Fake and Real News Detection

Abstract:

With the increasing number of Fraud website on the internet which increases the number of fake news day by day, people are profiting by clickbaits and publishing fake news on online. By clicking on a clickbait, users are led to a page that contains false information. More clicks contributes to more money for content publishers and fake news influences people perceptions.

The rise of Fake news has become a global problem that even major tech companies like Facebook and Google are struggling to solve. It can be difficult to determine whether a text is factual without additional context and human judgement.

Problem Statement:

The given news dataset consists of two folders consisting of True and Fake news articles. The objective is to develop a robust model that can accurately identify and classify a given news as Real or Fake. The project will involve the following steps:

1. **Data Preparation**: Preprocessing the dataset by Cleaning Data, Bag of Words, Stemming, Lemmatization, Tokenization, MultinomialNB Algorithm, TF-IDF, word vectorization, word2 vec and splitting the dataset into training and testing sets.
2. **Model Building**: Developing a Recurrent Neural Network (RNN) model from scratch that can accurately classify the text as Real or Fake news.
3. **Model Evaluation**: Evaluating the model's performance on the testing set to measure its accuracy and other performance metrics such as Precision, Recall, and F1-score. The model's generalization ability will also be evaluated by testing it on new and unseen news.

Scope of the project:

- Students will learn to load and prepare text data for building a binary classifier
- To perform feature extraction from text
- To build a robust binary classifier for text classification problem
- Given a new unseen text message, classify it as Fake or Real News.
- Students will gain better knowledge on Stemming, Lemmatization, MultinomialNB, Word2 vec, TF-IDF etc.