Fake News Detection Using Machine Learning

Abstract:

In our modern era where the internet is ubiquitous, everyone relies on various online resources for news. Along with the increase in the use of social media platforms like Facebook, Twitter, etc. news spread rapidly among millions of users within a very short span of time. The spread of fake news has far-reaching consequences like the creation of biased opinions to swaying election outcomes for the benefit of certain candidates. Moreover, spammers use appealing news headlines to generate revenue using advertisements via clickbait's.

In this project, we aim to perform binary classification of various news articles available online with the help of concepts pertaining to Artificial Intelligence, Natural Language Processing and Machine Learning. We aim to provide the user with the ability to classify the news as fake or real and also check the authenticity of the website publishing the news.

Existing System:

There was a company which is released their product white paper where their method for fake news detection supported the feedback for the precise news within the micro blogs. the authors actually develop two systems for deception detection supported support vector machines and Naive Bayes respectively. They collect the info by means of asking people to directly provide true or false information on several topics – abortion, execution and friendship.

The accuracy of the detection achieved by the system is around 70%. This text describes an easy fake news detection method supported one among the synthetic intelligence algorithms – naïve Bayes classifier, Random Forest and Logistic Regression. The goal of the research is to look at how these particular methods work for this particular problem given a manually labelled news dataset and to support (or not) the thought of using AI for fake news detection. The difference between these article and articles on the similar topics is that during this paper Logistic Regression was specifically used for fake news detection; also, the developed system was tested on a comparatively new data set which gave a chance to gauge its performance on a recent data.

Proposed System:

This Project explains the system which is developed in three parts. The first part is static which works on machine learning classifier. We studied and trained the model with 4 different classifiers and chose the best classifier for final execution. The second part is dynamic which takes the keyword/text from user and searches online for the truth probability of the news. The third part provides the authenticity of the URL input by user.

Software Tools:

1. VS Code
2. Jupyter Notebook
3. Heroku
4. Anaconda
5. Python3
6. Scikit-Learn
7. Pandas
8. Matplotlib

Hardware Tools:

1. Laptop
2. Operating System: Windows 11
3. RAM: 16GB
4. ROM: 4GB
5. Fast Internet Connectivity