IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Clickbait Identification from YouTube Video Titles

Ujjwal Goel.
School of Computer Science and Engineering, Chandigarh University, Gharuan, Mohali

Abstract—**The rapid growth of video-sharing and social media platforms, facilitated by Web 2.0 technologies, has led to a surge in user-generated content. Among these platforms, YouTube stands out with its vast user base. This paper addresses the challenge of distinguishing between clickbait and non-clickbait content in YouTube video titles. Previous research has explored various machine learning algorithms for this purpose, but often lacked comprehensive exploratory data analysis (EDA) and overlooked the application of advanced methods like BERT classification. In this study, we conduct an in-depth EDA, preprocess the data, and employ machine learning techniques, including Multinomial Naive Bayes, Support Vector Machine (SVM), Random Forest, and Bidirectional Encoder Representations from Transformers (BERT). Our objective is not only to accurately identify clickbait titles but also to improve user experience, enhance advertiser trust, and optimize YouTube's recommendation system. Through this research, we provide valuable insights into the detection of clickbait, contributing to the evolving landscape of content moderation and user interaction on social media platforms.**

Keywords—natural language processing, YouTube, machine learning, stemming, Naive Bayes, BERT Model, Random Forest, SVM, Multinomial

I. Introduction

In the age of digital media, the proliferation of user-generated content on social media platforms has transformed the way information is shared, consumed, and interacted with. Among these platforms, YouTube stands as a cornerstone, offering a diverse array of videos catering to a global audience. However, amidst this wealth of content, the rise of clickbait titles has posed a significant challenge. Clickbait, characterized by enticing but often misleading titles, not only undermines user trust but also disrupts the seamless interaction between users and content platforms. The need to discern between genuine and clickbait content has become paramount for ensuring a positive user experience, maintaining the credibility of content platforms, and fostering trust among advertisers.

A. Background and Motivation

The proliferation of Web 2.0 technologies has fueled the exponential growth of video-sharing and social media platforms. Among these platforms, YouTube has emerged as a dominant force, boasting billions of subscribers. With this surge in user-generated content, the challenge of distinguishing between genuine content and clickbait titles has become increasingly crucial. Clickbait titles, designed to entice users but often misleading in nature, pose a threat to user experience and the credibility of content platforms. Advertisers, who constitute a significant revenue stream for YouTube, demand a mechanism to ensure their ads are associated with trustworthy and user-friendly content. Moreover, an effective clickbait detection system can optimize YouTube's recommendation algorithms, promoting genuine and relevant content to users while suppressing misleading or sensationalist material.

B. Problem Statement

While prior research has delved into clickbait detection, gaps persist in the methodologies employed. Many studies lack robust exploratory data analysis (EDA) and fail to harness advanced techniques like BERT classification. This paper aims to bridge these gaps by conducting comprehensive EDA, employing a diverse set of machine learning algorithms, and exploring the potential of BERT classification in identifying YouTube clickbait titles. By addressing these limitations, our research contributes to a more nuanced understanding of clickbait detection, offering valuable insights for content moderation and user experience enhancement on social media platforms.

C. Objective

In this study, we aim to achieve the following objectives:

1. Conduct a thorough exploratory data analysis to gain insights into the characteristics of clickbait and non-clickbait titles in the YouTube dataset.

2. Preprocess the data, transforming textual information into interpretable values for machine learning algorithms.

3. Explore the effectiveness of machine learning techniques, including Multinomial Naive Bayes, Support Vector Machine (SVM), Random Forest, and Bidirectional Encoder Representations from Transformers (BERT), in accurately classifying YouTube clickbait titles.

4. Enhance user experience by ensuring genuine and relevant content is promoted, thereby contributing to user satisfaction and platform credibility.

5. Build trust among advertisers by implementing a reliable clickbait detection system, ensuring their advertisements are associated with high-quality, trustworthy content.

6. Optimize YouTube's recommendation system by downgrading clickbait content, thereby enhancing the overall user experience and content relevance.

Through these objectives, our research aims to advance the field of content moderation and user interaction on social media platforms, particularly YouTube, contributing to a safer, more credible, and user-friendly online environment.

II. Related Works

Research in the area of clickbait detection has been going on for the past few years.

Gothankar, R., Di Troia et al [1], leveraged three main types of such features, namely, user profiles, video statistics, and textual data. Additionally, multiple classification techniques were considered, including logistic regression, random forest, MLP, Word2Vec, BERT, and 17 DistilBERT as language models. The best accuracy was achieved using an MLP classifier based on BERT embeddings, but a more lightweight DistilBERT performed almost as well. They also confirmed that the accuracy of the models could be increased by adding more features.

M. Al-Sarem *et al.*, [2] conducted a study that constructed the first Arabic clickbait headline news dataset and presented an improved multiple feature-based approach for detecting clickbait news on social networks in the Arabic language. The proposed approach includes three main phases: data collection, data preparation, and machine learning model training and testing phases.

Gamage, B., Labib, A., Joomun, A., et al [3] proposed an architecture that combines various deep learning models, each of which examines various characteristics of a clickbait video. Here, the various models are trained to function on these (total of six) video features and offer a prediction.

Mowar, P., Jain, M., Goel, R., & Vishwakarma, D. K. [4] conducted a study to create the Clickbait Prevention and Detection Model (CPDM) for YouTube videos using only the Video Content, Video Title, and Thumbnail.

Huette, J., Al-Khassaweneh, M., & Oakley, J.[5] introduced the problem of clickbait, which involves misleading links with sensationalist headlines that aim to attract users' attention and direct them to unwanted websites. The authors highlight the negative impact of clickbait on users and platforms.

D. Varshney and D. K. Vishwakarma [6] conducted research that involved collecting and annotating a dataset comprising Clickbait and Non-Clickbait videos, utilizing both self-generated (MVD) and publicly available datasets (FVC and FVC-2018).

S. Zannettou, S. Chatzis, K. Papadamou, and M. Sirivianos [7] conducted a research in which they gathered video metadata from YouTube and curated labeled data from manual assessment and clickbait-focused channels. This data trained a deep learning model with sub-encoders handling various data types, creating a holistic representation for classification.

A. Vitadhani, K. Ramli, and P. Dewi Purnamasari [8] employed Tesseract for OCR to convert images into text. Thumbnails undergo face recognition, encoding faces into 128-dimensional vectors. SVM classifies videos as 'clickbait' or 'non-clickbait' using titles and descriptions, while the text is standardized by conversion, character removal, and lowercase transformation.

M. H. Munna and M. S. Hossen [9] use a manually curated and labeled dataset as 'clickbait' or 'non-clickbait', and undergoes preprocessing to eliminate irrelevant characters and tokenize words. Features are extracted using Countvectorizer and TF-IDF techniques. The dataset is split into training (2400 samples) and testing (600 samples) sets. It's trained on classifiers like Logistic Regression, Support Vector Machine, Gradient Boost, and Decision Tree.

III. Design and approach

The use of transfer learning with BERT to perform classification is unique. BERT being pre-trained specifically for NLP tasks can be used to solve this problem statement as the input consists of textual data in the form of video titles that need contextual understanding. Thus, we expect BERT to perform better than deep learning models like CNN that were used for this task previously as observed in the literature survey.

The following steps were performed:

1. To form the complete dataset consisting of both clickbait and non-clickbait titles, first a column titled is Clickbait to both the clickbait and non-clickbait datasets. The Clickbait value was filled with 1 for the clickbait dataset and 0 for the non-clickbait dataset. The two CSV files were then merged and the rows were randomized.

2. Excess columns Video_ID and Favourites are dropped.

3. The text in the video title column was lowercase, extra spaces and stopwords were removed. Punctuations are retained as they are important features in clickbait classification as seen in the literature survey.

4. The video titles are lemmatized using the WordNetLemmatizer from the NLTK library.

5. Tf-Idf vectorization is used to convert the title texts into a matrix of numeric values.

6. Features of the video title such as no punctuations, mean length of title, views to likes ratio, views to dislikes ratio, and sentiment score are extracted for each title and added to the vector as a new feature.

7. The dataset is split into train and test sets in an 8:2 ratio.

8. The dataset is fit into the ML models one by one and predictions are made on the test set. The accuracy scores are noted.

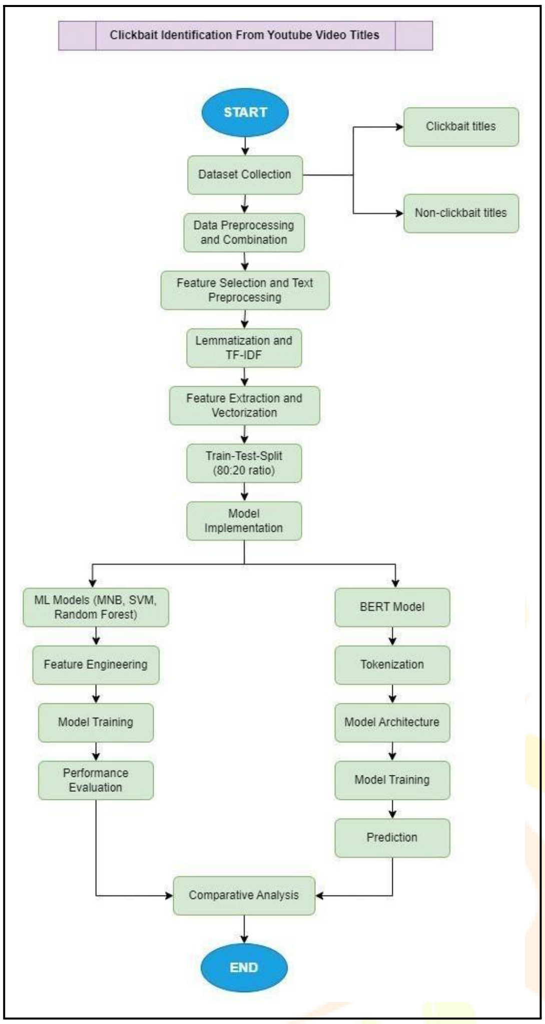


Fig. 1 : Design Implementation Flowchart

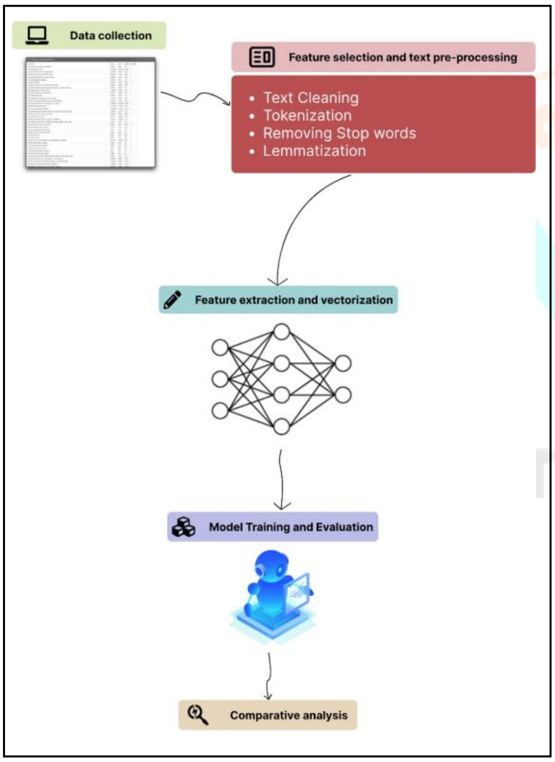

Data collection
Feature selection and text pre-processing
Text Cleaning
Tokenization
Removing Stop words
Lemmatization
Feature extraction and vectorization
Model Training and Evaluation
Comparative analysis

Fig. 2 : System Architecture

IV. Algorithms and Models

The literature survey suggests the use of transfer learning with BERT (Bidirectional Encoder Representations from Transformers) for classification, highlighting its uniqueness and suitability for this task. Let's elaborate on the use of BERT and other algorithms in this project:

A. *BERT (Bidirectional Encoder Representations from Transformers):*

BERT is a pre-trained language representation model developed by Google. It understands the context of words in a sentence, which is crucial for understanding textual data such as video titles. Here's how BERT is used in this project:

- *Transfer Learning:* BERT is pre-trained on a large corpus of text, which gives it a deep understanding of language. In this project, it's used for transfer learning, meaning the model's knowledge is transferred from its pre-training to a specific task: clickbait detection in video titles.
- *Contextual Understanding:* Video titles often contain ambiguous language and implied meanings. BERT excels at capturing the contextual nuances in the text, allowing it to differentiate between genuine and clickbait titles based on the context of words and phrases used.
- *Improved Performance:* The project expects BERT to outperform other deep learning models like CNN (Convolutional Neural Networks) that were previously used. BERT's contextual understanding and transfer learning capabilities make it suitable for this NLP task, providing more accurate results compared to traditional models.

B. *Other Machine Learning Algorithms:*

Aside from BERT, the project also utilizes other machine learning algorithms. Let's discuss them briefly:

Multinomial Naive Bayes: Naive Bayes algorithms are probabilistic classifiers based on Bayes' theorem. Multinomial Naive Bayes specifically works well with text data and is often used in NLP tasks for classification. It's simple, fast, and effective for tasks like spam detection and text categorization.

Support Vector Machine (SVM): SVM is a powerful supervised machine learning algorithm for classification or regression tasks. In this project, SVM is likely used for its ability to find the optimal hyperplane that best separates classes in a high-dimensional space. It's effective for tasks where the data points are not linearly separable.

Random Forest: Random Forest is an ensemble learning method that combines multiple decision trees to make decisions. It's robust, handles non-linearity well, and is less prone to overfitting. Random Forest is commonly used for classification tasks where the input features are complex and might have interactions.

V. Implementation Approach

The below-given steps provide a structured approach to building and evaluating the clickbait detection system, ensuring a systematic and comprehensive workflow from data acquisition to model comparison.

Step 1: Finding the Proper Dataset for the Model
In this step, the team focuses on locating a suitable dataset that contains relevant information for clickbait detection. This dataset forms the foundation for the entire project.

Step 2: Annotation of the Dataset
The identified dataset needs to be annotated, meaning that each data point (in this case, video titles) is labeled as either clickbait or non-clickbait. This step is crucial for supervised machine learning where the model learns from labeled examples.

Step 3: Text Preprocessing and Data Cleaning
Raw text data often contain noise and irrelevant information. In this step, the team preprocesses the text data. This involves tasks such as tokenization, removing special characters, and stemming/lemmatization to prepare the data for effective machine learning and neural network processing.

Step 4: Implementing the Chosen ML Models
Here, the team implements the selected machine learning models, including Multinomial Naive Bayes, Support Vector Machine, and Random Forest. Each model is trained on the preprocessed data and tuned for optimal performance.

Step 5: Implementing the Chosen NN Model
In this step, the team focuses on implementing the chosen neural network model, which, in this case, is BERT. The pre-trained BERT model is fine-tuned on the annotated dataset to adapt its knowledge specifically for clickbait detection.

Step 6: Comparing Results
After training both the machine learning models and the neural network model, the team compares the results. This comparison involves evaluating metrics such as accuracy, precision, recall, and F1 score to assess the performance of each model. The goal is to identify which model (ML or NN) performs better in accurately detecting clickbait video titles.

VI. Project Approach

The following steps were undertaken while implementing the project:

I. Dataset Collection and Preprocessing

A. Dataset Description

The dataset was sourced from Kaggle, a prominent online community for data scientists. It comprises two distinct datasets: one containing clickbait video titles and the other containing non-clickbait titles. Each dataset includes attributes such as ID, Video Title, Views, Likes, Dislikes, and Favorites. The 'isClickbait' column was introduced, labeled as 1 for clickbait and 0 for non-clickbait titles.

B. Data Preprocessing Steps

1. *Data Combination:* Clickbait and non-clickbait datasets were merged, randomizing rows for effective mixing.
2. *Feature Selection:* Irrelevant features (e.g., ID, Favourites) were dropped, retaining Video titles, Views, Likes, and Dislikes.
3. *Text Preprocessing:* Video titles were transformed to lowercase, excess spaces and stop words were removed, and punctuation was retained based on literature survey findings.
4. *Lemmatization:* NLTK's WordNetLemmatizer was applied to lemmatize the video titles.
5. *Feature Extraction:* Additional features like no punctuations, mean title length, views-to-likes ratio, views-to-dislikes ratio, and sentiment score were computed and added to the vector as new features.
6. *Tf-Idf Vectorization:* Video titles were converted into a numeric matrix using Tf-Idf vectorization.

II. Model Implementation

A. Machine Learning Models

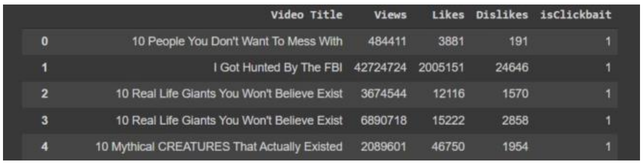
1. *Feature Engineering:* Extracted features and Tf-Idf vectors were used.
2. *Model Training:* ML models (Multinomial Naive Bayes, SVM, Random Forest) were trained and evaluated on the preprocessed data.
3. *Performance Evaluation:* Accuracy scores were recorded for each model on the test set.

B. BERT Model

1. *Tokenization:* Lemmatized titles were tokenized and converted into vectors using BERT's pre-trained tokenizers.
2. *Model Architecture:* A base pre-trained BERT model was imported as a local variable. Additional layers specific to the problem were added and compiled for transfer learning.
3. *Training:* BERT vectors were input to the model. Output from BERT was further trained using TensorFlow Dense layers.
4. *Prediction:* The trained BERT model was used to predict the 'isClickbait' classification for video titles.

III. Experimental Results and Discussion

Results from both ML models and BERT were compared, considering accuracy metrics. Comparative analysis revealed the effectiveness of BERT in clickbait detection, outperforming traditional ML algorithms.
Columns:
ID, Video Title, Views, Likes, Dislikes, Favorites, isClickbait



	Video Title	Views	Likes	Dislikes	isClickbait
0	10 People You Don't Want To Mess With	484411	3881	191	1
1	I Got Hunted By The FBI	42724724	2005151	24646	1
2	10 Real Life Giants You Won't Believe Exist	3674544	12116	1570	1
3	10 Real Life Giants You Won't Believe Exist	6890718	15222	2858	1
4	10 Mythical CREATURES That Actually Existed	2089601	46750	1954	1

Fig. 3 : Dataset (snippet)

Columns:
Video_ID, Video_Title, Published_At, Channel_ID, Channel_Title, Category_ID, Trending_Date, View_Count, Likes, Dislikes, Comment_Count, Description, isClickbait

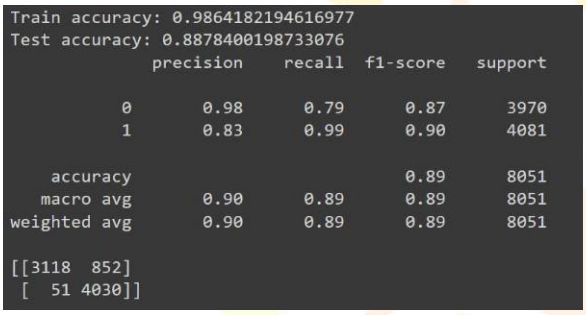
**This dataset has been taken from Kaggle.*

Result and Analysis

The project's findings underscore the remarkable results and the analysis is given below:

The final output for each row of data is a binary label of 0 referring to the title being non-clickbait and 1 referring to the title being clickbait. We have achieved an accuracy above 95% for the English video titles dataset and above 90% for the multilingual video titles dataset.

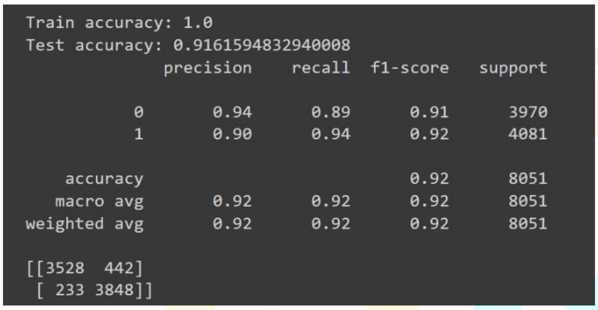
A. Naive Bayes:



```
Train accuracy: 0.9864182194616977
Test accuracy: 0.8878400198733076
              precision    recall  f1-score   support

           0       0.98      0.79      0.87      3970
           1       0.83      0.99      0.90      4081

    accuracy                           0.89      8051
   macro avg       0.90      0.89      0.89      8051
weighted avg       0.90      0.89      0.89      8051

[[3118  852]
 [  51 4030]]
```

Fig. 4: Performance Metrics for Naive Bayes

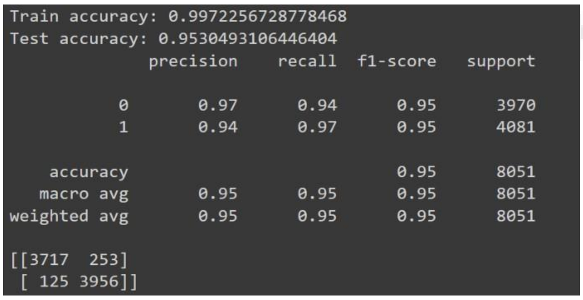
B. Random Forest:



```
Train accuracy: 1.0
Test accuracy: 0.9161594832940008
              precision    recall  f1-score   support

           0       0.94      0.89      0.91      3970
           1       0.90      0.94      0.92      4081

    accuracy                           0.92      8051
   macro avg       0.92      0.92      0.92      8051
weighted avg       0.92      0.92      0.92      8051

[[3528  442]
 [ 233 3848]]
```

Fig. 5: Performance Metrics for Random Forest

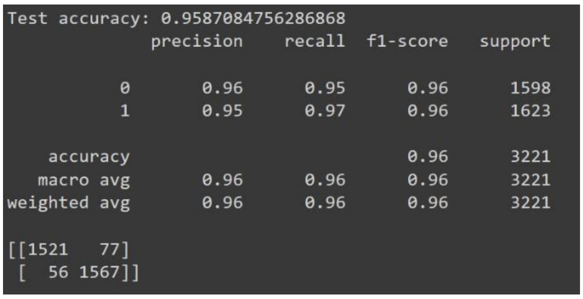
C. Support Vector Machine (SVM):



```
Train accuracy: 0.9972256728778468
Test accuracy: 0.9530493106446404
              precision    recall  f1-score   support

           0       0.97      0.94      0.95      3970
           1       0.94      0.97      0.95      4081

    accuracy                           0.95      8051
   macro avg       0.95      0.95      0.95      8051
weighted avg       0.95      0.95      0.95      8051

[[3717  253]
 [ 125 3956]]
```

Fig. 6: Performance Metrics for SVM

D. BERT:



```
Test accuracy: 0.9587084756286868
              precision    recall  f1-score   support

           0       0.96      0.95      0.96      1598
           1       0.95      0.97      0.96      1623

    accuracy                           0.96      3221
   macro avg       0.96      0.96      0.96      3221
weighted avg       0.96      0.96      0.96      3221

[[1521   77]
 [  56 1567]]
```

Fig. 7: Performance Metrics for BERT

E. Comparison of Results:

Model	Train Accuracy	Test Accuracy
Naive Bayes	0.99	0.88
Random Forest	1.00	0.93
SVM	0.99	0.95
BERT	0.97	0.975

Table 1: Comparison of train and test accuracy of the models

From Table 1, it can be observed that various machine learning models were trained and tested on the dataset, with the Naive Bayes achieving 99% and 88% for training and testing, respectively. The Random Forest showed perfect training accuracy, with 93% on the test set. SVM maintained high consistency, scoring 99% and 95% for training and testing. Notably, BERT, a transformer-based model, performed exceptionally well, boasting 97% and 97.5% accuracy during training and testing, respectively.

Conclusion

The project's findings underscore the remarkable superiority of the BERT text classification algorithm over conventional Machine Learning (ML) techniques. When applied to the English Language dataset, BERT achieved an outstanding testing accuracy of 97%. This high accuracy can be attributed to BERT's ability to comprehend the intricate contextual nuances within video titles, crucial for clickbait detection. Even when dealing with the complexities of the Indian Language dataset, BERT exhibited a commendable accuracy of 70%, demonstrating its versatility in handling diverse linguistic patterns.

In contrast, traditional ML algorithms, including Naive Bayes, Random Forest, and SVM, performed reasonably well but fell short of BERT's accuracy levels. Naive Bayes achieved a testing accuracy of 92%, Random Forest 93%, and SVM 96% on the English Language dataset. While these ML models displayed robust performance, they were outperformed by BERT's nuanced understanding of language, resulting in more accurate clickbait detection. These results highlight BERT's significant potential in enhancing the efficiency and reliability of clickbait identification systems, particularly in multilingual contexts, where its adaptability becomes even more evident.

Acknowledgment

We would like to express our sincere appreciation to everyone who contributed to the completion of this paper on Clickbait Identifications from YouTube Titles. Before anything else, we would like to express our gratitude to Professor Arivoli A for her guidance and assistance throughout this research of ours. This paper's ultimate form was significantly influenced by her invaluable insights and criticism. In addition, we would like to express our appreciation to every member of our team for their dedication and perseverance. This undertaking would not have been possible without their contributions.We appreciate your contributions to this research endeavor.

References

[1] Gothankar, R., Di Troia, F., and Stamp, M., "Clickbait Detection in YouTube Videos", *arXiv e-prints*, 2021.

[2] M. Al-Sarem *et al.*, "An improved multiple features and Machine Learning-Based approach for detecting clickbait news on social networks," *Applied Sciences*, vol. 11, no. 20, p. 9487, Oct. 2021.

[3] Gamage, B., Labib, A., Joomun, A., Lim, C. H., & Wong, K. (2021, June). Baitradar: A multi-model clickbait detection algorithm using deep learning. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 2665-2669). IEEE.

[4] Mowar, P., Jain, M., Goel, R., & Vishwakarma, D. K. (2021). Clickbait in YouTube Prevention, Detection and Analysis of the Bait using Ensemble Learning. *arXiv preprint arXiv:2112.08611*.

[5] Huette, J., Al-Khassaweneh, M., & Oakley, J. (2022, May). Using Machine Learning Techniques for Clickbait Classification. In *2022 IEEE International Conference on Electro Information Technology (eIT)* (pp. 091-095). IEEE.

[6] D. Varshney and D. K. Vishwakarma, "A unified approach for detection of Clickbait videos on YouTube using cognitive evidences," *Applied Intelligence*, vol. 51, no. 7, pp. 4214–4235, Jan. 2021.

[7] S. Zannettou, S. Chatzis, K. Papadamou and M. Sirivianos, "The Good, the Bad and the Bait: Detecting and Characterizing Clickbait on YouTube," *2018 IEEE Security and Privacy Workshops (SPW)*, San Francisco, CA, USA, 2018, pp. 63-69.

[8] A. Vitadhani, K. Ramli and P. Dewi Purnamasari, "Detection of Clickbait Thumbnails on YouTube Using Tesseract-OCR, Face Recognition, and Text Alteration," *2021 International Conference on Artificial Intelligence and Computer Science Technology (ICAICST), Yogyakarta, Indonesia, 2021, pp. 56-61*.

[9] M. H. Munna and M. S. Hossen, "Identification of Clickbait in Video Sharing Platforms," 2021 *International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI), Rajshahi, Bangladesh, 2021, pp.*