Paper Title:
Clickbait Post Detection using NLP for Sustainable Content

Paper Link:

1.Summary
1.1 Motivation:
The prevalent issue of clickbait in online content serves as the driving force behind this project, which aims to create a reliable method for identifying clickbait headlines. Content with little substance is frequently the result of clickbait, which is intended to tempt readers with spectacular claims. Encouraging users, saving them precious time, and enhancing the whole online content experience are the driving forces. In the end, by giving people the ability to distinguish between reliable and deceptive content, the research hopes to contribute to a more dependable as well as knowledgeable online environment.

1.2 Contribution:
The research makes a substantial contribution to the identification of clickbait by comparing and recommending the TF-IDF and Word2Vec methodologies. The Voting Classifier's better performance is emphasised, along with its limits and the critical significance that algorithm selection plays. Insightful information for practical applications is offered, and credibility is increased by using a clear dataset. The study provides fresh insights on the identification of clickbait, especially with regard to algorithmic selections and how they affect performance metrics.

1.3 Methodology:
This research uses a dataset of 32,519 articles and an accurate approach for clickbait detection. Tokenization, lemmatization, and the elimination of non-alphabetic characters are all part of preprocessing. Clickbait and non-clickbait headlines are organized into separate data frames using Word2Vec and TF-IDF feature representation. After data concatenation, data shuffling, and storing in an Excel sheet, machine learning methods such as Voting Classifier and Logistic Regression are applied.

1.4 Conclusion:
In summary, this study marks a major advancement in our knowledge of how to identify and evaluate clickbait content. The study has improved our understanding of how to discern between attention-grabbing headlines that may be deceptive and more reliable material by carefully analyzing a sizable dataset.

2. Limitations
2.1 First limitation:(Limited Understanding of Model Behaviour)

The lack of a thorough explanation of false positives and false negatives in the research makes it difficult to assess the clickbait detection models' true performance in various situations.

2.2 Second limitation:(Adaptability Issues)

The extent to which the suggested strategies are effective across various platforms, languages, and data types is not fully investigated in the research. This reduces our faith in the system's wider applicability.

3. Synthesis

The essence of this work is a thorough examination of clickbait identification using TF-IDF and Word2Vec techniques. Using important criteria, the study assesses model performance through rigorous preprocessing and several machine learning algorithms. Remarkably, the Voting Classifier exhibits consistently better performance in both Word2Vec and TF-IDF methods, with notable improvements in recall, accuracy, precision, and f1 score. Although the study has various limitations, it provides important insights into building trustworthy clickbait detection systems, highlighting the critical role that algorithm selection has in producing accurate and well-rounded results.