IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 7, July 2022

DOI: 10.17148/IJARCCE.2022.11735

Bait Detector: YouTube Video Recommendation

Ankush P Gowda[1], Ananya Alse A R[2], Chethan G S[3], Adarsha Ujjanimatha[4], Santosh E[5]

Students, Maharaja Institute of Technology, Mysore, Karnataka[1-4]

Assistant Professor, Maharaja Institute of Technology, Mysore, Karnataka[5]

Abstract: We attempt to detect the clickbait with our design model. YouTube videos often include captivating descriptions and intriguing thumbnails designed to increase the number of views, and thereby increase the revenue for the person who posted the video. Initially, in the proposed system, we gather data like the audio transcript from YouTube along with Title, Comments, likes, views, and Statistics. we train, pre-process and evaluate the data sets. In Multi-Model Architecture, we apply the SVM algorithm for titles, Comments, likes, and Statistics. According to the output obtained by this algorithm, we classify video as clickbait or not.

Keywords: Scikit-learn (Sklearn), Regular Expression (Regex)

I. INTRODUCTION

Click-baiting is to attract readers attention to click it. Click-baiting is a growing phenomenon on the internet, and it is defined as a method of exploiting cognitive biases to attract online viewership, that is, to attract "clicks." In other words, these headlines contain text which leaves the reader curious about what the article contents might be, or they contain text about topics not really covered in the article itself. In order to avoid readers from clicking it, we attempt to detect the clickbait with our design model. YouTube videos often include captivating descriptions and intriguing thumbnails designed to increase the number of views, and thereby increase the revenue for the person who posted the video. This creates an incentive for people to post clickbait videos, in which the content might deviate significantly from the title, description, or thumbnail. In effect, users are tricked into clicking on clickbait videos. In this project, we consider the challenging problem of detecting clickbait YouTube videos. We experiment with multiple state-of-the-art machine learning techniques using a variety of textual features.

The identification of currency depends on the characteristics of currency notes of a particular country. Due to use for a long time, currency notes may be contaminated by noises. To Identify whether the currency is authentic or not there are many features. Although it may not be practically possible to accurately identify a counterfeit in a paper currency which can only be identified by an intelligent machine.

requires a system that will recognize currency. It has various potential applications that includes banknote counting machines, money exchange machines, assisting blind persons, electronic banking, currency monitoring systems etc. The recognition of currency is a very important need for visually impaired people. They are not being able to differentiate between currencies correctly, so it is very easy for them to be cheated by the others. Therefore, there is an urgent need to design a system that will recognize the currency authenticity and its value.

Currency duplication also known as counterfeit currency is a vulnerable threat on economy. It is now a common phenomenon due to advanced printing and scanning technology. The possible solutions are to use either chemical

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 7, July 2022

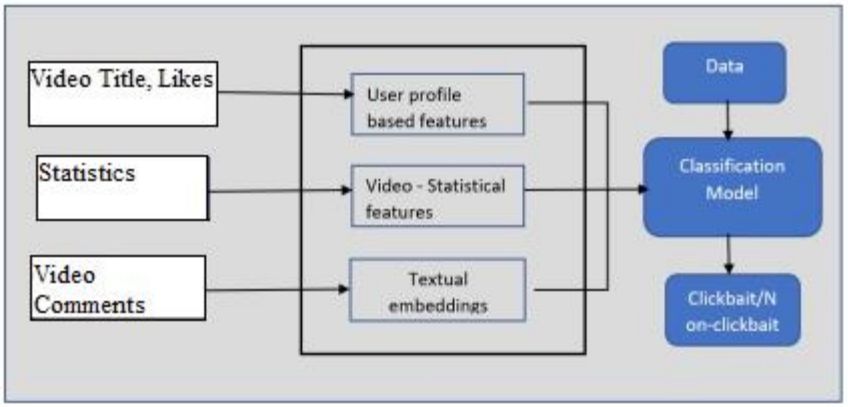
DOI: 10.17148/IJARCCE.2022.11735

properties of the currency or to use its physical appearance.
Some of these are implemented using image mapping techniques but are not very accurate. To overcome the limitations of the already implemented techniques, in this paper, we propose an efficient and cost-effective counterfeit currency detection through a Mobile application that uses Machine Learning technique Convolutional Neural Network to authenticate the banknotes.

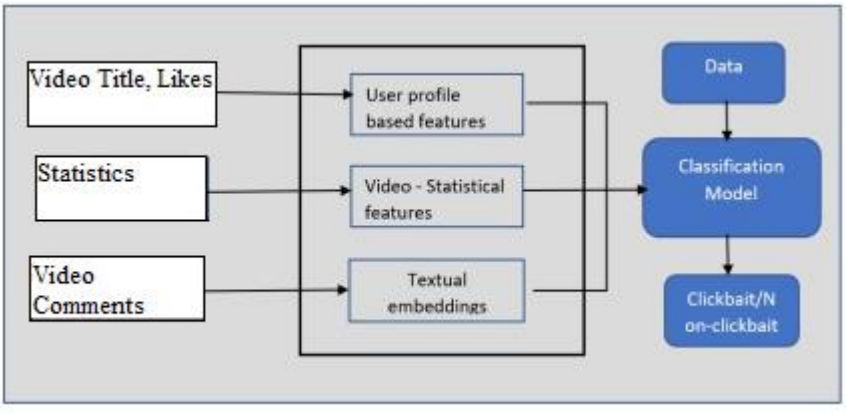
I. PROPOSED METHODOLOGY

Proposed System:

- Initially in proposed system, we gather data like Title from YouTube along with Comments, likes, views and Statistics.
- Next, we train, pre-process and evaluate the data sets.
- In Multi-Model Architecture we apply SVM algorithm for Title, Comments, likes and Statistics.
- According to the output obtained by this algorithm we classify video is clickbait or not.



Methodology:
The project is implemented in modular approach. Each module is coded as per the requirements and tested, and this process is iterated till the all the modules have been thoroughly implemented.
Steps to work Procedure:
1. The first step includes the data set collection where it takes the link of the video as input.
2. Then fetches the statistics of video from the YouTube API for processing the data.
3. Statistics include the data like video title, likes, comments, views.
4. The collected datasets undergo tokenization, genism and word2vec processes.
5. SVM algorithm is applied to these datasets after all the processes mentioned above.
6. Next, SVM model is built to predict the output.
7. Finally, the model will predict whether the given video is click-Bait or not.



IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 7, July 2022

DOI: 10.17148/IJARCCE.2022.11735

II. SYSTEM DESIGN IMPLIMENTATION

System Analysis

The process of studying a procedure or business to identify its goals and purposes and create systems and procedures that will achieve them in an efficient way. Another view sees system analysis as a problem-solving technique that breaks down a system into its component pieces for the purpose of the studying how well those component parts work and interact to accomplish their purpose.

Study of the system:

The system follows a minimal and intuitive approach. The front end is designed such that its straight to the point and easy to navigate for anyone who knows how to use computers. The user can input the sample images and can get the output directly on console or can use button to get the individual test results.

Operating Environment:

Python Python is an interpreted, high-level, general-purpose programming language. Created by Guido van Rossum and first released in 1991. Python's design philosophy emphasizes code readability with its notable use of significant whitespace. It's language constructs and object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects.

Matplotlib Matplotlib is a cross-platform, data visualization and graphical plotting library for Python and its numerical extension NumPy. As such, it offers a viable open-source alternative to MATLAB. Developers can also use matplotlib's APIs (Application Programming Interfaces) to embed plots in GUI applications

Numpy: NumPy is a Python package which stands for 'Numerical Python'. It is the core library for scientific computing, which contains a powerful n-dimensional array object, provide tools for integrating C, C++ etc. It is also useful in linear algebra random number capability etc. NumPy array can also be used as an efficient multi-dimensional container for generic data. Now, let me tell you what exactly is a python numpy array.

O S Module: The OS module in Python provides functions for interacting with the operating system. OS comes under Python's standard utility modules. This module provides a portable way of using operating system-dependent functionality. The OS and os.path modules include many functions to interact with the file system.

Pickle: The OS module in Python provides functions for interacting with the operating system. OS comes under Python's standard utility modules. This module provides a portable way of using operating system-dependent functionality. The os and os.path modules include many functions to interact with the file system.

Argprase: The argparse module makes it easy to write user-friendly command-line interfaces. It parses the defined arguments from the sysargv. The argparse module also automatically generates help and usage messages, and issues errors when users give the program invalid arguments. The argparse is a standard module; we do not need to install it. A parser is created with Argument Parser and a new parameter is added with add_argument. Arguments can be optional, required, or positional.

Pandas: Pandas is an open-source library that is made mainly for working with relational or labeled data both easily and intuitively. It provides various data structures and operations for manipulating numerical data and time series. This library is built on top of the NumPy library. Pandas is fast and it has high performance & productivity for users.

Sklearn: Scikit-learn (Sklearn) is the most useful and robust library for machine learning in Python. It provides a selection of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction via a consistence interface in Python. It is the most useful and robust library for machine learning in Python. It provides a selection of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction via a consistence interface in Python.

Regex: A RegEx, or Regular Expression, is a sequence of characters that forms a search pattern. RegEx can be used to check if a string contains the specified search pattern.Python has a builtin package called re, which can be used to work with Regular Expressions. A regular expression (shortened as regex or regexp sometimes referred to as rational expression is a sequence of characters that specifies a search pattern in text. Usually such patterns are used.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 7, July 2022

DOI: 10.17148/IJARCCE.2022.11735

III. TEST CASES

Test case nummber	Testing Scenario	Expected result	Result
TC – 01	Giving wrong image format	System will not process the given data	Pass
TC – 02	Loading the wrong image data	System will produce the wrong prediction	Pass
TC – 03	Loading the correct image	System will process the image and produce grayscale image	Pass
TC – 04	Feature extraction	System will extract the features from the grayscale image	Pass
TC – 05	Feature Comparison	Extracted features will be compared with the model	Pass
TC – 06	Detection	Model will detect the disease based on the features	Pass
TC-07	Final Result	User can view the disease detected from the provided image with blood group	Pass

IV. CONCLUSION WITH FUTURE ENHANCEMENTS.

This project has helped us to solve one of the major issues looked by the YouTube system, which is the clickbait issue. Initially, the data was scraped using the YouTube API since a dataset was not readily available. To the best information and research did, the misleading content issues have been done in the field of news, twitter tweets, however, have been not carefully considered for the YouTube System. This project aims to bridge this gap and help in improving the user experience and aides in expanding the trust of the investors by detecting the misleading videos. Multiple machine learning algorithms have been used for this research out of which the best accuracy of 90% was achieved by the SVM.

As future work, we will explore the performance of the model on various kinds of media and aim to understand the different insights from the model interpretability to better clickbait videos. Alongside image analysis on the thumbnail of the videos could likewise be considered as a distinctive factor since clickbait videos are bound to use user attracting thumbnails. These factors would further help in distinguishing the videos more precisely. Although machine learning algorithm have been considered for this project, more deep learning algorithms such as Recurrent Neural Network (RNN) could also be additionally considered.

REFERENCES

[1] Kit Smith, “46 fascinating and incredible youtube statistics,” 2019.
[2] Eric Rosenberg, “How youtube ad revenue works,” 2018.

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 7, July 2022

DOI: 10.17148/IJARCCE.2022.11735

[3] George Loewenstein, "The psychology of curiosity: A review and reinterpretation," Psychological Bulletin, vol. 116, no. 1, pp. 75, 1994.
[4] Ankesh Anand, Tanmoy Chakraborty, and Noseong Park, "We used neural networks to detect clickbaits: You won't believe what happened next!," arXiv.org, 2016.
[5] Hai-Tao Zheng, Jin-Yuan Chen, Xin Yao, Arun Kumar Sangaiah, Yong Jiang, and Cong-Zhi Zhao, "Clickbait convolutional neural network," vol. 10, no. 5, pp. 138, 2018.
[6] Ryan Kiros, Yukun Zhu, Ruslan Salakhutdinov, Richard Zemel, Antonio Torralba, Raquel Urtasun, and Sanja Fidler, "Skip-thought vectors," arXiv.org, 2015.
[7] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean, "Efficient estimation of word representations in vector space," arXiv.org, 2013.
[8] Wang Yu Wang Zhichao, Weng Nan, "Research of title party news identification technology based on topic sentence similarity," , no. 11, pp. 48–53, 2011.
[9] Martin Potthast, Sebastian Kopsel, Benno Stein, and ¨ Matthias Hagen, "Clickbait detection," 2016, Advances in Information Retrieval, pp. 810–817, Springer International Publishing.
[10] Abhijnan Chakraborty, Bhargavi Paranjape, and Niloy Ganguly, "Stop clickbait: Detecting and preventing clickbaits in online news media," arXiv.org, 2016.
[11] Yimin Chen, Niall J. Conroy, and Victoria L. Rubin, "Misleading online content: Recognizing clickbait as "false news"," in Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection, New York, NY, USA, 2015, WMDD '15, p. 15–19, Association for Computing Machinery.
[12] Hai-Tao Zheng, Xin Yao, Yong Jiang, Shu-Tao Xia, and Xi Xiao, "Boost clickbait detection based on user behavior analysis," 2017, Web and Big Data, pp. 73–80, Springer International Publishing.
[13] S. Zannettou, S. Chatzis, K. Papadamou, and M. Sirivianos, "The good, the bad and the bait: Detecting and characterizing clickbait on youtube," in 2018 IEEE Security and Privacy Workshops (SPW), 2018, pp. 63–69.