***K. J. Somaiya College of Engineering, Mumbai-77***

***(Autonomous College Affiliated to University of Mumbai)***

*Source adaptive disinformation detection*

***Team member*** *- Heet sakaria1 Vedang Parasnis2 Hritik Jaiswal3*

***Mentor*** - Irfan Siddavatam, Head of I.T. Department

*1Department of Information technology, K J Somaiya engineering college, vidyavihar, Somaiya university*



**Keywords**: Quality Requirements, Machine Learning, Requirements Classification, Natural Language Processing, Analytics, Fake news, Classification, Topic modeling, Text analytics.

## **1. INTRODUCTION**

Fake news is playing an increasingly dominant role in spreading misinformation by influencing people’s perceptions or knowledge to distort their awareness and decision-making. The growth of social media and online forums has spurred the spread of fake news causing it to easily blend with truthful information. This study provides a novel text analytics–driven approach to fake news detection for reducing the risks posed by fake news consumption.

* 1. **Purpose**

The purpose the document to build the intelligent system which will detect the disinformation present in the news

* 1. **Overview**

The speed with which the Fake News spread today has encouraged work in various areas to

Minimize the damage and the public insecurity caused by their proliferation.

The purpose of the work is to come up with a solution that can be utilized by users to detect

And filter out sites containing false and misleading information. We will be using simple and carefully selected features of the title and post to accurately identify fake posts/content/headline.

* 1. **Objective**
* Gathering and comparing news headlines from the genuine resources via crawlers and web scrapers.
* Classifying the information based on the coherence, consistency and acceptability of

headlines.

* Measure for classification will be similarity, orientation, context and semantics.
* Evaluate the credibility of the headline and identify the similarity among headlines.
* Manipulation of disinformation detected by the cognitive behavioral process which is based on sentiment analysis via Natural language processing. Further identifying the most accurate information among them. Finally manipulate the exaggeration of News on the basis of the explicit information predicted by the algorithms used.

**1. 4: Scope**

* 1. **System overview**

The product mainly consists of google chrome extension which will extract the necessary features (headline, date, summary/content) from the site and after that we will apply suitable model over input features to classify or identify whether the particular news is fake or not. The main purpose of application is to provide a degree of disinformation present in the news through various parameters like credibility, confidence score.

## **1.5 Some of the important Terminology**

**Raw data**: It’s basically the input data which we will be fed to the model.

**Database**: Collection of information on different topics related to each other.

**Preprocessing**: Preprocessing includes various methods (tokenize, word check, stemming, tagging, topic collection, verb collection, topic vectorization) which will help us to convert the raw data into useful feature vectors.

**Linguistic approach**: This approach (e.g., natural language processing or NLP) are focused on news content, and aim to investigate fake news patterns by analyzing underlying semantics.

**Network approach**: The network approaches leverage existing knowledge networks to check facts of news.

**Classifier**: An algorithm that implements classification, especially in a concrete datasets. The classifier will help to classify the input into various categories.

**Model**: Model here is basically can be machine learning model (Logistic regression, SVM), Deep learning model (Pretrained AlexNet, VGGNet, ResNet, MobileNetv2), NLP (Knowledge graph, Custom model using CNN , BERT, OpenAI’s GPT-3 , XLNet , Capture score Integration (CSI) Model, Ensemble technique).

**Deceptive**: giving an appearance or impression different from the true one; misleading.

**1.5 Assumptions and Dependencies**

1. The raw text and designed algorithm is for english language and the text the application parses must write in this language.

2. We assume that the news which is categorized as fake is not compelling to most portion of population having similar ideological beliefs

3. We assume that that the ground truth of the data used for training of our model is not a disdain but from well reputed source.

**2. OVERALL DESCRIPTION**

2.1 Overall description of proposed solution

1. Hybrid Based approach For characterizing of Fake News based on

* Feedback Based Identification

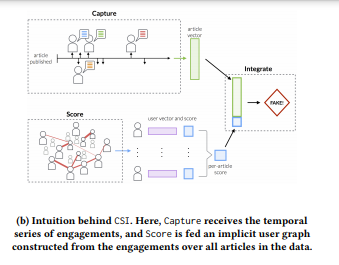
1. **Hand-crafted features**
2. **Propagation pattern analysis**
3. **Temporal pattern analysis**
4. **Response text analysis**
5. **Response user analysis**

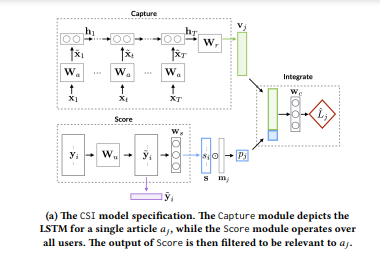
**User features**. User features can be obtained based on two types of information. One is the features extracted from user profiles on social networks. The second is features extracted based on user behaviors from content sharing and response patterns of the user.For performing and analyzing the User features we have used Capture-Score-Integration (CSI) Mode.

**1. Capture-Score-Integration (CSI) Model**

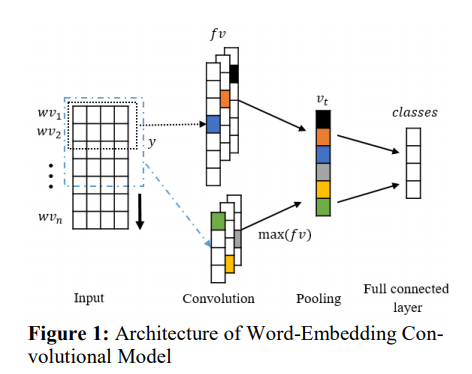
* Capture: To extract temporal representations of articles we use a Recurrent Neural Network (RNN)
* Score: To compute a score si and representation yi , user features are fed into a fully connected layer and a weight is applied to produce the scores vectors s.
* Integrate: The outputs of the two modules are concatenated and the resultant vector is used for classification

Dataset used: **Twitter and Weibo**





* First, we use word embedding to convert each word in the tweet into a vector. We randomly initialize the word embedding matrix. Each row of this matrix is a vector that represents a word in the vocabulary.
* Then we learn the embedding weights during the training process. Second, we concatenate these word vectors to produce a matrix representing the sentence.



**2. Content Based Identification**

1. **Cue and feature methods**
2. **Linguistic analysis methods**
3. **Deep learning content-based**

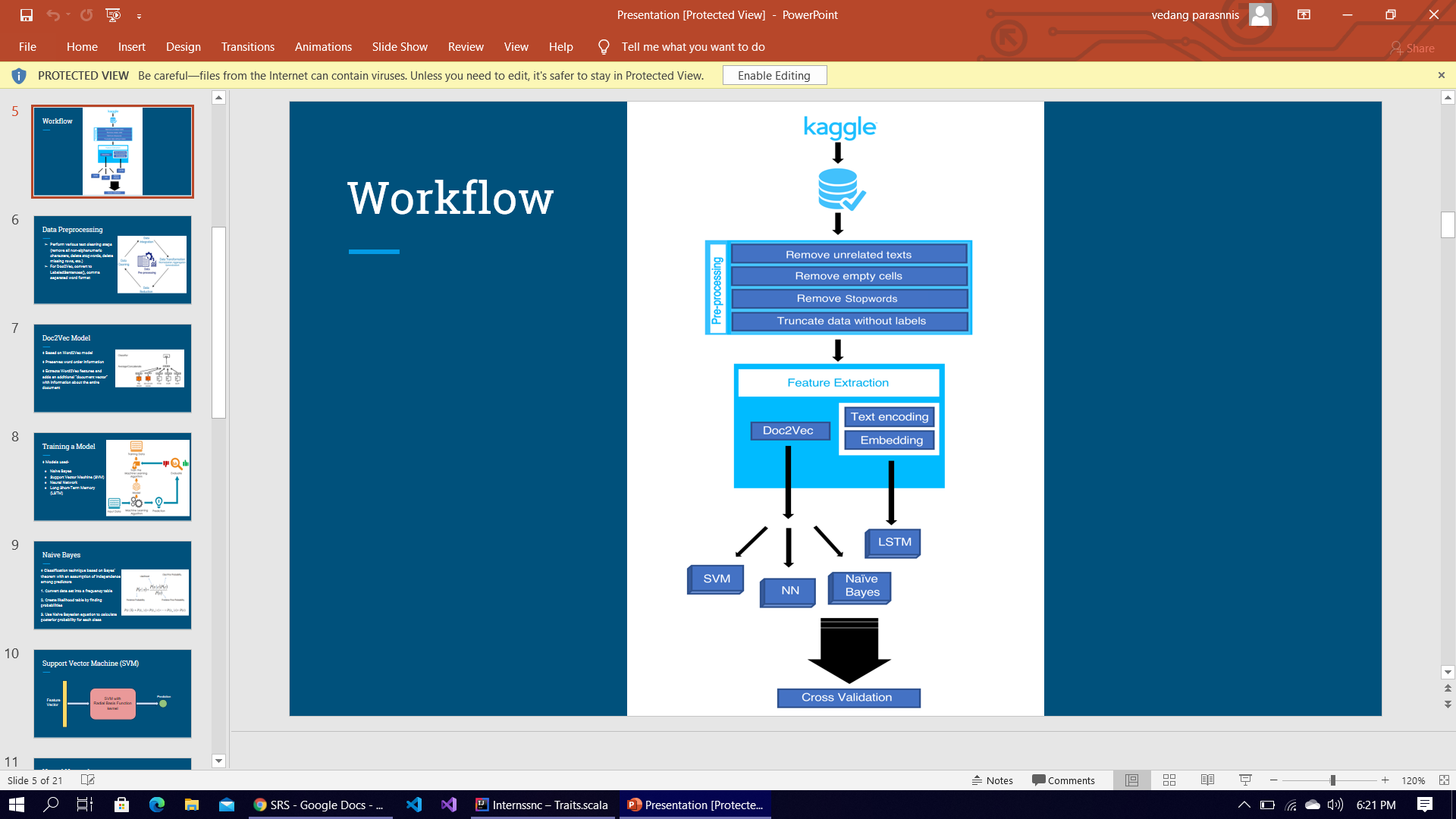
**Ensemble technique**

The Ensemble technique enables us to alleviate the shortcomings of linguistic analysis based methods by automatic feature extraction, being able to extract both simple features and more complex features that are difficult to specify. Ensemble learning based methods have demonstrated significant advances in text classification and analysis and are powerful methods for feature extraction and classification with their ability to capture complex patterns relevant to the task.

The given Ensemble techniques are used for developing machine learning and deep learning models for fabricated fake news detection and intent upto which is fabricated.We mainly use Doc2vec as our feature extraction technique and provide it as input to our Neural Network

We have a special focus on feature engineering technique using conditional and shared Bidirectional Long Short-Term Memory (BiLSTM) and utilized the effectiveness of fully connected neural networks with dropout.

Understanding the effectiveness of Gradient Boosting-based models for shared tasks, we trained a classifier using LightGBM technique which boosted the performance of our models and thereby providing a cross validation based on the efficiency of different models to maximum accuracy with our test dataset.



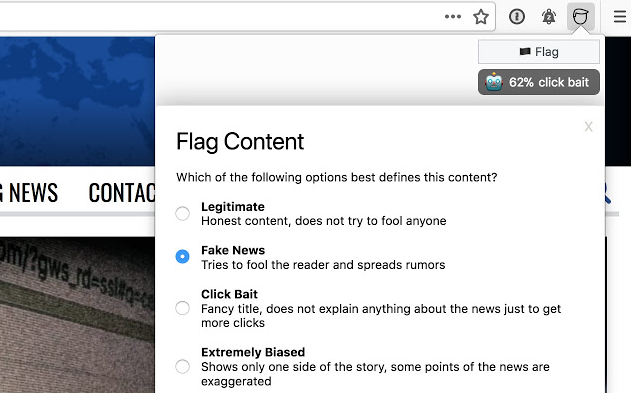
**Dataset** : Fake news dataset (kaggle)

**Current Work:** To enhance the accuracy of this approach we are trying to implement High performing **Natural Language Interface (NLI)** models are independently trained and ensembled with a fine-tuned BERT model to determine soft labels, which are then used to fine-tune the original Natural Language Interface models, BERT, and the Decomposable Attention model

**2.2** **User Interfaces**

The user interfaces will be an icon in the browser. While using the browser, the user's click will trigger a panel containing 1 button. With these buttons the user will activate the extension from the current website. The extension will fetch or crawl the necessary information (headling, date, content, url) from the website. After that raw data will be fed to the model as input which will classify whether the news is fake or not and also give credibility score.

**Sample Design**



**2.2.1 Hardware Interfaces**

Not applicable.

**2.2.2** **Software interfaces**

**Operating system:** windows

**Web client**: chrome

**Database :** NoSQL (In memory database)

**Cloud infrastructure support:** AWS EC2 instance

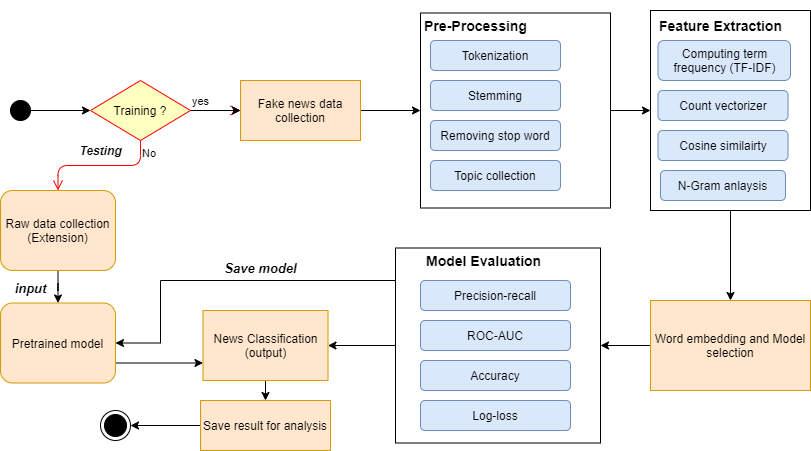
**Machine learning model deploy :** Docker, Amazon SageMaker - Amazon AWS

**Programming language**: Python, Javascript

**Library :** Tensorflow, keras, Flask, React.js

**platform** : Google colab, Jupyter notebook

**Activity diagram**



**3. Specific requirements**

**3.1 Functional Requirements**

As mentioned at section 2.2, these requirements are categorized by use cases. For any specific use case, there are specific requirements which are detailed below.

**High Priority**

1. The system should help in determining whether a given news article is fake based either on content of the news or on the basis of feedback to the given news by other users in a social environment
2. The system would be agile to identify and mark any specific source of the news to be fake or real based on number of fake articles generated from that source
3. The System will provide an graphical user friendly interface for better understanding of results
4. The system shall provide reliable and accurate results to the end user.
5. The System will be adaptable and able to learn by itself and is able to adapt to various types of disinformation like Satire,Parody,Clickbait etc

**Low priority :**

1. System should analyze the type of content and information generated by the sources (news website). if the news generated by the source is false and crosses the particular threshold then the source is added in fake news website collection.
2. The system shall maintain confidentiality and not try to leak private information of the user.
3. The system shall allow the user's status to be stored for the next time he returns to the web site. This will save the user minutes per visit by not having to provide the same details again.
4. The system shall translate web pages into the languages of the countries where the company's products are available. This will improve customer service and reduce the number of support calls from foreign customers.

**3.2 Non-functional Requirements**

**3.2.1 Balancing timeliness v/s detection accuracy**: Early detection and mitigation are critical goals of any effective system. However, the available information for detection increases as time progresses, with only the content of the article being available at the start, followed by increasing user responses as propagation continues. Most existing methods either rely on content only or on user responses only, or do not utilize responses incrementally. Detection systems must aim to utilize incrementally available information to trade-off confidence in detection accuracy v/s timeliness of the detection and mitigation effort.

**3.2.2 Prioritization and cost-effectiveness**. The ability to optimally decide which contents to factcheck at what time, can equip the system in providing better responses by being able to quickly remove false information that can have a potentially larger and faster impact than those that might have a negligible or slower impact if allowed to propagate further in time. Also, human involvement in fact-checking increases not only the delay but also the cost of intervention, which necessitates the need for prioritization of information to manually fact-check, until reliable automated methods can be sought.

**3.2.3** **Robustness, scalability and interpretability**. The high stakes and consequences of fake news necessitates the need for reliability in detection. Mistakenly removing true information from the platform, or not detecting and removing potentially viral false information would become problematic in practice. To move from manual and semi-automated solutions to fully automated ones will not be possible without robust and also interpretable predictions.

**3.2.4 Accuracy & Performance.**

Most ML work reports on algorithm accuracy (often precision and recall), i.e., how “correct” the output is compared to reality. Further work looks more broadly at algorithm performance i.e including comparisons of performance in specific contexts.