FAKE NEWS DETECTION USING ML APPROACHES: A SYSTEMATIC REVIEW

The easy access and exponential growth of the information available on social media networks has made it intricate to distinguish between false and true information. The easy dissemination of information by way of sharing has added to exponential growth of its falsification. The credibility of social media networks is also at stake where the spreading of fake information is prevalent. Thus, it has become a research challenge to automatically check the information viz a viz its source, content and publisher for categorizing it as false or true. Machine learning has played a vital role in classification of the information although with some limitations. This paper reviews various Machine learning approaches in detection of fake and fabricated news. The limitation of such and approaches and improvisation by way of implementing deep learning is also reviewed

**Existing system:**

* The core task of detecting fake news involves identifying the language (set of words or sentences) which is used to deceive the readers. The idea of classifying fake news by learning word-level meaning is a very challenging task under the skin.
* Detecting fake news is hard for many reasons. First, manual task of identifying fake news is very subjective. Assessing the veracity of a news story is a complex and cumbersome task, even for trained experts. News is not only spread through traditional media outlets anymore but also through various social media channels. Automated solution requires understanding the natural language processing which is difficult and complex. These complexities make it a daunting task to classify text as fake news

**Proposed system:**

* The input features for our model consists of Tf-Idf word vector representations of article-headline pair, Cosine similarity between article-headline pairs represented using Tf-Idf, and cosine similarity between article-headline pair represented using Google’s word vectors (Word2Vec). We computed Tf-Idf scores on unigrams and bigrams. To avoid bias due to unbalance in dataset, we did not consider words which appeared in more than 50% of all training documents and excluded the words which appeared in less than 50 documents.
* Our model attempts to capture the relative importance of a word present in articleheadline pairs locally (how important is the word for that specific headline-article pair) and globally (how common or uncommon that specific word is in relation to all the words in the corpus). In order to capture the similarity between the headlinearticle pair, we calculated the Cosine similarity between Headline- Article Tf-Idf pairs

**SYSTEM SPECIFICATION:**

**HARDWARE REQUIREMENTS:**

* **System :** intel core i3.
* **Hard Disk :** 1TB.
* **Ram :** 4gb.

**SOFTWARE REQUIREMENTS:**

* **Operating system :** Windows family.
* **Coding Language :** Python.
* **Front-End :** Python.
* **Designing :** Html, css, javascript.
* **Data Base :** MySQL.