**INNOVATION DOCUMENT: FAKE NEWS DETECTION FOR USING NLP**

**A Proposed Ensemble Voting Model for Fake News Detection:**

Fake news or rumors are a phenomenon that significantly influences our social lives. Politicians in the political world usually rely on fake news as a powerful mechanism to change public opinion. Fake news spread through the media poses a real threat to the credibility of information, and the detection of fake news has attracted increased attention in recent years. Therefore, it becomes highly necessary to develop a method to identify fake news. This paper proposes a new ensemble voting model for detecting fake news in online text using a hybrid of machine learning and deep learning algorithms. Our ensemble model consists of three algorithms, namely, Convolution Neural Network (CNN) Gated Recurrent Unit (GRU) model of Recurrent Neural Network (RNN) and Random Forest.

**DEEP LEARNING ARCHITECHTURE:**

Fake news is defined as a made-up story with an intention to deceive or to mislead. In this paper we present the solution to the task of fake news detection by using Deep Learning architectures. Gartner research [1] predicts that “By 2022, most people in mature economies will consume more false information than true information”. The exponential increase in production and distribution of inaccurate news presents an immediate need for automatically tagging and detecting such twisted news articles. However, automated detection of fake news is a hard task to accomplish as it requires the model to understand nuances in natural language. Moreover, majority of the existing fake news detection models treat the problem at hand as a binary classification task, which limits model’s ability to understand how related or unrelated the reported news is when compared to the real news. To address these gaps, we present neural network architecture to accurately predict the stance between a given pair of headline and article body. Our model outperforms existing model architectures by 2.5% and we are able to achieve an accuracy of 94.21% on test data.

**PREDICTIVE SYSTEM ACCRACY:**

The fake news on social media and various other media is wide spreading and is a matter of serious concern due to its ability to cause a lot of social and national damage with destructive impacts. A lot of research is already focused on detecting it. This paper makes an analysis of the research related to fake news detection and explores the traditional machine learning models to choose the best, in order to create a model of a product with supervised machine learning algorithm, that can classify fake news as true or false, by using tools like python scikit-learn, NLP for textual analysis. This process will result in feature extraction and vectorization; we propose using Python scikit-learn library to perform tokenization and feature extraction of text data, because this library contains useful tools like Count Vectorizer and Tiff Vectorizer. Then, we will perform feature selection methods, to experiment and choose the best fit features to obtain the highest precision, according to confusion matrix results.

**ROBUSTNESS IN FAKE NEWS DETECTION:**

Fake news content on social media platforms has significantly increased due to the limited control of its propagation. This news, generated at high volume and speed has very few of them annotated by experts as true news or fake news. In a bid to ensure that fake news on social media is detected in a timely manner, a novel semi-supervised deep learning pipeline is proposed to learn from the limited labeled data as well as the vast amount of unlabeled data available. The pipeline consists of a shared deep neural network layer that extracts the low-level features that exists in both labeled and unlabeled data together. These features are fed into a two-path deep neural network where one path is implemented for supervised learning, while the other path is implemented for unsupervised learning. The supervised path learns from the available labeled data, while the unsupervised path learns from the unlabeled data which is often more in proportion. This pipeline is implemented using recurrent neural networks (RNN) as base models, and joint optimization of the two paths is implemented to complete semi-supervised learning. This pipeline is evaluated by testing the implemented models on two benchmark datasets: LIAR and PHEME. Results from the experiments demonstrate that the pipeline is robust and adaptive to different models and hyperparameters, and the models implemented using the semi-supervised pipeline proposed effectively detects fake news even with very few labeled data FAKE NEWS DETECTION