FEATURE EXTRACTION AND EXPERIMENTAL INVESTIGATION OF CLICKBAIT IN YOUTUBE VIDEOS

A PROJECT REPORT

Submitted by

Nithin P	**Reg.No. SCM18CS052**
Sharath K Nambiar	**Reg.No. SCM18CS069**
Shehzad Ibrahim	**Reg.No. SCM18CS070**
Yedu Krishna J	**Reg.No. SCM18CS085**

To

The APJ Abdul Kalam Technological University in partial
fulfillment of the requirements for the award of the Degree
Of
Bachelor of Technology
In
Computer Science and Engineering

Department of Computer Science and Engineering

SCMS SCHOOL OF ENGINEERING AND TECHNOLOGY

(Affiliated to APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY)

VIDYA NAGAR, PALISSERY, KARUKUTTY ERNAKULAM – 683 582

MAY 2022

DECLARATION

I undersigned hereby declare that the project report Feature Extraction and Experimental Investigation of Clickbait in YouTube Videos, submitted for partial fulfilment of the requirements for the award of degree of Master of Technology of the APJ Abdul Kalam Technological University, Kerala is a bonafide work done by me under supervision of Ms. Nisha S Raj, Assistant Professor, Department of Computer Science and Engineering, SCMS SCHOOL OF ENGINEERING AND TECHNOLOGY, Karukutty, Ernakulum. This submission represents my ideas in my own words and where ideas or words of others have been included, I have adequately and accurately cited and referenced the original sources. I also declare that I have adhered to ethics of academic honesty and integrity and have not misrepresented or fabricated any data or idea or fact or source in my submission. I understand that any violation of the above will be a cause for disciplinary action by the institute and/or the University and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been obtained. This report has not been previously formed the basis for the award of any degree, diploma or similar title of any other University.

Place:

Date:

Nithin P
Sharath K Nambiar
Shehzad Ibrahim
Yedukrishna J

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

SCMS SCHOOL OF ENGINEERING AND TECHNOLOGY

VIDYA NAGAR, PALISSERY, KARUKUTTY ERNAKULAM

CERTIFICATE

This is to certify that the report entitled “**Feature Extraction and Experimental Investigation of Clickbait in YouTube Videos**“ submitted by Nithin P (SCM18CS052), Sharath K Nambiar (SCM18CS069), Shehzad Ibrahim (SCM18CS070), Yedukrishna J (SCM18CS085) to the APJ Abdul Kalam Technological University in partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING, SCMS SCHOOL OF ENGINEERING AND TECHNOLOGY VIDYA NAGAR, PALISSERY, KARUKUTTY ERNAKULAM is a bonafide record of the project work carried out by him/her under my/our guidance and supervision. This report in any form has not been submitted to any other University or Institute for any purpose.

Internal Supervisor(s) External Supervisor (if any)

HEAD OF THE DEPARTMENT

ACKNOWLEDGEMNT

We are greatly indebted to Dr. Praveensal C J, Principal, SCMS SCHOOL OF ENGINEERING AND TECHNOLOGY, Karukutty, Ernakulam and Dr. Varun G Menon, Head of Department, Department of Computer Science and Engineering, SCMS SCHOOL OF ENGINEERING AND TECHNOLOGY, Karukutty, Ernakulam who whole heartedly permitted me to conduct this seminar.

I would like to thank my guide, Ms. Nisha S Raj, Assistant Professor, Department of Computer Science and Engineering, SCMS SCHOOL OF ENGINEERING AND TECHNOLOGY, Karukutty, Ernakulam who has given me valuable guidance and support throughout the seminar. Also, I would like to thank my seminar coordinator, Mrs. Josna Philomina, Assistant Professor, Department of Computer Science and Engineering SCMS SCHOOL OF ENGINEERING AND TECHNOLOGY, Karukutty, Ernakulam, who supported and instructed me all the way.

I would like to express my sincere gratitude to all the teachers of the Computer Science Department who gave us moral and technical support. I would like to thank the supporting staff in the computer lab whose dedicated work kept the lab working smoothly, thus enabling me to have access to various resources which helped me understand more about the seminar topic. I would also like to thank my friends and family members for providing me with the necessary resources and support. Last but not least I like to thank God Almighty for helping me to conduct the seminar hassle-free.

ABSTRACT

Social media sites such as Twitter, Facebook, and YouTube make it easy for users to express their thoughts, due to this lot of inaccurate and unreliable audio-visual information is frequently produced and shared through these channels.

YouTube is one of the most popular video-sharing platforms, with billions of users accounting for about one-third of the internet population. As a result, it is rife with videos that do not accurately depict the situation that it refers to.

The titles and thumbnails of these videos are purposely designed to attract the user’s attention and make them curious to follow the link and read, view, or listen to the attached content. Such videos are called Clickbait Videos.

The aim is to detect these clickbait YouTube Videos using multiple evidences that is selected from its metadata.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 GENERAL BACKGROUND

In the internet era that we’re living in, YouTube doesn’t actually require a formal introduction. YouTube is so popular that it’s almost impossible to find a person who doesn’t know about it. It covers a wide range of applications from tutorials about millions of topics to entertainment. It’s possible to say that YouTube can easily claim that no other app has attained such a high level of scope. It has such an influence on people that, it’s not possible for a person to spend a day without it. The wide horizon of success is due to the fact that it doesn’t inflict any financial cost to the viewers nor the content publishers. The platform provides freedom to the content creators without imposing any regulations to the legitimacy of the content. It also offers monetary profits to the content publishers or the channel admin, as we say in technical terms according to the viewership and the support a video gets. The support or viewership of a video can easily be boosted by sharing a link of the video further increasing the income earned by a video publisher. Thus, content creators started creating misleading but attractive contents to gain more viewership. The users use various parameters like thumbnail, description, title etc. to decide whether to watch the video or not. Content creators increase their view count by posting videos with thumbnails which do not match their content but potentially attract a wide range of audience. Clickbait videos are infecting the internet in a daily wide spread basis. Users get confused about the reality of facts and are misled into lies they’re made to believe. This creates a toxic environment in the internet. A system is needed for the detection of such clickbait videos.

YouTube removed the public dislike count from all its videos in November 2021. While creators can still see the number of dislikes they received for their video, viewers can only see the number of likes. This created a hole in the existing clickbait detection system with the number of dislikes being removed. Like to dislike ratio was the most promising parameter when it comes to detecting whether the video is a clickbait or not. This research thus focuses on finding the best and most accurate parameter from view-like ratio, comments, description and title. The research also focuses on the detection of clickbait with the absence of dislike count using conventional machine learning approaches like KNN, Decision Tree, LSTM, Sentimental Analysis and Logistic Regression. The dataset used for training and testing the algorithms was constructed from the scratch containing a list of 500 videos with all its necessary parameters.

1.2 OBJECTIVE

The objectives of this research are to

a) Find out the accuracy, precision and recall of each parameter and determine the best replacement for dislike.

b) Make a dataset containing 500 YouTube videos

c) Use the most accurate parameters and dataset to train and test KNN, naïve bayes, decision tree, LSTM, Sentimental Analysis and Logistic Regression.

CHAPTER 2

LITERATURE SURVEY

2.1 Clickbait detection

This paper proposes a new model for the detection of clickbait, i.e., short messages that lure readers to click a link. Clickbait is primarily used by online content publishers to increase their readership, whereas its automatic detection will give readers a way of filtering their news stream. The first clickbait corpus of 2992 Twitter tweets is compiled, 767 of which are clickbait, and, by developing a clickbait model based on 215 features that enables a random forest classifier to achieve 0.79 ROC-AUC at 0.76 precision and 0.76 recall.

2.2 Relationship between game theory and clickbait

Clickbait media is thought to be created through social media platforms' algorithmic curation. Despite the fact that clickbait can be risky, especially for heritage news organisations, it is widely used. The provision of clickbait is seen as a revised game with an arbitrary limit. It is observed that the behaviour of 37 German legacy news providers following algorithm changes using machine learning and analysis of Facebook posts and Twitter messages over 54 months. The findings show that clickbait is used infrequently, with few heavier-using sources, and that clickbait performance has a reversed U-shaped relationship, with the quantity of clickbait and the number of people engaging form a U-shaped relationship in reverse. News organisations as a whole adjust to an industry-wide clickbait standard. While it is not possible to prove that algorithmic curation generates clickbait, it can be shown that Facebook's regulatory involvement to reduce clickbait disperses uneven supply trends. It can contribute to a better understanding of editorial decision making in competitive situations that are subject to platform regulation.

2.3 Stylized Headline Generation.

Clickbaits are enticing social media postings or spectacular headlines designed to entice readers to click on them. Clickbait is all over social media, and it can have serious consequences for both consumers and the media industry. To address this difficulty, we suggest using recent advances in deep generative models to build synthetic headlines with specified styles and investigate their use in improving clickbait identification. The authors suggest, in particular, using style transfer to generate styled headlines from original papers. Furthermore, because generating stylized headlines is difficult due to issues such as the discrete nature of texts and the need to preserve the document's semantic meaning while achieving style transfer, proposed a novel solution called Stylized Headline Generation (SHG) that can not only generate readable and realistic headlines to enlarge original training data, but also aids supervised learning's classification capacity. SHG's success in generating high-quality and high-utility headlines for clickbait detection is demonstrated by experimental findings on real-world datasets.

2.4 OVCP

Online video sharing services (such as YouTube and Vimeo) have become a popular way for individuals to view video content. On internet video sharing platforms, clickbait video, whose substance plainly deviates from its title/thumbnail, has developed as a major issue. Current clickbait detection systems that primarily rely on evaluating the title text, thumbnail image, or video content have been demonstrated to be ineffective in detecting online clickbait videos. In this research, the authors present Online Video Clickbait Protector (OVCP), and unique content-agnostic approach for detecting clickbait films by examining the comments left by the audience that watched the video. OVCP, unlike other systems, does not directly evaluate the video's content and pre-click information but instead, goes through the different features that can be extracted from the different comments that are left by different users who have watched the video and, in the end, using these features we finally classify the video as Clickbait or not.

2.5 Identifying Clickbait Posts on social media with an Ensemble of Linear Models

Making a link so alluring that people click on it is the goal of clickbait. However, the content of such publications frequently has little to do with the title, exhibits poor quality, and ultimately dissatisfies the reader. The creators of the clickbait challenge (http://www.clickbait-challenge.org/) invited the participants to create a machine learning model for grading articles based on their "clickbaitness" in order to benefit the readers. The strategy was successfully evaluated in the challenge, where it demonstrated excellent performance of 0.036 MSE and placed third out of all the solutions to the challenge. In this research, they proposed to address the clickbait problem with an ensemble of Linear SVM models.

2.6 Machine Learning Based Detection of Clickbait Posts in Social Media

In this study, a dataset containing over 21,000 headlines and titles from the 2017 Clickbait Challenge (clickbait-challenge.com), each of which is annotated with at least five crowdsourced assessments of its clickbaitness. Develop a reliable computational clickbait detection model. For our final model, we chose the 60 most crucial features out of a total of 331 features after filtering out many features to prevent overfitting and speed up learning. On the clickbait class, Random Forest Regression produced the following findings using these features: MSE=0.035 MSE, Accuracy=0.82, and F1-sore=0.61.

2.7 Clickbait Detection in Tweets Using Self-Attentive Network

A model which is capable of evaluating each tweet's level of click baiting. We first reformat the regression problem as a multi-classification problem, based on the annotation scheme. To perform multi-classification, we apply a token-level, self-attentive mechanism on the hidden states of bi-directional Gated Recurrent Units (biGRU), which enables the model to generate tweets' task-specific vector representations by attending to important tokens. The self-attentive neural network can be trained end-to-end, without involving any manual feature engineering.

2.8 Towards a Regression Model for Clickbait Strength

Malicious content publishers misuse social media to manipulate as many users as possible to visit their websites using clickbait messages. Machine learning technology may help to handle this problem, giving rise to automatic clickbait detection. To accelerate progress in this direction, they organized the Clickbait Challenge 2017, a shared task inviting the submission of clickbait detectors for a comparative evaluation. A total of 13 detectors have been submitted, achieving significant improvements over the previous state of the art in terms of detection performance. Also, many of the submitted approaches have been published open source, rendering them reproducible, and a good starting point for newcomers.

2.9 Detecting Clickbait in Online Social Media

This paper proposes a machine learning approach to detect clickbait posts published in social media. Clickbait posts are short, catchy phrases pointing into a longer online article. Users are encouraged to click on these posts to read the full article in many cases. The suggested approach differentiates between clickbait and legitimate posts based on training mainstream machine learning (ML) classifiers.

2.10 Detecting and preventing clickbaits in online news media

In this work, they attempt to automatically detect clickbait videos and then build a browser extension which warns the readers of different media sites about the possibility of being baited by such headlines. The extension also offers each reader an option to block clickbaits she doesn't want to see. Then, using such reader choices, the extension automatically blocks similar clickbaits during her future visits.

CHAPTER 3

METHODOLOGY

3.1. Data Collection

Our dataset consists of 500 videos (Clickbait, Non-Clickbait) gathered from YouTube. In this section, it is explained how clickbait videos are causing a problem for the audience on YouTube. Then, it is explained how the platform and publisher selection is made for the dataset, and the necessary details about these publishers are presented. Finally, it is explained how data is extracted from this platform and pre-processed for the analysis.

3.1.1. Platform Selection

YouTube is the leading platform that is plagued with videos that don’t faithfully represent the situation that it refers to. The user has to add a title, a description, and a thumbnail before uploading a video. These data become crucial parameters on which the users can base their decision to watch a video or not. Many YouTube content creators use clickbait titles and thumbnails that might deviate from the actual content to increase viewership for a video, and generate more revenue. The freedom in creating content gives an incentive for people to post clickbait videos, in which the content might deviate significantly from the title, description, or thumbnail.

3.2. Dataset Construction

We built two datasets. Our first dataset consists of details of 500 videos (Clickbait, Non-Clickbait) gathered from YouTube. The data was collected using a program which extracts videos based on certain keywords. The Google Sheet programming environment was used to code the program that extracts this video details. The keywords we used were top clickbait words like ‘10 Reasons Why’, ‘Wow’, ‘Amazing’, ‘No Way’ etc. The program uses YouTube Data API v3 as a Library and uses it to scrape and analyse videos relating to this query. We extracted 30 first hit videos of each clickbait keyword. The extracted video details include: Title, YouTube Link, Channel ID, Description, Tags, View Count, Like Count, Favourite Count and Comment Count. To get certain details like Video Length, Category ID and Channel Verification we used the website mattw.io which extracts every Metadata detail of the a given video.

The second database consists of all the comments of the videos which are present in the Metadata Dataset. The Comments of these YouTube Videos were extracted using the YouTube Data API v3 using the Google Sheet Programming platform. The program accepts the Video ID as input and it outputs all the Comments extracted from the video into the Google Sheet. The extracted details include: Video ID, Comment, number of likes, number of replies. Extracted comments of 500 Videos which consists of nearly 10 lakhs of rows of data.

3.2.1. Dataset Validation

In order to evaluate the accuracy of labelling the headlines in the dataset as clickbait and non-clickbait, we examined the video thumbnail and the video content. This information coming from humans is essential because it was used for validating the labels of the dataset and validating the results of the machine learning models. A column was constructed labelled as Clickbait/Non-Clickbait in the Metadata Dataset.

We also found new factors for classification View Like ratio, View Comment Ratio using the Like, View and Comment columns. We found that for majority of the videos if the View Like Ratio was above a value of 126 then the chance that the video is a clickbait was very high. This was used as another column of Clickbait/Non-Clickbait classification using Metadata details.

3.3. Data Analysis

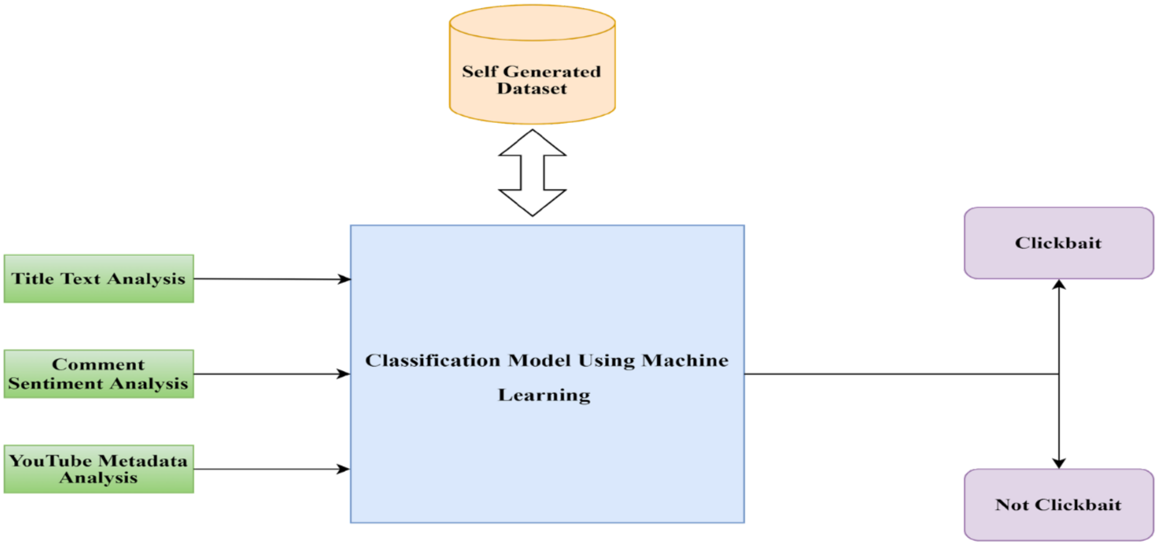

Self Generated Dataset
Title Text Analysis
Comment Sentiment Analysis
YouTube Metadata Analysis
Classification Model Using Machine Learning
Clickbait
Not Clickbait

Fig. 3.1

The model we constructs uses three factors for classification of the video as either a Clickbait or a Non- Clickbait. The first factor is the Title Text Analysis which includes scanning and analysing the title text of the YouTube Video.

The second factor is Comment Sentiment Analysis. It includes Sentiment Analysis of every video comments and coming to a conclusion of whether a video comment is mostly positive negative or neutral.

The third factor is the Metadata Details. Metadata is a set of data that describes and gives information about other data. It includes values like View count, Like count and Comment Count. These were analysed and different conclusions were extracted.

3.3.1. Feature Selection

Due to the removal of Dislike count from YouTube videos, it is required to develop a model based on other features. Our aim is to find some of the other features which can effectively detect clickbait videos on YouTube.

3.3.1.1 Title Text Analysis

The title of the videos was subjected to text analysis and categorised based on misleading titles. Developed a classification model using Naive Bayes Algorithm and LSTM and Trained using self-built Dataset which classifies videos based on the title. In the pre-processing stage the title of the YouTube video was converted to tokens. The data is tokenized i.e., split into tokens which are the smallest or minimal meaningful units. The data is split into words. This is then converted into lowercase to avoid ambiguity between same words in different cases like 'NLP', 'nlp' or 'Nlp'. The punctuations are removed to increase the efficiency of the model. They are irrelevant because they provide no added information. Lemmatization in linguistics is the process of grouping together the inflected forms of a word so they can be analysed as a single item, identified by the word's lemma, or dictionary form. It involves the morphological analysis of words. In lemmatization we find the root word or base form of the word rather than just clipping some characters from the end e.g., *is, are, am* are all converted to its base form *be* in Lemmatization. Here lemmatization is done using NLTK library. Term frequency-Inverse Data Frequency method is used to convert the text into features. The pre-processed data is then trained using Naive Bayes and LSTM and the accuracy measures were calculated.

3.3.1.2 YouTube Metadata Analysis

By using view count and like count we derived a parameter called like-view ratio and developed a hypothesis that a clickbait video should have a high like view ratio than a non-clickbait video. Using this hypothesis, we classified videos as clickbait/non-clickbait. We were able to assess its performance by calculating Accuracy, Precision, F1 Score and Recall using Naïve Bayes, K Nearest Neighbour and Decision Tree Algorithm. Different attributes used were Like Count, View Count, Category ID and Comment Count.

3.3.1.3 Comment Sentiment Analysis

Finally, comments of each and every single video in the dataset were extracted using YouTube Data API v3 and we developed an algorithm for Sentiment Analysis which evaluates each comment and classifies them into positive, negative and neutral comments. The algorithm uses the python library TextBlob. It has an inbuilt function that does sentiment analysis on a given input. The program accepts two datasets, one containing video and the other containing comments of the videos, and produces an excel sheet containing the number of positive, negative and neutral comment count of each video. We developed a hypothesis that a video with more positive comments than negative comments are more likely to be not clickbait.

3.3.2. Performance Analysis

In addition to developing machine learning models that can successfully distinguish clickbait sentences from non-clickbait sentences, it is also critical to understand how these models decide and their confidence in their decisions. The explainability of models is essential for understanding the mechanism of the clickbait strategy and its linguistic structure. Examining the results of the models by discretization does not provide an accurate assessment of the performance of the models. In this way, it is not possible to observe which sentences the models call clickbait or non-clickbait with what degree of Certainty.

CHAPTER 4

RESULTS AND DISCUSSIONS

4.1. Title Text Analysis

4.1.1 Naïve Bayes

The Naïve Bayes was an algorithm used on the 500-video dataset in this study. We used 100 videos at first and got an accuracy of 80% and upon increasing the number of videos by 100, we saw a slight change in accuracy. This model performs with a mean accuracy of 69.28%.

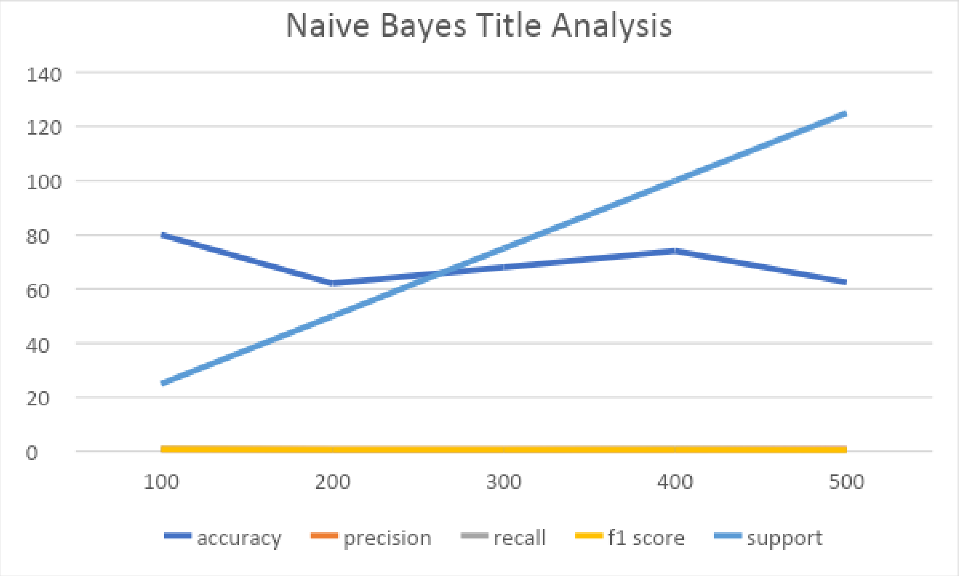


Fig. 4.1

4.1.2 Long Short-Term Memory

The LSTM was developed using a Deep Learning library, Keras. This model was trained with the 500-video dataset as in the previous algorithms, and it performs with a mean accuracy of 66.25%. The performance of the LSTM can be Found in Figure.

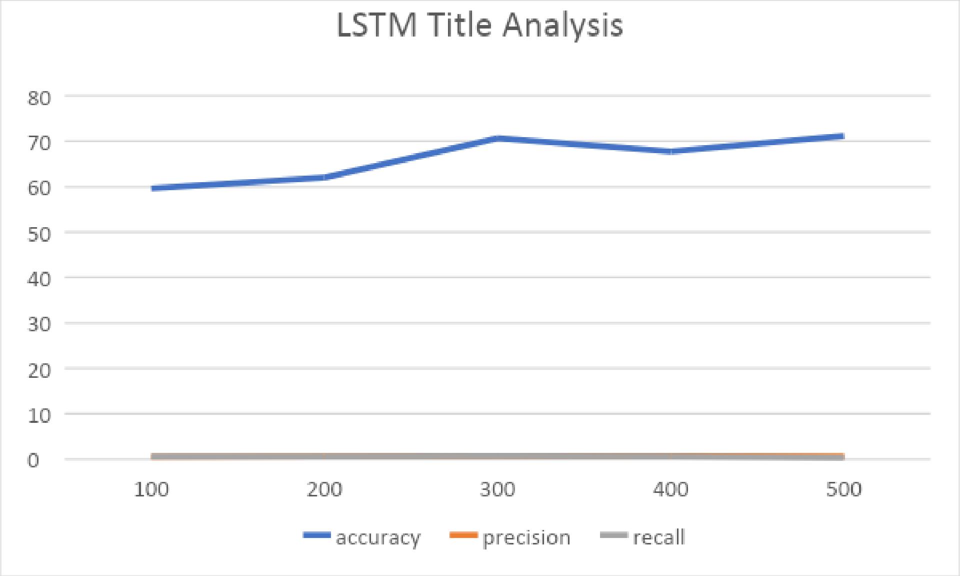


Fig. 4.2

4.2. YouTube Metadata Analysis

4.2.1 Naïve Bayes

The Naïve Bayes was an algorithm used on the 500-video dataset in this study. According to feature importance of this model (Figure 4.3), the number of views, the number of likes, number of comments and category id are distinctive for clickbait detection. We used 100 videos at first and used only the number of views and got an accuracy of 55% and when we combined view count with like count, we got an accuracy of 57.49%. Then we combined view, like and comment count and got an accuracy of 50% and finally combined all parameters and got an accuracy of 60%. Then we increased the number of videos by 100 and obtained the following results.

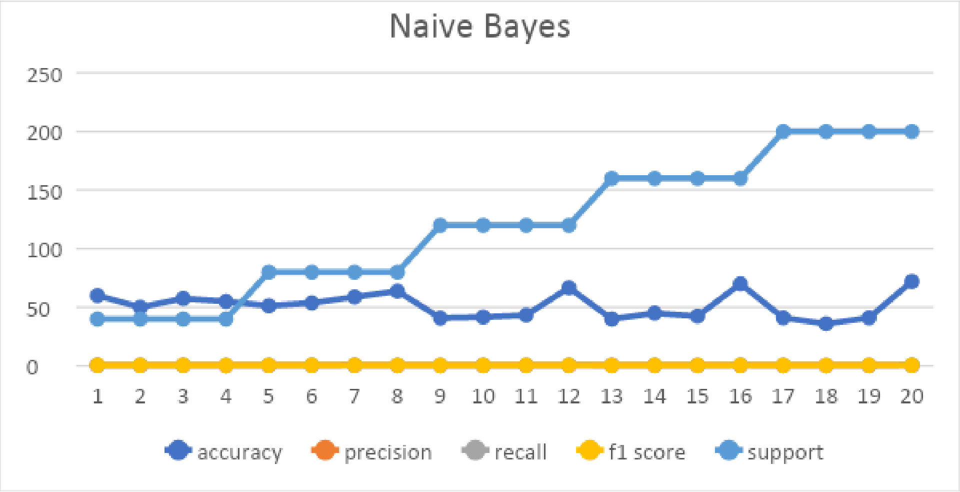


Fig 4.3

4.2.2 Decision Tree

The Decision Tree was an algorithm used on the 500-video dataset in this study. According to feature importance of this model (Figure 4.4), the number of views, the number of likes, number of comments and category id are distinctive for clickbait detection. We used 100 videos at first and used only the number of views and got an accuracy of 50% and when we combined view count with like count, we got an accuracy of 45%. Then we combined view, like and comment count and got an accuracy of 55% and finally combined all parameters and got an accuracy of 60%. Then we increased the number of videos by 100 and obtained the following results.

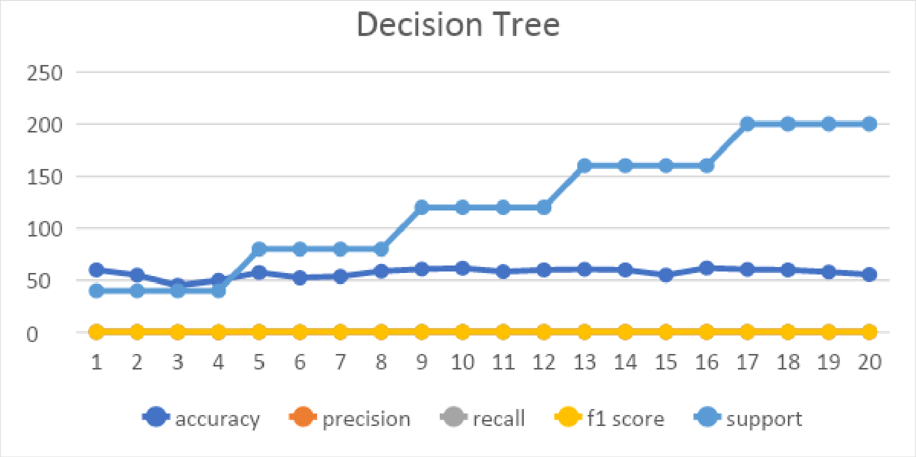


Fig. 4.4

4.2.3 K-Nearest Neighbour

The Naïve Bayes was an algorithm used on the 500-video dataset in this study. According to feature importance of this model (Figure 4.5), the number of views, the number of likes, number of comments and category id are distinctive for clickbait detection. We used 100 videos at first and used only the number of views and got an accuracy of 55% and when we combined view count with like count, we got an accuracy of 52.5%. Then we combined view, like and comment count and got an accuracy of 50% and finally combined all parameters and got an accuracy of 47.55%. Then we increased the number of videos by 100 and obtained the following results.

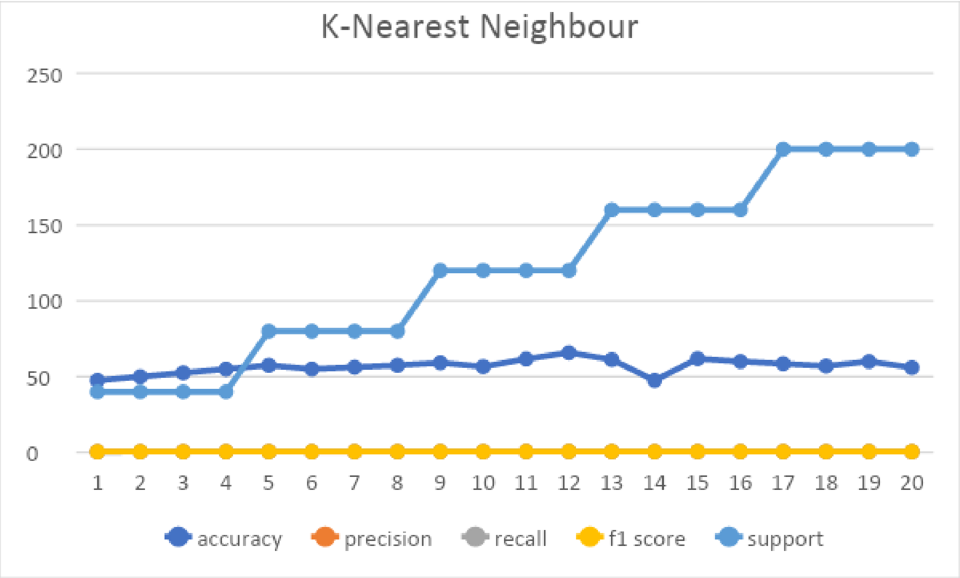


Fig. 4.5

Algorithm Used	Parameters Used	Precision	Recall	F1 Score	Accuracy
Naïve Bayes	View	0.48	0.69	0.57	69.37%
	View, Comment	0.49	0.70	0.58	70.0%
	View, Comment, Category	0.50	0.71	0.58	70.62%
	View, Like	0.57	0.68	0.60	65.17%
	View, Comment, Like, Category	0.42	0.65	0.51	65.0%
Decision Tree	View	0.62	0.62	0.62	62.03%
	View, Like	0.63	0.64	0.63	63.8%
	View, Comment, Category	0.64	0.66	0.65	65.74%
	View, Like, Comment, Category	0.60	0.61	0.60	61.11%
K- Nearest Neighbour	View	0.49	0.51	0.47	50.92%
	View, Like	0.53	0.54	0.53	53.70%
	View, Comment, Category	0.60	0.62	0.61	62.03%
	View, Comment, Like, Category	0.64	0.66	0.65	53.70%

Table 4.1

4.3. Comment Sentiment Analysis

4.3.1. Naïve Bayes

Comment Sentiment Analysis, also known as opinion mining is a powerful tool that can easily be put to work on a large database to automatically understand the emotions behind the comments posted by users. We used Naïve Bayes to perform Comment Sentiment Analysis. Each comment of the videos is evaluated and are classified as negative, positive or neutral on the basis of the certain keywords to analyse the emotion of user behind the comment using an algorithm. The algorithm uses Python Library TextBlob. Two datasets are accepted by the algorithm, one containing video and the other containing comments of the videos. The output will be an excel sheet containing the number of positive, negative and neutral videos. Classification of clickbait videos is carried out based on the assumption that, If the number of Negative comments is more than the number of Positive comments, the video is possibly a clickbait video. The performance of the algorithm is assessed by using the parameters Accuracy, Precision, Recall and F1 Score.

CHAPTER 5

DISCUSSION AND CONCLUSION

5.1 DISCUSSION

With the increase in accessibility of internet, there was a boom in video streaming platforms like YouTube. With the increase in users all around the globe, the platform started flooding with videos, some of which do not faithfully represent what it is supposed to. The content creator has to add a title, a description, and a thumbnail before uploading a video. These data become crucial parameters on which the users can base their decision to watch a video or not. For assessing the performance of each of the algorithms, we chose Accuracy, F1 Score, Precision and Recall as the parameters. These parameters help in determining how much of the retrieved data is accurately predicted as Clickbait or not using features of a YouTube video like View Count, Like Count, Comment Count, Category ID and a combination of all those features. After training and testing various machine learning algorithms, it was found that the accuracy of View Count, Like Count and Category ID was higher than the other parameters in detecting clickbait. There was a steady increase in accuracy with the increase in the entries in database.

5.2 CONCLUSION

The emergence of YouTube as a platform for video streaming created a revolution in the internet era. Soon YouTube was crowned as the best Video Streaming Platform powered by Google. YouTube became so popular that each and every internet user trusted it. One of the reasons for its high demand of usage and its popularity is that it pays the content creators for the videos they create based on its insight, that is its view count and other video parameters. Since it’s an easy source of income people started finding shortcuts without any effort of creating content, one of which is clickbaited videos.

The escalation of Clickbait videos in YouTube has created a sense of deceitfulness in public which might harm the honest contents. Clickbait Detection is imperative to maintain the reliability and quality of the platform. Like-Dislike ratio used to be the most effective feature in detecting clickbait in the past. But as dislike count got hidden, there came the necessity to rely on the other features of the video such as title, comments, like-view, category id etc. Research has been conducted to assess the performance of each feature and to determine the most effective one in detecting whether the video is a clickbait or not. This research thus defines a pathway to a reliable exploration of YouTube platform and thereby refining YouTube as a platform.

5.3 FUTURE WORK

Clickbait Detection is a project of infinite potential. With the content publishers finding new and creative ways of baiting viewers for income, we ought to find more ways and create stronger algorithm for its prevention. Some of the few planned future developments are :-

- Discovering new features to classify videos as clickbait or non-clickbait with an accuracy higher than the other features
- Increase the number of videos in the dataset.
- Use thumbnail for the detection of clickbait using image processing.
- Use comment depth analysis for better accuracy.
- Compare audio and text extracted from the video.

REFERENCES

1. D. Molina, M., Sundar, S. S., Rony, M. M. U., Hassan, N., Le, T., & Lee, D. (2021, May). Does clickbait actually attract more clicks? Three clickbait studies you must read. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1-19).

2. Genç, Ş. (2021). *Turkish clickbait detection in social media via machine learning algorithms* (Master's thesis, Middle East Technical University).

3. Zheng, H. T., Chen, J. Y., Yao, X., Sangaiah, A. K., Jiang, Y., & Zhao, C. Z. (2018). Clickbait convolutional neural network. *Symmetry*, *10*(5), 138.

4. Varshney, D., & Vishwakarma, D. K. (2021). A unified approach for detection of Clickbait videos on YouTube using cognitive evidences. *Applied Intelligence*, *51*(7), 4214-4235.

5. Bourgonje, P., Schneider, J. M., & Rehm, G. (2017, September). From clickbait to fake news detection: an approach based on detecting the stance of headlines to articles. In *Proceedings of the 2017 EMNLP workshop: natural language processing meets journalism* (pp. 84-89).

6. Zannettou, S., Chatzis, S., Papadamou, K., & Sirivianos, M. (2018, May). The good, the bad and the bait: Detecting and characterizing clickbait on youtube. In *2018 IEEE Security and Privacy Workshops (SPW)* (pp. 63-69). IEEE.

7. Qu, J., Hißbach, A. M., Gollub, T., & Potthast, M. (2018). Towards Crowdsourcing Clickbait Labels for YouTube Videos. In *HCOMP (WIP&Demo)*.

8. Potthast, M., Köpsel, S., Stein, B., & Hagen, M. (2016, March). Clickbait detection. In *European Conference on Information Retrieval* (pp. 810-817). Springer, Cham.

9. Agrawal, A. (2016, October). Clickbait detection using deep learning. In *2016 2nd international conference on next generation computing technologies (NGCT)* (pp. 268-272). IEEE.

10. Shang, L., Zhang, D. Y., Wang, M., Lai, S., & Wang, D. (2019). Towards reliable online clickbait video detection: A content-agnostic approach. *Knowledge-Based Systems*, *182*, 104851.

11. Grigorev, A. (2017). Identifying clickbait posts on social media with an ensemble of linear models. *arXiv preprint arXiv:1710.00399*.

12. Cao, X., & Le, T. (2017). Machine learning based detection of clickbait posts in social media. *arXiv preprint arXiv:1710.01977*.

13. Zhou, Y. (2017). Clickbait detection in tweets using self-attentive network. *arXiv preprint arXiv:1710.05364*.

14. Potthast, M., Gollub, T., Hagen, M., & Stein, B. (2018). The clickbait challenge 2017: Towards a regression model for clickbait strength. *arXiv preprint arXiv:1812.10847*.

15. Gothankar, R., Di Troia, F., & Stamp, M. (2021). Clickbait Detection in YouTube Videos. *arXiv preprint arXiv:2107.12791*.

16. Kaothanthong, N., Kongyoung, S., & Theeramunkong, T. (2021). Headline2Vec: A CNN-based Feature for Thai Clickbait Headlines Classification. *INTERNATIONAL SCIENTIFIC JOURNAL OF ENGINEERING AND TECHNOLOGY (ISJET)*, *5*(1), 20-31.

17. Hansrajh, A., Adeliyi, T. T., & Wing, J. (2021). Detection of Online Fake News Using Blending Ensemble Learning. *Scientific Programming*, *2021*.

18. Lischka, J. A., & Garz, M. (2021). Clickbait news and algorithmic curation: A game theory framework of the relation between journalism, users, and platforms. *New Media & Society*, 14614448211027174.

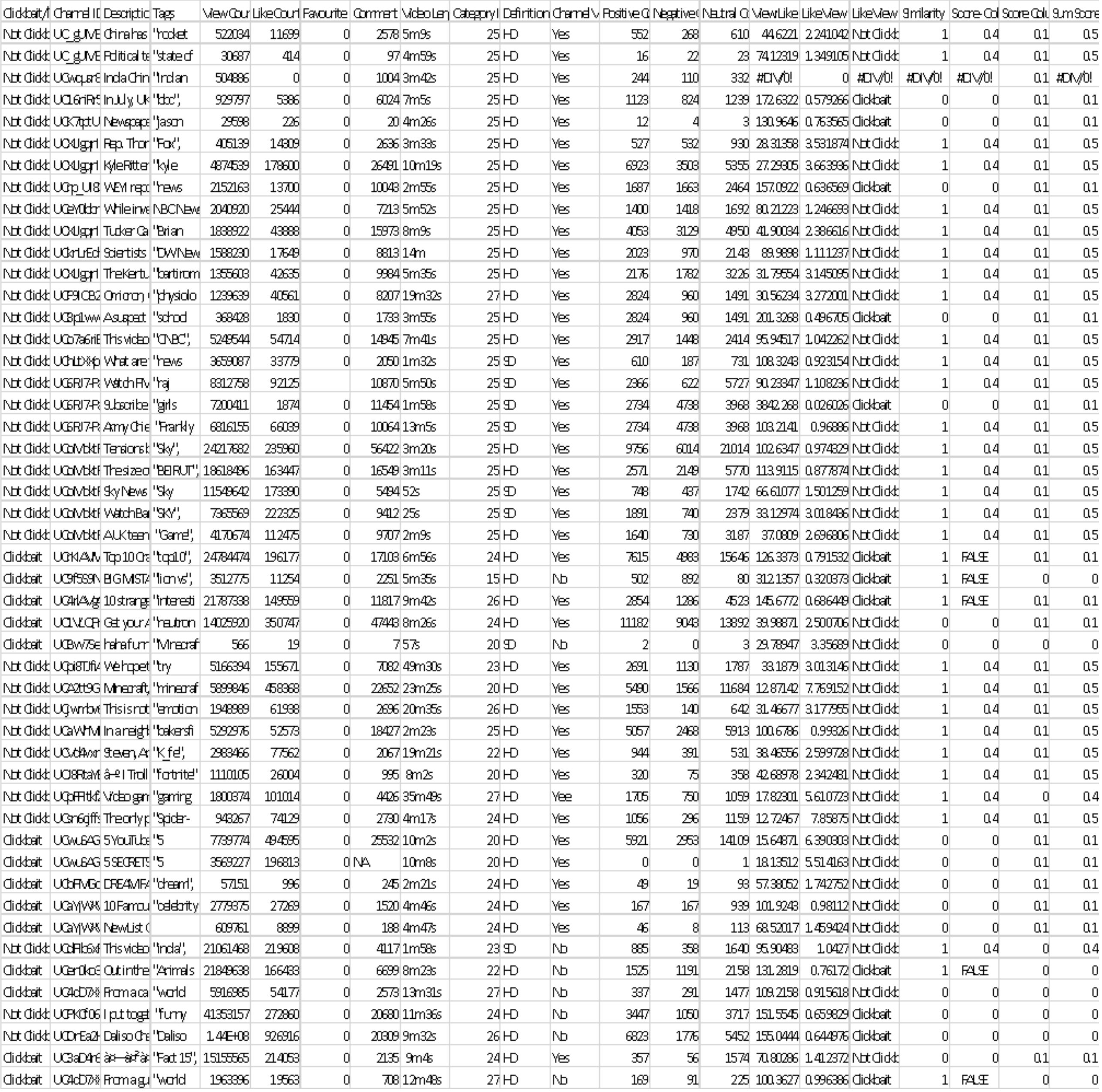
19. Shu, K., Wang, S., Le, T., Lee, D., & Liu, H. (2018, November). Deep headline generation for clickbait detection. In *2018 IEEE International Conference on Data Mining (ICDM)* (pp. 467-476). IEEE.

APPENDICES

APPENDIX A

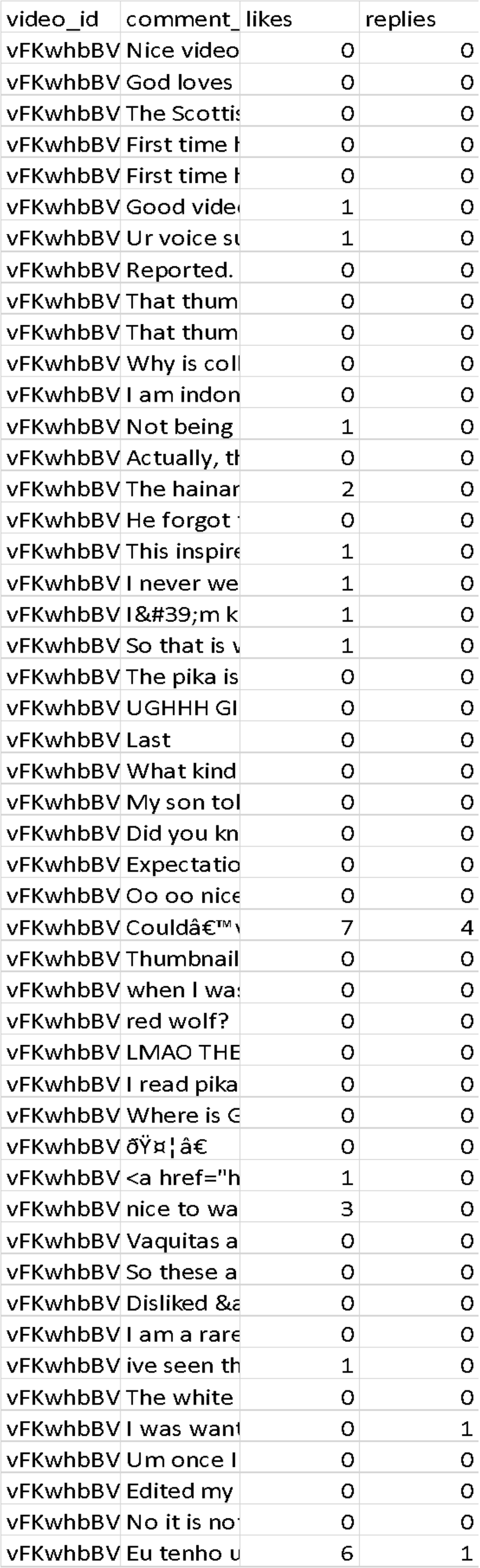
THE ORGANIZATION OF THE DATASET

Appendix A.1. The organization of the dataset. A view of the clickbait data from the Self-generated dataset.



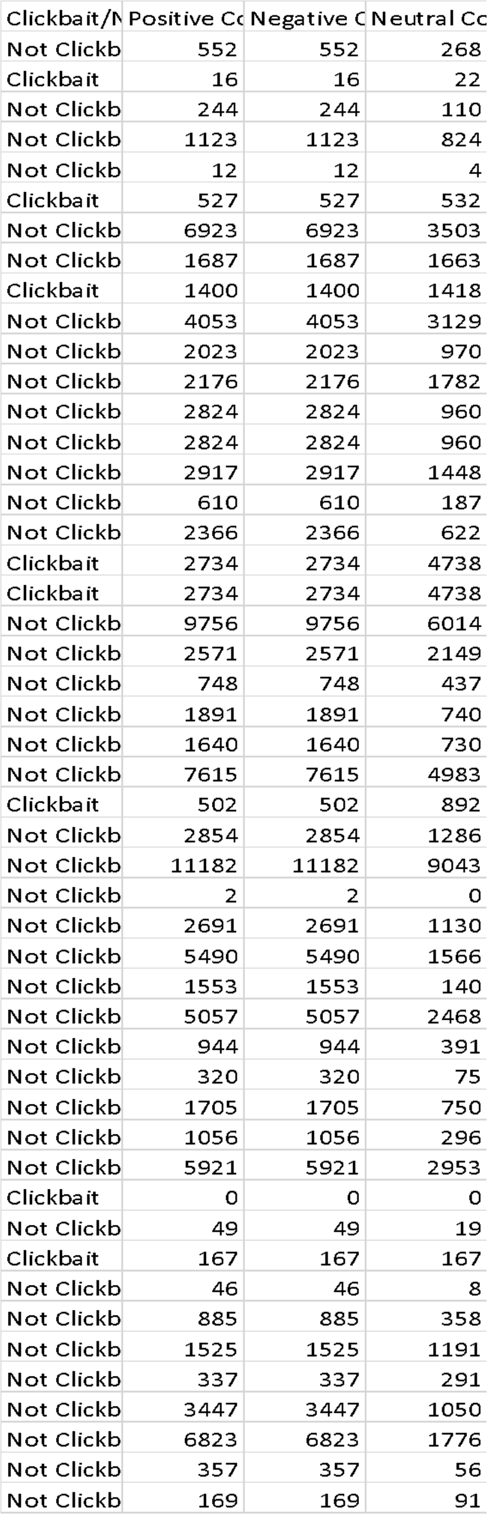
Clickbait/I	Channel ID	Descriptio	Tags	ViewCour	LikeCount	Favourite	Comment	VideoLen	Category I	Definition	Channel V	Positive C	Negative C	Neutral C	ViewLike	LikeView	LikeView	Similarity	Score-Col	Score Colu	Sum Score
Not Clickb	UC_gUME	China has	"rocket	522034	11699	0	2578	5m9s	25	HD	Yes	552	268	610	44.6221	2.241042	Not Clickb	1	0.4	0.1	0.5
Not Clickb	UC_gUME	Political te	"state of	30687	414	0	97	4m59s	25	HD	Yes	16	22	23	74.12319	1.349105	Not Clickb	1	0.4	0.1	0.5
Not Clickb	UCwqusr8	India Chin	"Indian	504886	0	0	1004	3m42s	25	HD	Yes	244	110	332	#DIV/0!	0	#DIV/0!	#DIV/0!	#DIV/0!	0.1	#DIV/0!
Not Clickb	UC16niRr5	In July, Uk	"bbc",	929797	5386	0	6024	7m5s	25	HD	Yes	1123	824	1239	172.6322	0.579266	Clickbait	0	0	0.1	0.1
Not Clickb	UCK7qtU	Newspape	"jason	29598	226	0	20	4m26s	25	HD	Yes	12	4	3	130.9646	0.763565	Clickbait	0	0	0.1	0.1
Not Clickb	UCXIJgqn1	Rep. Thor	"Fox",	405139	14309	0	2636	3m33s	25	HD	Yes	527	532	930	28.31358	3.531874	Not Clickb	1	0.4	0.1	0.5
Not Clickb	UCXIJgqn1	Kyle Ritter	"kyle	4874539	178600	0	26491	10m19s	25	HD	Yes	6923	3503	5355	27.29305	3.663936	Not Clickb	1	0.4	0.1	0.5
Not Clickb	UCrp_UI8	WEM repo	"news	2152163	13700	0	10043	2m55s	25	HD	Yes	1687	1663	2464	157.0922	0.636569	Clickbait	0	0	0.1	0.1
Not Clickb	UCeY0bbn	While inve	NBCNews	2040920	25444	0	7213	5m52s	25	HD	Yes	1400	1418	1692	80.21223	1.246693	Not Clickb	1	0.4	0.1	0.5
Not Clickb	UCXIJgqn1	Tucker Ca	"Brian	1838922	43888	0	15973	8m9s	25	HD	Yes	4053	3129	4950	41.90034	2.386616	Not Clickb	1	0.4	0.1	0.5
Not Clickb	UCknLrEd	Scientists	"DWNew	1588230	17649	0	8813	14m	25	HD	Yes	2023	970	2143	89.9898	1.111237	Not Clickb	1	0.4	0.1	0.5
Not Clickb	UCXIJgqn1	The Kentu	"bartirom	1355603	42635	0	9984	5m35s	25	HD	Yes	2176	1782	3226	31.79554	3.145095	Not Clickb	1	0.4	0.1	0.5
Not Clickb	UCP9ICB2	Omicron	"physiolo	1239639	40561	0	8207	19m32s	27	HD	Yes	2824	960	1491	30.56234	3.272001	Not Clickb	1	0.4	0.1	0.5
Not Clickb	UCBp1ww	A suspect	"school	368428	1830	0	1733	3m55s	25	HD	Yes	2824	960	1491	201.3268	0.496705	Clickbait	0	0	0.1	0.1
Not Clickb	UCo7a6riE	This video	"CNBC",	5249544	54714	0	14945	7m41s	25	HD	Yes	2917	1448	2414	95.94517	1.042262	Not Clickb	1	0.4	0.1	0.5
Not Clickb	UChLtXKp	What are	"news	3659087	33779	0	2050	1m32s	25	SD	Yes	610	187	731	108.3243	0.923154	Not Clickb	1	0.4	0.1	0.5
Not Clickb	UCGRJ7-R	Watch PV	"raj	8312758	92125		10870	5m50s	25	SD	Yes	2366	622	5727	90.23347	1.108236	Not Clickb	1	0.4	0.1	0.5
Not Clickb	UCGRJ7-R	Subscribe	"girls	7200411	1874	0	11454	1m58s	25	SD	Yes	2734	4738	3968	3842.268	0.026026	Clickbait	0	0	0.1	0.1
Not Clickb	UCGRJ7-R	Army Chie	"Frankly	6816155	66039	0	10064	13m5s	25	SD	Yes	2734	4738	3968	103.2141	0.96886	Not Clickb	1	0.4	0.1	0.5
Not Clickb	UCbMbktf	Tensions b	"Sky",	24217682	235960	0	56422	3m20s	25	HD	Yes	9756	6014	21014	102.6347	0.974329	Not Clickb	1	0.4	0.1	0.5
Not Clickb	UCbMbktf	The size o	"BEIRUT",	18618496	163447	0	16549	3m11s	25	HD	Yes	2571	2149	5770	113.9115	0.877874	Not Clickb	1	0.4	0.1	0.5
Not Clickb	UCbMbktf	Sky News	"Sky	11549642	173390	0	5494	52s	25	SD	Yes	748	437	1742	66.61077	1.501259	Not Clickb	1	0.4	0.1	0.5
Not Clickb	UCbMbktf	Watch Ba	"SKY",	7365569	222325	0	9412	25s	25	SD	Yes	1891	740	2379	33.12974	3.018436	Not Clickb	1	0.4	0.1	0.5
Not Clickb	UCbMbktf	A UK teen	"Game",	4170674	112475	0	9707	2m9s	25	HD	Yes	1640	730	3187	37.0809	2.696806	Not Clickb	1	0.4	0.1	0.5
Clickbait	UCK1AWv	Top 10 Cra	"top10",	24784474	196177	0	17103	6m56s	24	HD	Yes	7615	4983	15646	126.3373	0.791532	Clickbait	1	FALSE	0.1	0.1
Clickbait	UC9f5S9N	BIG MISTA	"lion vs",	3512775	11254	0	2251	5m35s	15	HD	No	502	892	80	312.1357	0.320373	Clickbait	1	FALSE	0	0
Clickbait	UC4rlAvg	10 strange	"interesti	21787338	149559	0	11817	9m42s	26	HD	Yes	2854	1286	4523	145.6772	0.686449	Clickbait	1	FALSE	0.1	0.1
Clickbait	UC1VLCR	Get your A	"neutron	14025920	350747	0	47443	8m26s	24	HD	Yes	11182	9043	13892	39.98871	2.500706	Not Clickb	0	0	0.1	0.1
Clickbait	UCBw7Se	haha fun	"Minecraf	566	19	0	7	57s	20	SD	No	2	0	3	29.78947	3.35689	Not Clickb	0	0	0	0
Not Clickb	UCpi8TJfi	We hope t	"try	5166394	155671	0	7082	49m30s	23	HD	Yes	2691	1130	1787	33.1879	3.013146	Not Clickb	1	0.4	0.1	0.5
Not Clickb	UCA2tt9G	Minecraft,	"minecraf	5899846	458368	0	22652	23m25s	20	HD	Yes	5490	1566	11684	12.87142	7.769152	Not Clickb	1	0.4	0.1	0.5
Not Clickb	UCjwmbw	This is not	"emotion	1948989	61938	0	2696	20m35s	26	HD	Yes	1553	140	642	31.46677	3.177955	Not Clickb	1	0.4	0.1	0.5
Not Clickb	UCaWHM	In a neigh	"bakersfi	5292976	52573	0	18427	2m23s	25	HD	Yes	5057	2468	5913	100.6786	0.99326	Not Clickb	1	0.4	0.1	0.5
Not Clickb	UCvd4wr	Steven, Ar	"K_fe",	2983466	77562	0	2067	19m21s	22	HD	Yes	944	391	531	38.46556	2.599728	Not Clickb	1	0.4	0.1	0.5
Not Clickb	UCI8RtaM	â–º I Troll	"fortnite"	1110105	26004	0	995	8m2s	20	HD	Yes	320	75	358	42.68978	2.342481	Not Clickb	1	0.4	0.1	0.5
Not Clickb	UCpFFtkf	Video gam	"gaming	1800374	101014	0	4426	35m49s	27	HD	Yee	1705	750	1059	17.82301	5.610723	Not Clickb	1	0.4	0	0.4
Not Clickb	UCsn6gjffs	The only p	"Spider-	943267	74129	0	2730	4m17s	24	HD	Yes	1056	296	1159	12.72467	7.85875	Not Clickb	1	0.4	0.1	0.5
Clickbait	UCwu6AG	5 YouTube	"5	7739774	494595	0	25532	10m2s	20	HD	Yes	5921	2953	14109	15.64871	6.390303	Not Clickb	0	0	0.1	0.1
Clickbait	UCwu6AG	5 SECRETS	"5	3569227	196813	0	NA	10m8s	20	HD	Yes	0	0	1	18.13512	5.514163	Not Clickb	0	0	0.1	0.1
Clickbait	UCbFVGc	DREAMFA	"dream",	57151	996	0	245	2m21s	24	HD	Yes	49	19	93	57.38052	1.742752	Not Clickb	0	0	0.1	0.1
Clickbait	UCaYjWA	10 Famou	"celebrity	2779375	27269	0	1520	4m46s	24	HD	Yes	167	167	939	101.9243	0.98112	Not Clickb	0	0	0.1	0.1
Clickbait	UCaYjWA	New List (		609761	8899	0	188	4m47s	24	HD	Yes	46	8	113	68.52017	1.459424	Not Clickb	0	0	0.1	0.1
Not Clickb	UCdRb6x	This video	"india",	21061468	219608	0	4117	1m58s	23	SD	No	885	358	1640	95.90483	1.0427	Not Clickb	1	0.4	0	0.4
Clickbait	UCer0ko3	Out in the	"Animals	21849638	166433	0	6699	8m23s	22	HD	No	1525	1191	2158	131.2819	0.76172	Clickbait	1	FALSE	0	0
Clickbait	UC4cD7X	From a ca	"world	5916985	54177	0	2573	13m31s	27	HD	No	337	291	1477	109.2158	0.915618	Not Clickb	0	0	0	0
Not Clickb	UCPKOf06	I put toget	"funny	41353157	272860	0	20680	11m36s	24	HD	No	3447	1050	3717	151.5545	0.659829	Clickbait	0	0	0	0
Not Clickb	UCDrEa2	Daliso Cha	"Daliso	1.44E+08	926916	0	20309	9m32s	26	HD	No	6823	1776	5452	155.0444	0.644976	Clickbait	0	0	0	0
Clickbait	UC3aD4r6	à¤—à¤²à	"Fact 15",	15155565	214053	0	2135	9m4s	24	HD	Yes	357	56	1574	70.80286	1.412372	Not Clickb	0	0	0.1	0.1
Clickbait	UC4cD7X	From a gu	"world	1963396	19563	0	708	12m48s	27	HD	No	169	91	225	100.3627	0.996386	Clickbait	1	FALSE	0	0

Appendix A.2. The organization of the dataset. A view of the comment data from the comment dataset.



video_id	comment_	likes	replies
vFKwhbBV	Nice video	0	0
vFKwhbBV	God loves	0	0
vFKwhbBV	The Scottis	0	0
vFKwhbBV	First time I	0	0
vFKwhbBV	First time I	0	0
vFKwhbBV	Good vide	1	0
vFKwhbBV	Ur voice su	1	0
vFKwhbBV	Reported.	0	0
vFKwhbBV	That thum	0	0
vFKwhbBV	That thum	0	0
vFKwhbBV	Why is coll	0	0
vFKwhbBV	I am indon	0	0
vFKwhbBV	Not being	1	0
vFKwhbBV	Actually, th	0	0
vFKwhbBV	The hainar	2	0
vFKwhbBV	He forgot	0	0
vFKwhbBV	This inspire	1	0
vFKwhbBV	I never we	1	0
vFKwhbBV	I&#39;m k	1	0
vFKwhbBV	So that is v	1	0
vFKwhbBV	The pika is	0	0
vFKwhbBV	UGHHH GI	0	0
vFKwhbBV	Last	0	0
vFKwhbBV	What kind	0	0
vFKwhbBV	My son tol	0	0
vFKwhbBV	Did you kn	0	0
vFKwhbBV	Expectatio	0	0
vFKwhbBV	Oo oo nice	0	0
vFKwhbBV	Couldâ€™v	7	4
vFKwhbBV	Thumbnail	0	0
vFKwhbBV	when I wa	0	0
vFKwhbBV	red wolf?	0	0
vFKwhbBV	LMAO THE	0	0
vFKwhbBV	I read pika	0	0
vFKwhbBV	Where is G	0	0
vFKwhbBV	ðŸ¤¦â€	0	0
vFKwhbBV	<a href="h	1	0
vFKwhbBV	nice to wa	3	0
vFKwhbBV	Vaquitas a	0	0
vFKwhbBV	So these a	0	0
vFKwhbBV	Disliked &a	0	0
vFKwhbBV	I am a rare	0	0
vFKwhbBV	ive seen th	1	0
vFKwhbBV	The white	0	0
vFKwhbBV	I was want	0	1
vFKwhbBV	Um once I	0	0
vFKwhbBV	Edited my	0	0
vFKwhbBV	No it is no	0	0
vFKwhbBV	Eu tenho u	6	1

Appendix A.3. The organization of the dataset. A view of the clickbait data from the sentiment analysis.

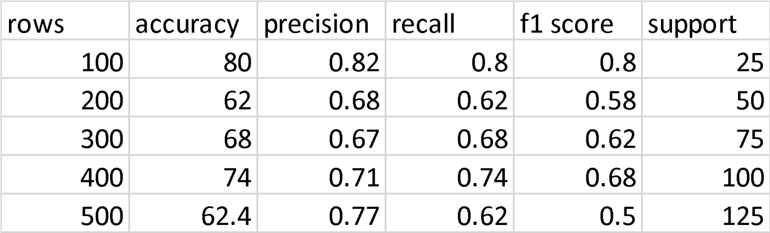


Clickbait/N	Positive Cc	Negative C	Neutral Cc
Not Clickb	552	552	268
Clickbait	16	16	22
Not Clickb	244	244	110
Not Clickb	1123	1123	824
Not Clickb	12	12	4
Clickbait	527	527	532
Not Clickb	6923	6923	3503
Not Clickb	1687	1687	1663
Clickbait	1400	1400	1418
Not Clickb	4053	4053	3129
Not Clickb	2023	2023	970
Not Clickb	2176	2176	1782
Not Clickb	2824	2824	960
Not Clickb	2824	2824	960
Not Clickb	2917	2917	1448
Not Clickb	610	610	187
Not Clickb	2366	2366	622
Clickbait	2734	2734	4738
Clickbait	2734	2734	4738
Not Clickb	9756	9756	6014
Not Clickb	2571	2571	2149
Not Clickb	748	748	437
Not Clickb	1891	1891	740
Not Clickb	1640	1640	730
Not Clickb	7615	7615	4983
Clickbait	502	502	892
Not Clickb	2854	2854	1286
Not Clickb	11182	11182	9043
Not Clickb	2	2	0
Not Clickb	2691	2691	1130
Not Clickb	5490	5490	1566
Not Clickb	1553	1553	140
Not Clickb	5057	5057	2468
Not Clickb	944	944	391
Not Clickb	320	320	75
Not Clickb	1705	1705	750
Not Clickb	1056	1056	296
Not Clickb	5921	5921	2953
Clickbait	0	0	0
Not Clickb	49	49	19
Clickbait	167	167	167
Not Clickb	46	46	8
Not Clickb	885	885	358
Not Clickb	1525	1525	1191
Not Clickb	337	337	291
Not Clickb	3447	3447	1050
Not Clickb	6823	6823	1776
Not Clickb	357	357	56
Not Clickb	169	169	91

APPENDIX B

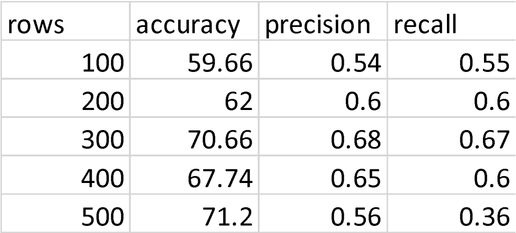
THE PERFORMANCE-BASED FEATURES OF THE MODELS

Appendix B.1. performance based important features of the Naïve Bayes model using title of the video.



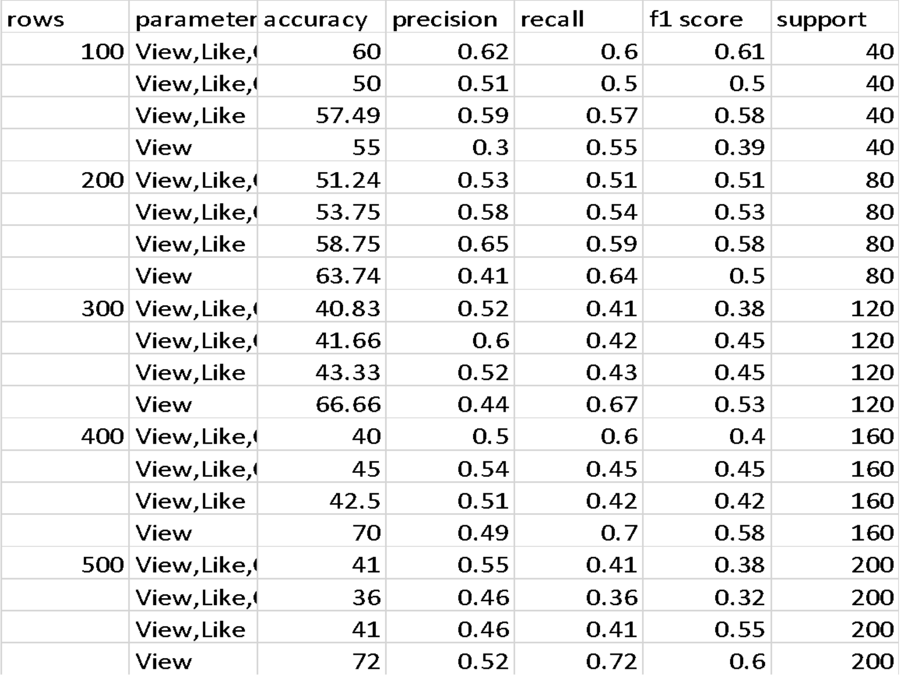
rows	accuracy	precision	recall	f1 score	support
100	80	0.82	0.8	0.8	25
200	62	0.68	0.62	0.58	50
300	68	0.67	0.68	0.62	75
400	74	0.71	0.74	0.68	100
500	62.4	0.77	0.62	0.5	125

Appendix B.2. performance based important features of the Long Short-Term Memory model using title of the video.



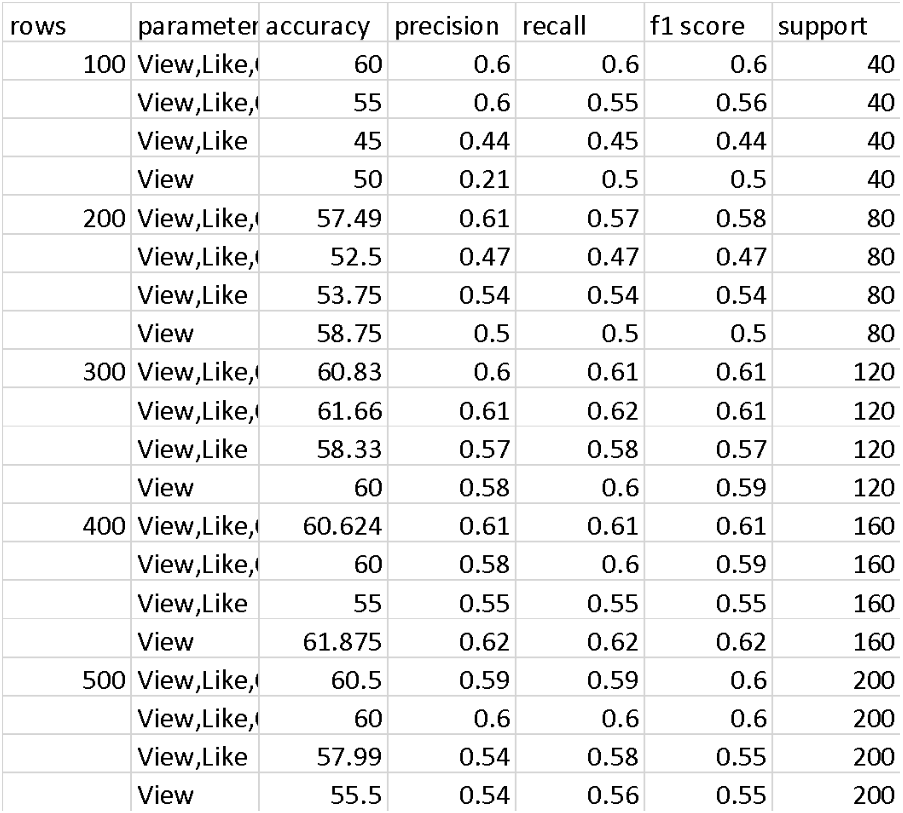
rows	accuracy	precision	recall
100	59.66	0.54	0.55
200	62	0.6	0.6
300	70.66	0.68	0.67
400	67.74	0.65	0.6
500	71.2	0.56	0.36

Appendix B.3. performance based important features of the Naïve Bayes model using video metadata.



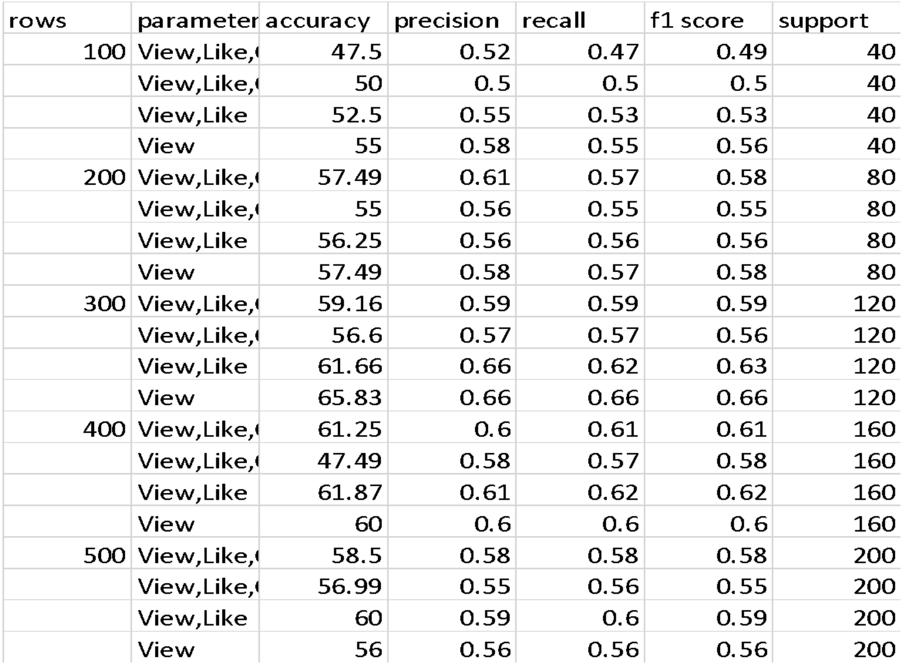
rows	parameter	accuracy	precision	recall	f1 score	support
100	View,Like,(	60	0.62	0.6	0.61	40
	View,Like,(	50	0.51	0.5	0.5	40
	View,Like	57.49	0.59	0.57	0.58	40
	View	55	0.3	0.55	0.39	40
200	View,Like,(	51.24	0.53	0.51	0.51	80
	View,Like,(	53.75	0.58	0.54	0.53	80
	View,Like	58.75	0.65	0.59	0.58	80
	View	63.74	0.41	0.64	0.5	80
300	View,Like,(	40.83	0.52	0.41	0.38	120
	View,Like,(	41.66	0.6	0.42	0.45	120
	View,Like	43.33	0.52	0.43	0.45	120
	View	66.66	0.44	0.67	0.53	120
400	View,Like,(	40	0.5	0.6	0.4	160
	View,Like,(	45	0.54	0.45	0.45	160
	View,Like	42.5	0.51	0.42	0.42	160
	View	70	0.49	0.7	0.58	160
500	View,Like,(	41	0.55	0.41	0.38	200
	View,Like,(	36	0.46	0.36	0.32	200
	View,Like	41	0.46	0.41	0.55	200
	View	72	0.52	0.72	0.6	200

Appendix B.3. performance based important features of the Decision Tree model using video metadata.



rows	parameter	accuracy	precision	recall	f1 score	support
100	View,Like,(	60	0.6	0.6	0.6	40
	View,Like,(	55	0.6	0.55	0.56	40
	View,Like	45	0.44	0.45	0.44	40
	View	50	0.21	0.5	0.5	40
200	View,Like,(	57.49	0.61	0.57	0.58	80
	View,Like,(	52.5	0.47	0.47	0.47	80
	View,Like	53.75	0.54	0.54	0.54	80
	View	58.75	0.5	0.5	0.5	80
300	View,Like,(	60.83	0.6	0.61	0.61	120
	View,Like,(	61.66	0.61	0.62	0.61	120
	View,Like	58.33	0.57	0.58	0.57	120
	View	60	0.58	0.6	0.59	120
400	View,Like,(	60.624	0.61	0.61	0.61	160
	View,Like,(	60	0.58	0.6	0.59	160
	View,Like	55	0.55	0.55	0.55	160
	View	61.875	0.62	0.62	0.62	160
500	View,Like,(	60.5	0.59	0.59	0.6	200
	View,Like,(	60	0.6	0.6	0.6	200
	View,Like	57.99	0.54	0.58	0.55	200
	View	55.5	0.54	0.56	0.55	200

Appendix B.4. performance based important features of the K-nearest Neighbour model using video metadata.



rows	parameter	accuracy	precision	recall	f1 score	support
100	View,Like,(	47.5	0.52	0.47	0.49	40
	View,Like,(	50	0.5	0.5	0.5	40
	View,Like	52.5	0.55	0.53	0.53	40
	View	55	0.58	0.55	0.56	40
200	View,Like,(	57.49	0.61	0.57	0.58	80
	View,Like,(	55	0.56	0.55	0.55	80
	View,Like	56.25	0.56	0.56	0.56	80
	View	57.49	0.58	0.57	0.58	80
300	View,Like,(	59.16	0.59	0.59	0.59	120
	View,Like,(	56.6	0.57	0.57	0.56	120
	View,Like	61.66	0.66	0.62	0.63	120
	View	65.83	0.66	0.66	0.66	120
400	View,Like,(	61.25	0.6	0.61	0.61	160
	View,Like,(	47.49	0.58	0.57	0.58	160
	View,Like	61.87	0.61	0.62	0.62	160
	View	60	0.6	0.6	0.6	160
500	View,Like,(	58.5	0.58	0.58	0.58	200
	View,Like,(	56.99	0.55	0.56	0.55	200
	View,Like	60	0.59	0.6	0.59	200
	View	56	0.56	0.56	0.56	200