Machine and Big Data." 2016 39th International Conference on Telecommunications and Signal Processing (TSP), 2016. https://doi.org/10.1109/tsp.2016.7760939
Deep Learning Algorithms for Detecting Fake News in Online Text	RNN models (Vanilla RNN, GRU) and LSTMs	LAIR dataset from POLITIFACT.COM	83.5% 81.4%	Shery Girgis, Eslam Amer, and Mahmoud Gadallah. "Deep Learning Algorithms for Detecting Fake News in Online Text." 2018 13th International Conference on Computer Engineering and Systems (ICCES), 2018. https://doi.org/10.1109/icces.2018.8639198.
Evaluating Machine Learning Algorithms for Fake News Detection	SVM, Bounded Decision tree, Random Forests and Gradient Boosting	Signal Media, OpenSource.com	73.6% 67.6% 64.8% 65.7%	Gilda, Shlok. "Evaluating Machine Learning Algorithms for Fake News Detection." 2017 IEEE 15th Student Conference on Research and Development (SCOReD), 2017. https://doi.org/10.1109/scored.2017.8305411.

4 Project Management and Deliverables

4.1 Tasks and Time Plan

Task	FROM	To
Announce proposal for students	1 July 2019	15th of July 2019
Register to projects	-	till end of July
Lecture writing paper skills	Second week of September 2019	
Proposal Evaluation	First week of October 2019	
SRS Evaluation	Second week of December 2019	
SDD Evaluation	Third week of February 2020	
Prototype Evaluation	Three days after midterm exams	
Technical Evaluation	1st week of May 2019	
Final Thesis	Last 10 days in June 2020	
Ceremony	24th of June 2020	

5 References

[1] Kaliyar, Rohit Kumar. "Fake News Detection Using A Deep Neural Network." 2018 4th International Conference on Computing Communication and Automation (ICCCA), 2018. https://doi.org/10.1109/ccaa.2018.8777343.

[2]Jain, Akshay, and Amey Kasbe. "Fake News Detection." 2018 IEEE Inter-

national Students Conference on Electrical, Electronics and Computer Science (SCEECS), 2018. https://doi.org/10.1109/sceecs.2018.8546944.

[3]Povoda, Lukas, Radim Burget, and Malay Kishore Dutta. "Sentiment Analysis Based on Support Vector Machine and Big Data." 2016 39th International Conference on Telecommunications and Signal Processing (TSP), 2016. https://doi.org/10.1109/tsp.2016.7760939.

[4]Girgis, Sherry, Eslam Amer, and Mahmoud Gadallah. "Deep Learning Algorithms for Detecting Fake News in Online Text." 2018 13th International Conference on Computer Engineering and Systems (ICCES), 2018. https://doi.org/10.1109/icces.2018.8639198.