

NAME OF THE PROJECT

# Fake news project

Submitted by:

Neha kumari

**ACKNOWLEDGMENT**

This includes mentioning of all the references, research papers, data sources, professionals and other resources that helped you and guided you in completion of the project.

I took help from google, Kaggle, Github and my previous project

**INTRODUCTION**

* Business Problem Framing

Fake news has become one of the biggest problems of our age. It has serious impact on our online as well as offline discourse. One can even go as far as saying that, to date, fake news poses a clear and present danger to western democracy and stability of the society.

* Motivation for the Problem Undertaken

In the era of news in our lives, it is the people’s responsibility to not to share any misleading information as there are many sources available now-a-days. The fraud news such as spam messages, funding news or any false information to be fall out or reach to the people we consider it as a serious issue although it is extremely complicated to find out which is fraud and which is not a fraud profile or users in social media, they replicate the information as the original one. As the technology evolved and the machine intelligence has come into existence everyone tends to use available sources for creating and dissemination of fraud news. People who are illiterate might be new to digital media as they are inexperienced, so they are the ones who believe that fraud news easily and makes it practical in their lives. To a minimum, we have deviled а simple web аррliсаtiоn which statistically detects false information, and also real news.

**Analytical Problem Framing**

* Mathematical/ Analytical Modeling of the Problem

# Naive Bayes

# Logistic Regression

# Decision tree

# Random Forest

* Data Sources and their formats

Data source from google, data in csv formate

* Data Preprocessing Done

we are cleaning our text by steaming, lemmatization, remove stopwords, remove special symbols and numbers, etc. After cleaning the data we have to feed this text data into a vectorizer which will convert this text data into numerical features.

* Hardware and Software Requirements and Tools Used

Hardware -laptop, **keyboard, mouse**

**Software – Jupiter Notbook**

**Tools -** Scikit-learn,scipy.stats

**library –** using python library – numpy , pandas, matplotlib,sns,

**machine algorithm –** Naïve bayes,logistic regression, decision tree classifier, random forest classifier

**Model/s Development and Evaluation**

* Identification of possible problem-solving approaches (methods)

Describe the approaches you followed, both statistical and analytical, for solving of this problem.

* Testing of Identified Approaches (Algorithms)

Listing down all the algorithms used for the training and testing.

* Run and Evaluate selected models

Describe all the algorithms used along with the snapshot of their code and what were the results observed over different evaluation metrics.

* Key Metrics for success in solving problem under consideration

What were the key metrics used along with justification for using it? You may also include statistical metrics used if any.

* Visualizations

Mention all the plots made along with their pictures and what were the inferences and observations obtained from those. Describe them in detail.

If different platforms were used, mention that as well.

* Interpretation of the Results

Give a summary of what results were interpreted from the visualizations, preprocessing and modelling.

**CONCLUSION**

* Key Findings and Conclusions of the Study

The model that we developed is not particularly belongs to one media and in our dataset all the data consists of news reports from various digital media that means our model understanding could be applied to any digital media to know what is fraud and what is fraud not. Our model’s future work is to develop a dynamic model so that our users can download our app and can easily detect any news and any fake URL’s and we are also thinking to develop a model so that it can detect any fake profiles present in any media such as Facebook, Instagram, Stack Overflow and also any fraud reviews for duplicate products. Not only these but there are also many unsettled controversies and topics about celebrity’s fraud news and the news articles about the world is always a concern to each and everyone. By considering above matters professionals have to share out with them. For example, recently many frauds news articles about covid vaccine created an immense effect on people. My opinion people must research about news if any new news is in front of us. Now-a-days many news are being forwarded to WhatsApp application about funding and it is especially important to point out the major sources of that news and have knowledge about them and share to people so that everyone can able to understand about a particular news. I think everything is there in people’s hand, if we

* Learning Outcomes of the Study in respect of Data Science

we are cleaning our text by remove stopwords, remove special symbols and numbers, etc.

we have used 5 machine larning algorithm

best algorithm is decision tree classifer has given the best score around 99%

* Limitations of this work and Scope for Future Work

What are the limitations of this solution provided, the future scope? What all steps/techniques can be followed to further extend this study and improve the results.