**PROJECT REPORT**

**on**

**Fake News Detection Using Machine Learning**

**(CSE VI Semester project )**

**2021-2022**



**Submitted to: Submitted by:**

Mr. UMANNG GARG Mr. HIMANSHU CHAND

(CC-CSE-A-VI-Sem) Roll. No.: 1918375

**Guided by:** CSE-A-VI-Sem

**DEPARTMENT OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY**

**GRAPHIC ERA HILL UNVERSITY, DEHRADUN**

**Fake News Detection Using Machine Learning**

ABSTRACT

Consuming news from social media is becoming increasingly popular nowadays. Social media brings benefits to users due to the inherent nature of fast dissemination, cheap cost, and easy access. However, the quality of news is considered lower than traditional news outlets, resulting in large amounts of fake news. Detecting fake news becomes very important and is attracting increasing attention due to the detrimental effects on individuals and the society. The performance of detecting fake news only from content is generally not satisfactory, and it is suggested to incorporate user social engagements as auxiliary information to improve fake news detection. Thus it necessitates an in- depth understanding of the correlation between user profiles on social media and fake news. I perform a comparative analysis over explicit and implicit profile features between these user groups, which reveals their potential to differentiate fake news.

**INTRODUCTION**

With the increasing popularity of social media, more and more people consume news from social media instead of traditional news media. Fake news is now viewed as one of the greatest threats to democracy, and freedom of expression. It has weakened public trust in government. The reach of fake news will be best highlighted all the time. The extensive spread of fake news has the potential for extremely negative impacts on individual and society. A type of yellow journalism, fake news encapsulates pieces of news that may be hoaxes and is generally spread through social media and other online media. This is often done to further or impose certain ideas and is often achieved with political agendas. Such news items may contain false and/or exaggerated claims, and may end up being verbalized by algorithms, and users may end up in a filter bubble.

**MOTIVATION**

In today’s modern world we are so much connected to social media and this often helps us in gathering hot topics and current affairs on regular basis.

But sometimes the scenario is bit different when we encounter two contradicting news on different platforms. At this time we got confused in believing on particular side as we don’t know which news is real and which one is fake. To overcome this situation one needs a platform where he/she can search for the news for its correctness.

So, the idea of creating such platform encouraged me to develop this project and help the society as well with this kind of project.

**WORKING AND METHODOLOGY**

Primarily, we take input in the form of a dataset through the user and then the dataset is handled using the model build based on the count vectorizer or a tf-idf matrix (word tallies how often they are used in other articles in our dataset).[1]

Now the next step is to extract the most optimal features for count vectorizer or tfidf-vectorizer, this is done by using a n-number of the most used words mainly removing the stop words which are common words such as “the”, “when”, and “there” and only using those words that appear at least a given number of times in a given text dataset.

Since this problem is a kind of text classification, Implementing **passive aggressive algorithm** will be best as this is standard for text-based processing. The actual goal is in developing a model which was the text transformation (count vectorizer vs tfidf vectorizer) and choosing which type of text to use.

Finally, we are obtaining the accuracy and printing the true and false positives and negatives using confusion matrix. Confusion matrix is nothing but a table used to describe the performance of classification model or classifier on set of data set. We obtain the accuracy based on how well our classifiers are working and the data set is bet fit in it. Accuracy is calculated based on true and false positives and negatives.

The final accuracy achieved by the model was **94.71%**.

After successfully building the ML model it was then exported using pickle python library along with as ‘finalized\_model.pkl’ and ‘vectorizer.pkl’ respectively.

Later on this model is then used with flask to create a web application to handle the new fake news search and provide a beautiful user interface with the help of HTML , CSS and Bootstrap.

**SCREENSHOTS OF THE PROJECT**

Text

Description automatically generated

Text

Description automatically generatedflask file:-

Background pattern

Description automatically generatedCollab notebook :-

Text

Description automatically generatedHTML File:-

# REFERENCES

1.-This project is inspired by Adam Geitgey

[Data science playlist by codejay](https://youtu.be/CUkggjNNoWs)

2.GeeksForGeeks

[Passive Aggressive Classifiers - GeeksforGeeks](https://www.geeksforgeeks.org/passive-aggressive-classifiers/)

3.Udemy Machine learning courses

[Machine Learning A-Z (Python & R in Data Science Course) | Udemy](https://www.udemy.com/course/machinelearning/)