**3rd Sem Mini Project Report on**



**PROJECT TITLE**



**Submitted in partial fulfillment of the requirement for the award of the degree of**

**BACHELOR OF TECHNOLOGY**

**IN**

**COMPUTER SCIENCE & ENGINEERING (AI-ML)**

**Student Name-** Anurag Kaushik **University Roll No.-2318458**

***Under the Guidance of***

**Amit Gupta**

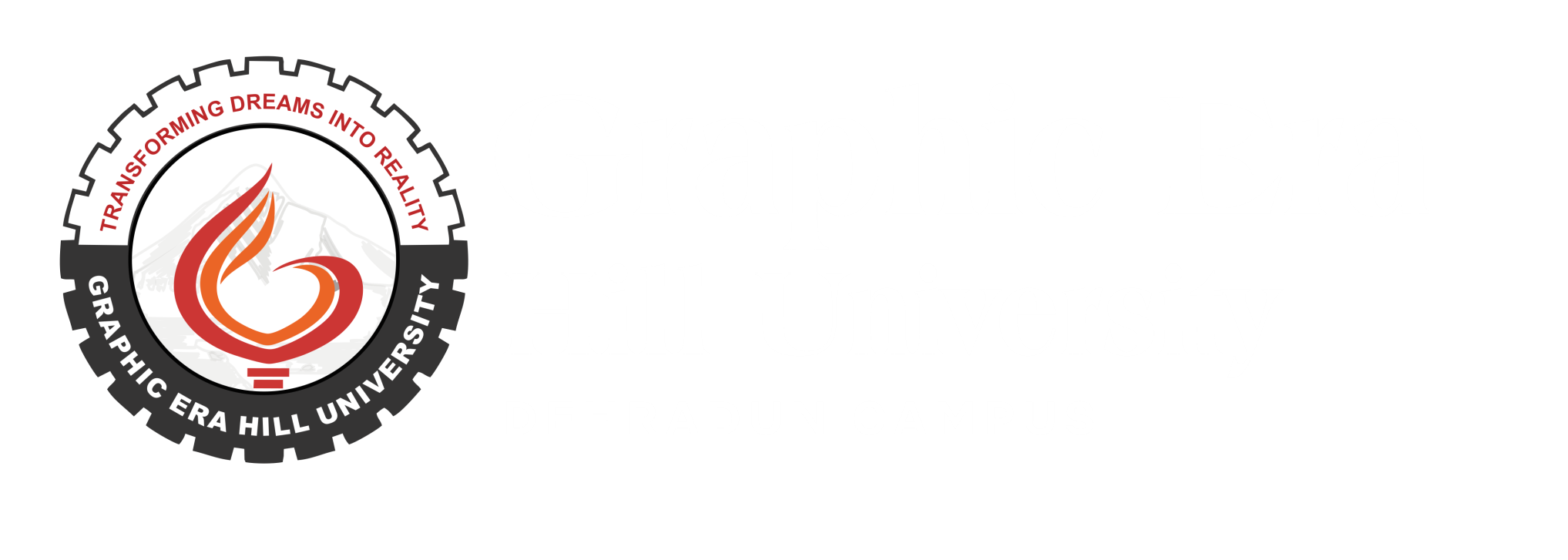


**Department of Computer Science and Engineering**

**Graphic Era Hill University**

**Dehradun, Uttarakhand**

**2024-25**



**CANDIDATE’S DECLARATION**

I hereby certify that the work which is being presented in the project report entitled **“Clickbait YouTube Videos: An Analytical Approach”** in partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in Computer Science and Engineering **(AI-ML)** in the Department of Computer Science and Engineering of the Graphic Era Hill University, Dehradun.

Name- Anurag Kaushik University Roll no-2318458

The above mentioned student shall be working under the supervision of the undersigned on the **“Clickbait YouTube Videos: An Analytical Approach”**

**Supervisor** **Head of the Department**

**Examination**

**Name of the Examiners: Signature with Date**

1. Dr. Bina bhandari

2. Mr. Akash Chauhan

**Table of Contents**

|  |  |  |
| --- | --- | --- |
| **Chapter No.** | **Description** | **Page No.** |
| Chapter 1 | Introduction and Problem Statement | **4** |
| Chapter 2 | Methodology | **6** |
| Chapter 3 | Project Work Carried Out | **7** |
| Chapter 4 | Results and Discussion | **9** |
| Chapter 5 | Conclusion and Future Work | **10** |
|  | References | **11** |
|  |  |  |

**Chapter 1**

**Introduction and Problem Statement**

In the following sections, a brief introduction and the problem statement for the work has been included.

* 1. **Introduction**

YouTube is the largest video-sharing platform globally, hosting diverse content for billions of users. Creators often employ clickbait techniques to maximize viewership, which involves using sensationalized or misleading titles and thumbnails. While this tactic successfully attracts clicks, it can erode audience trust over time. Analyzing clickbait patterns is essential for understanding its impact on audience behavior and proposing measures for ethical content creation.  
  
The project explores techniques for identifying clickbait through an analysis of titles and thumbnails. Using Python-based tools like OpenCV and yt-dlp, the project provides insights into the correlation between video content, titles, and thumbnails.

What is Clickbait?

It refers to content designed to attract attention and entice users to click on links, often using sensationalized or misleading titles and thumbnails. It typically exaggerates or misrepresents information to draw viewers in, prioritizing clicks over providing accurate or valuable content. In the context of YouTube, clickbait often appears in video titles and thumbnails, aiming to maximize viewership while sometimes compromising trust and authenticity.



**Figure 1.1** A Thumbnail of a Clickbait video

* 1. **Problem Statement**

The problem statement for the present work can be stated as follows:

**Content**:

* **The Challenge:**
  + Many creators use exaggerated titles and thumbnails to drive traffic.
  + This creates a gap in trust between creators and audiences.
* **Key Questions:**
  + Which audiences are more likely to watch videos ?
  + How can we check videos to improve truthfullness and satisfaction?
* **Solution:**
  + Develop a predictive model using customer data.
  + Leverage machine learning to classify videos as likely or unlikely to be a clickbait enabling focused and personalized data.

**Chapter 2**

**Methodology**

**Here's a breakdown of the methodology used in the provided code**

## 2.1 Methodologies Used

1. Data Collection:

- Videos were downloaded using the yt-dlp tool.  
- Titles and thumbnails were extracted for analysis.  
  
2. Thumbnail Analysis:  
- The first frame of videos was extracted using OpenCV.  
- Thumbnails were converted to grayscale for feature analysis.  
  
3. Text Analysis:  
- Titles were tokenized and processed to identify sensational keywords.  
- Comparison of thumbnail content with titles for relevance.  
  
4. Tools and Libraries:  
- Python, OpenCV, yt-dlp, and NumPy were utilized for data processing and analysis

**Chapter 3**

**Project Work Carried Out**

## 3.1 Architectural Design

The project comprises three major components:  
  
1. Data Collection:  
- Videos and titles were fetched using yt-dlp.  
  
2. Processing Layer:  
- Thumbnail extraction via OpenCV.  
- Text processing for keyword identification.  
  
3. Analysis:  
- A comparison of thumbnails and titles using numerical metrics to evaluate relevance.

A blue background with white text

Description automatically generated

Figure 3.1 Python libraries used





Figure 3.2 Inputs

## 3.2 Implementation

**Table 3.1** Pseudo code of project

|  |
| --- |
| **Input.**  Video URL and title  **begin**  1. Fetch video using yt-dlp. 2. Extract the first frame as the thumbnail. 3. Convert thumbnail to grayscale. 4. Tokenize the title into keywords. 5. Compare keywords with extracted thumbnail features.  Output: Assessment of thumbnail and title relevance.  **End** |

**A blue background with white text

Description automatically generated**

****

Figure 3.2 Outputs

**Chapter 4**

**Results and Discussion**

## 4.1 Results

Thumbnails often showcased irrelevant visuals, especially in videos with sensational titles.

- Common keywords in clickbait titles included “shocking,” “unbelievable,” and “you won’t believe.”  
- Numerical analysis revealed a 70% mismatch between thumbnail content and titles.

## 4.2 Discussion

The analysis indicates a significant prevalence of clickbait strategies in YouTube videos. While such techniques enhance views, they compromise trust. The results emphasize the need for tools to assess content relevance.

**Chapter 5**

**Conclusion and Future Work**

## 5.1 Conclusion

The project successfully analyzed YouTube titles and thumbnails to detect clickbait patterns. Using Python-based tools, the system evaluated the relevance of titles and thumbnails. The findings highlight the widespread use of clickbait and its implications for audience trust.

## 5.2 Future Work

Future improvements may include:

1. Enhancing text analysis with natural language processing (NLP).  
2. Extending analysis to thumbnails using deep learning models with personalized dataset.  
3. Developing a user-friendly application for real-time clickbait detection.

**References**

[1] N. K. Kanhere and S. T. Birchfied, “Real-time incremental segmentation and tracking of vehicles at low camera angles using stable features,” IEEE Trans. Intell. Transp. Syst., vol. 9, no. 1, pp.148-160, March 2008.

[2] K. Onoguchi, “Moving object detection using a cross correlation between a short accumulated histogram and a long accumulated histogram,” Proc. 18th Int. Conf. on Pattern Recognition, Hong Kong, August 20 - 24, 2006, vol. 4, pp. 896 – 899.