Fake News Detection

TEAM MEMBERS

- Andrew Paul Gardner (E20CSE118).
- Aryan Singh Chaudhary (E20CSE120).
- Harshit Sharma (E20CSE136).
- Sahil Saurabh Jain (E20CSE144).

Abstract

Misleading content presented under the guise of legitimate journalism is a global information accuracy and completeness problem that influences opinion-forming, decision-making, and voting patterns. The fake news articles that initially circulated across social media platforms share key linguistic traits, including unsubstantiated hyperbole and excessive use of unsourced and quoted content. Most of the so-called “fake news” first spreads through social media channels such as Facebook and Twitter and then flows into mainstream media platforms such as traditional TV and radio news. In this project, results of a fake news identification study documenting the performance of fake news classifiers are presented.

INTRODUCTION

Deliberately misleading content under the guise of legitimate journalism (or “fake news” as it is commonly known) is the use of global information to influence opinion-forming, decision-making, and voting patterns. It's a matter of accuracy and completeness. Most fake news is first disseminated through social media channels such as Facebook and Twitter and then flows into mainstream media platforms such as traditional TV and radio news. Initially disseminated via social media platforms, fake news shares key linguistic traits, such as unsubstantiated exaggeration and excessive use of unsourced and quoted content. In this white paper, we present and discuss the results of a fake news identification study documenting the performance of fake news classifiers.

EXISTING SYSTEM

Enforce user privacy preferences. Securing data. Users aren't skilled enough to translate privacy requirements into privacy preferences. User might have various difficulties in setting up privacy preferences.

We rely on Manual-Fact checking in the current system which can be very much inaccurate and may cause major issues as fake news can affect the system in a bad

Fact checking websites, like Politifact and various other websites, rely fully on human judgement to decide whether a certain news is truthful or not. Human judgment can be clouded by emotions as compared to a machine which does not have any bias.

DISADVANTAGES

Digitally produced data is scattered Information is managed by many providers, social media etc. Users are losing control of their data. Data cannot be exploited as different sources have different views.

There are many cons of fake news, which can escalate further at a wider level if not tackled on time. From a personal level to global, these Disadvantages of Fake News can affect social or economic harmony.

1.Change in Public Opinion

Incidents such as the exposure of fraud and the discovery of flaws in the opposition system are common examples of changes in public opinion.

2.Defamation

Everyone has the right to express themselves. This right is often abused to defame influential and influential figures, such as business leaders and celebrities. Fake news is the best way to spread misinformation across borders.

3.Fake News Cost lives

As the coronavirus pandemic spreads, so too have deaths linked to fake coronavirus news skyrocketed

4.False Perception

When an influential person says it, people tend to believe it. Getting the wrong perception about someone is a big drawback of Fake His News. Every time a strong

personality with millions of fans misleads people about an incident, they will blindly believe it.

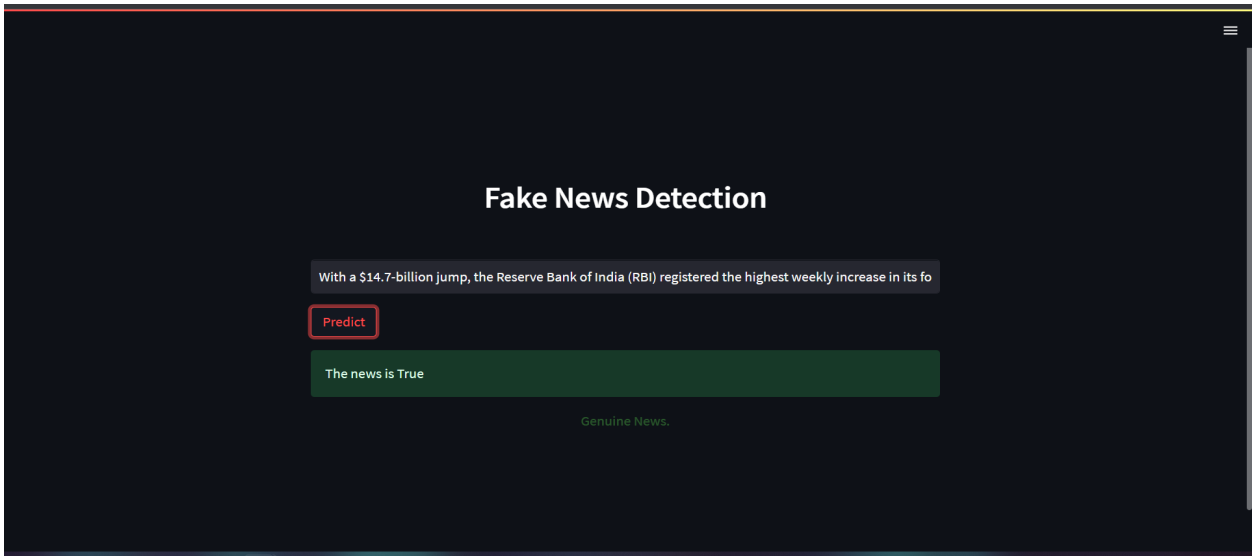
CONCLUSION

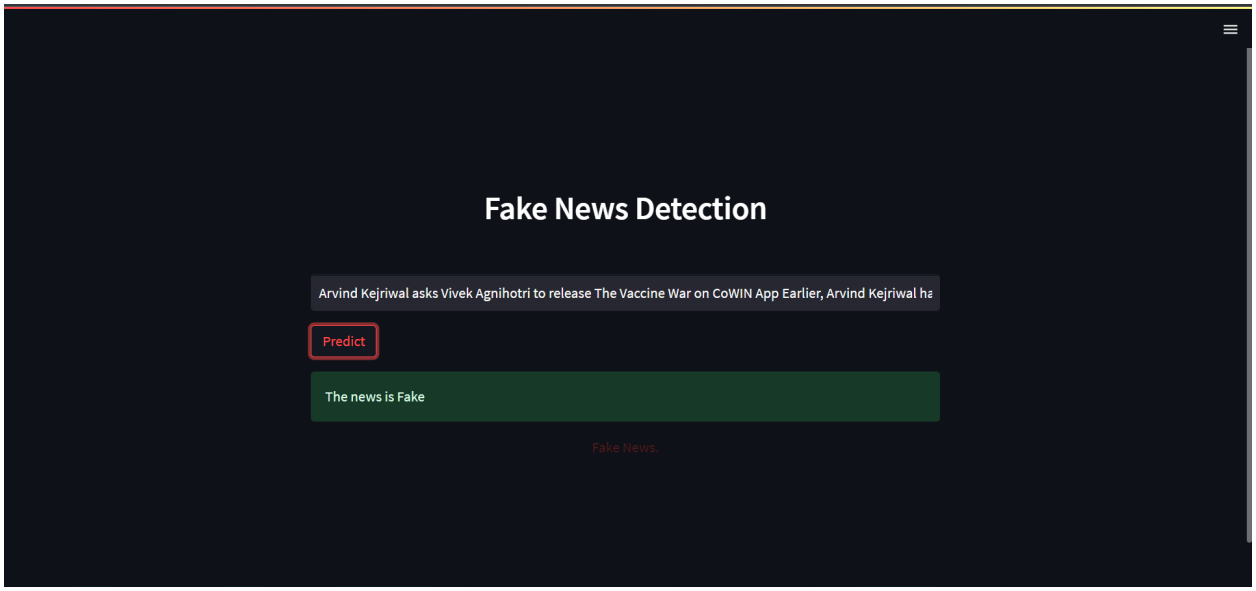
This project presented the results of research that created a limited fake news detection system. The work presented here is novel in the field in that it presents the results of a comprehensive research project that began with qualitative observations and led to working quantitative models.

The work presented in this project is promising because it demonstrates a relatively effective level of machine learning classification of large fake news documents with only one extractor. Finally, additional research and work is underway to identify and create additional taxonomy grammars for fake news, which should result in more sophisticated taxonomy schemes for both fake news and direct citations.

Approach

So we used two datasets to for this model, one of which was the Fake. csv and other being True.csv files containing true and fake news. We subsequently had to clean our dataset so that it could be properly used on our model. We removed the city names and reuters fields from our dataset since they were useless and caused overfitting. We eliminated two additional fields, date and subject, because they were useless. Since, we had two datasets that needed to be combined, we needed to label them. So we made a new column and appended 1 for true news and 0 for fake news.
Then, for the processing part of this project, we created a pipeline that used vectorization, TF-IDF, and the SVC model to automatically enter the output of the Count Vectorizer, among other things, into the TF-IDF.
TFIDF was utilised to tell us if the news we were submitting was real or false. The Count Vectorizer was used to convert a text into a matrix. We utilised a bag of words to find the word(s) that were commonly used in false news. Finally, because we employed supervised machine learning, we used the SVC model to categorise the news as phoney or true.


Fake News Detection
With a $14.7-billion jump, the Reserve Bank of India (RBI) registered the highest weekly increase in its fo
Predict
The news is True
Genuine News.


Fake News Detection
Arvind Kejriwal asks Vivek Agnihotri to release The Vaccine War on CoWIN App Earlier, Arvind Kejriwal ha
Predict
The news is Fake
Fake News.