

MADANAPALLE INSTITUTE OF TECHNOLOGY & SCIENCE
(UGC – AUTONOMOUS)

COMPUTER SCIENCE & ENGINEERING

COURSE: 20CSE606 AI TOOLS, TECHNIQUES AND APPLICATIONS

A Project Assignment 1 On

YOUTUBE TRENDING VIDEO ANALYSIS

Submitted By

G. Manju Pujitha	-	20691A05C4
V. Sahana	-	20691A05E8
A. Sri Harsha	-	20691A05H1

Submitted To

Dr. S. Kusuma, Ph. D
Assistant Professor
Department of Computer Science & Engineering

Faculty In charge

INDEX

S.NO	TOPIC	PAGE NO.
1.	INTRODUCTION	1
	1.1 Motivation	1
	1.2 Problem Definition	1
	1.3 Objective of the Project	1
	1.4 Limitations of Project	2
2	DESIGN	3
	2.1 Introduction	3
3.	IMPLEMENTATION AND RESULTS	4
	3.1 Introduction	4
	3.2 Implementation of key functions	4
	3.3 Method of Implementation	5
	3.4 Output Screens and Result Analysis	12
	3.5 Conclusion	
4.	CONCLUSION	10
	4.1 Conclusion	10
8.	REFERENCES	11

ABSTRACT

Online video streaming platforms are heavily used nowadays. Websites such as YouTube offers content creators a great platform to share their knowledge, ideas and interesting information to their viewers. For a video to reach to maximum people, YouTube offers a trending page on website that shows videos which are trending at that particular time. Other than few viral videos that achieve high view count which are predictable to end up in trending section, rest of the videos cannot be predicted. Corporate companies are using social media for improving their businesses, the data mining and analysis are very important in these days. This paper deals with analysis of YouTube Data on Trending Videos. The analysis is done using user features such as Views, Comments, Likes, and Dislikes. Analysis can be performed using algorithms like Linear Regression, classification and other Machine learning models and python libraries like pandas, matplotlib library to classify the YouTube Data and obtain useful information.

CHAPTER-1

INTRODUCTION

1.1 MOTIVATION

The motivation behind the project "YouTube Trending Video Analysis Using AI" is to explore the potential of artificial intelligence in analyzing online content and to provide insights into the factors that contribute to the success of YouTube's trending videos. By understanding the characteristics of successful videos, content creators and marketers can optimize their videos to increase their reach, engagement, and ultimately, their revenue. Furthermore, the project aims to demonstrate the usefulness of AI in the field of digital marketing and to showcase its potential for improving business performance. Ultimately, the project seeks to contribute to the development of AI technologies and their practical applications in the digital domain.

1.2 PROBLEM DEFINITION

The problem definition of the project "YouTube Trending Video Analysis Using AI" is to identify the factors that contribute to the success of YouTube's trending videos and provide insights into their characteristics. Although YouTube provides data on the views, likes, and comments of videos, it does not provide information on the specific factors that lead to a video's success. This lack of information makes it difficult for content creators and marketers to optimize their videos for maximum performance. Therefore, the project aims to address this problem by using artificial intelligence to analyze a range of factors, such as video length, category, keywords, and engagement metrics, to identify patterns that contribute to a video's success. By providing insights into the characteristics of successful videos, the project can help content creators and marketers to optimize their videos for better performance and increase their audience reach and engagement.

1.3 OBJECTIVE OF THE PROJECT

- To understand characteristics that makes a video end up in trending section on YouTube.
- To grasp knowledge about YouTube's algorithm and extract key aspects that can be used to improve growth of YouTube channels.

- Provide content creators and other digital businesses with information about features required to make a video that trends and reaches to maximum people.
- To accelerate a channels growth and increase their subscriber count by making trending videos.
- By increasing subscriber count and views this can help you youtuber's to earn more money, sponsorships and opportunities to excel in their social media carrier.

1.4 LIMITATIONS OF PROJECT

There are several limitations of the project "YouTube Trending Video Analysis Using AI" that should be considered:

- 1) Data availability: The project relies on data collected through YouTube API, which may not always be comprehensive or up-to-date. Additionally, the API may not provide access to all the data points required for the analysis, which could limit the accuracy of the results.
- 2) Sample bias: The project focuses on analyzing YouTube's trending videos, which represent only a small proportion of all the videos on the platform. As a result, the insights gained from the project may not be representative of the platform as a whole.
- 3) Algorithmic limitations: The accuracy of the analysis depends on the quality of the machine learning algorithms used. While modern AI algorithms are highly advanced, they may still produce inaccurate results due to biases in the data or limitations in the algorithm's design.
- 4) Interpretation of results: The project presents results in the form of data visualizations and a dashboard. However, the interpretation of the results may require domain expertise or additional context, which may not always be available to non-experts.
- 5) Dynamic nature of the platform: YouTube is a dynamic platform that is constantly evolving, with changes to algorithms, user behavior, and content creation practices. This means that the insights gained from the analysis may become outdated over time and require ongoing monitoring and updates.

CHAPTER-2 DESIGN

2.1 INTRODUCTION

The world of online video content has exploded in recent years, with YouTube being one of the most popular platforms for sharing and consuming videos. With millions of videos uploaded every day, it can be challenging for content creators and marketers to stand out and gain traction. This is where the project "YouTube Trending Video Analysis Using AI" comes in, which aims to use artificial intelligence to identify the characteristics of successful YouTube videos.

By analyzing a range of factors, such as video length, category, keywords, and engagement metrics, the project seeks to provide insights into the factors that contribute to a video's success on the platform. Through this analysis, content creators and marketers can optimize their videos to increase their reach and engagement, ultimately leading to greater revenue.

The project uses machine learning algorithms to analyze YouTube's trending videos and identify patterns that contribute to their success. The results are presented through data visualizations and a dashboard, providing an easy-to-under

CHAPTER-3

IMPLEMENTATION AND RESULTS

3.1 INTRODUCTION

The implementation of the project "YouTube Trending Video Analysis Using AI" involves applying machine learning algorithms to identify the factors that contribute to a video's success on the platform. The goal of the implementation is to provide insights into the characteristics of successful videos and help content creators and marketers to optimize their videos for maximum performance.

The implementation of the project involves several stages, including data collection, data cleaning, feature extraction, model training, and results presentation. Data is collected from the YouTube API, and then cleaned and processed to remove irrelevant or incorrect information. Feature extraction algorithms are then used to identify key features that contribute to a video's success, such as video length, category, and keywords.

Machine learning algorithms are trained on the data, and the resulting models are evaluated to determine their accuracy in predicting the success of videos. The results are then presented through data visualizations and a dashboard, providing insights into the characteristics of successful YouTube videos.

The implementation of the project has the potential to provide significant benefits to content creators and marketers, helping them to optimize their videos for better performance and increase their audience reach and engagement. With the explosion of online video content, understanding the factors that contribute to success on YouTube is more important than ever, and the implementation of this project can help to unlock the secrets of successful videos on the platform.

3.2 IMPLEMENTATION OF KEY FUNCTIONS

The implementation of the project "YouTube Trending Video Analysis Using AI" involves several key functions, including:

1. Data collection: The first step in the implementation process is to collect data from the YouTube API, including information about video views, likes, comments, and other relevant data.
2. Data cleaning: The collected data is then cleaned and processed to remove any irrelevant or incorrect information.
3. Feature extraction: Feature extraction algorithms are used to identify key features that contribute to a video's success, such as video length, category, and keywords.

4. Model training: Machine learning algorithms are used to train models on the data, which can predict the success of a video based on its features.
5. Model evaluation: The trained models are evaluated to determine their accuracy in predicting the success of videos.
6. Results presentation: The results are then presented through data visualizations and a dashboard, providing insights into the characteristics of successful YouTube videos.
7. Optimization: The project can be optimized to improve its performance, such as by using more advanced algorithms or incorporating additional features.
8. Deployment: The project can be deployed as a web application or integrated into other software tools, allowing content creators and marketers to easily access its insights and optimize their videos.

These key functions are essential for the successful implementation of the project and can help to provide significant benefits to content creators and marketers looking to improve the performance of their videos on YouTube.

3.3 METHOD OF IMPLEMENTATION AND OUTPUTS

```
import pandas as pd
import numpy as np
import matplotlib as mpl
from matplotlib import pyplot as plt
import seaborn as sns

import warnings
from collections import Counter
import datetime
import wordcloud
import json

df = pd.read_csv("USvideos.csv")
PLOT_COLORS = ["#268bd2", "#0052CC", "#FF5722", "#b58900", "#003f5c"]
pd.options.display.float_format = '{:.2f}'.format
sns.set(style="ticks")
plt.rc('figure', figsize=(8, 5), dpi=100)
plt.rc('axes', labelpad=20, facecolor="#ffffff", linewidth=0.4, grid=True, labelsz=14)
plt.rc('patch', linewidth=0)
plt.rc('xtick.major', width=0.2)
plt.rc('ytick.major', width=0.2)
plt.rc('grid', color='#9E9E9E', linewidth=0.4)
plt.rc('font', family='Arial', weight='400', size=10)
plt.rc('text', color='#282828')
plt.rc('savefig', pad_inches=0.3, dpi=300)
```



```
df["description"] = df["description"].fillna(value="")
df.describe()
```

↳

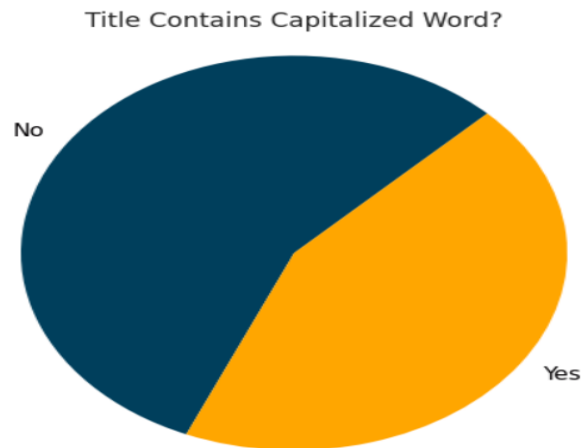
	category_id	views	likes	dislikes	comment_count
count	40949.00	40949.00	40949.00	40949.00	40949.00
mean	19.97	2360784.64	74266.70	3711.40	8446.80
std	7.57	7394113.76	228885.34	29029.71	37430.49
min	1.00	549.00	0.00	0.00	0.00
25%	17.00	242329.00	5424.00	202.00	614.00
50%	24.00	681861.00	18091.00	631.00	1856.00
75%	25.00	1823157.00	55417.00	1938.00	5755.00
max	43.00	225211923.00	5613827.00	1674420.00	1361580.00

```
def contains_capitalized_word(s):
    for w in s.split():
        if w.isupper():
            return True
    return False
```

#we want to see how many trending video titles have at least one uppercase word

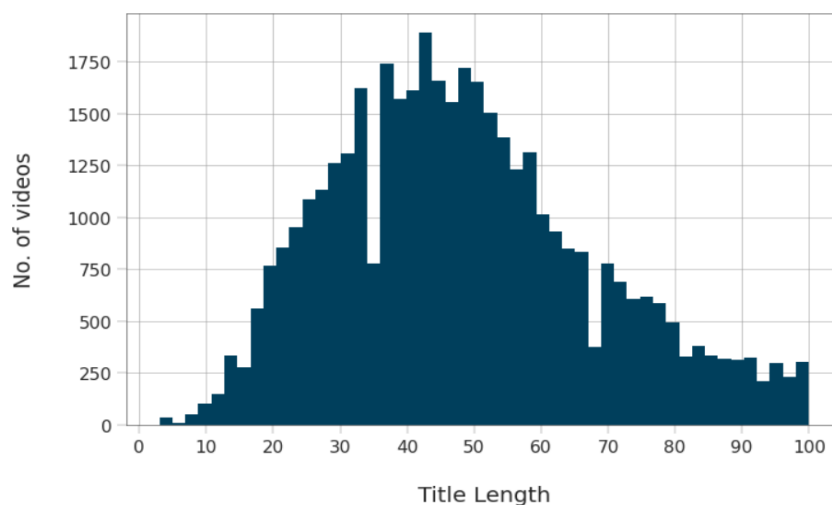
```
df["contains_capitalized"] = df["title"].apply(contains_capitalized_word)

value_counts = df["contains_capitalized"].value_counts().to_dict()
fig, ax = plt.subplots()
_ = ax.pie([value_counts[False], value_counts[True]], labels=['No', 'Yes'],
           colors=['#003f5c', '#ffa600'], textprops={'color': '#040204'}, startangle=45)
_ = ax.axis('equal')
_ = ax.set_title("Title Contains Capitalized Word?")
```



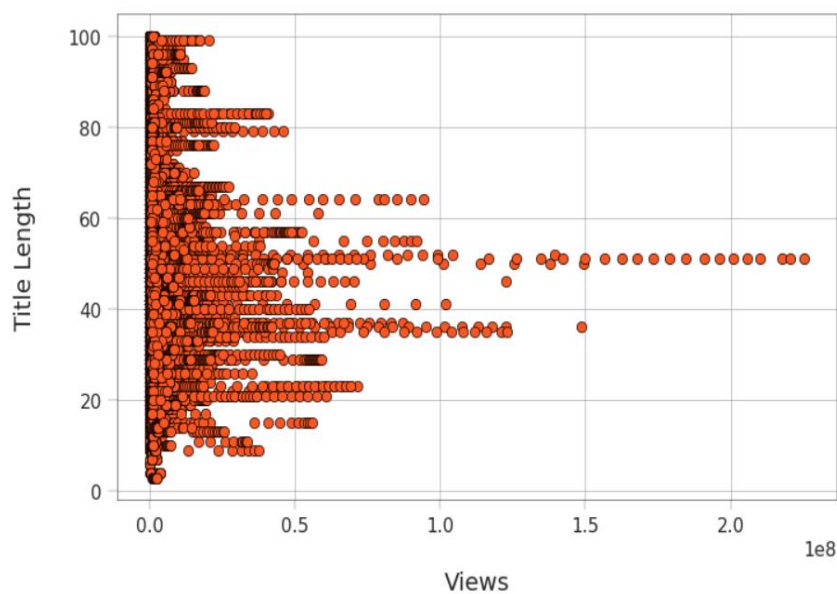
#We can see that 44% of trending video titles contain at least one word in all caps. We will use our added variable later to analyze the correlation between the variables. Let's add another column to our dataset to analyze the length of titles of videos, then plot the title length histogram to get an idea of the length of trending video titles:

```
df["title_length"] = df["title"].apply(lambda x: len(x))
fig, ax = plt.subplots()
_ = sns.distplot(df["title_length"], kde=False, rug=False,
                 color=PLOT_COLORS[4], hist_kws={'alpha': 1}, ax=ax)
_ = ax.set(xlabel="Title Length", ylabel="No. of videos", xticks=range(0, 110, 10))
```



#We can see that the videos title length distribution looks like a normal distribution, where most videos have a title length of around 30-60 characters. Now, let's draw a scatter plot to analyze the relationship between the title length and the number of views:

```
fig, ax = plt.subplots()
_ = ax.scatter(x=df['views'], y=df['title_length'], color=PLOT_COLORS[2], edgecolors="#000000", linewidths=0.5)
_ = ax.set(xlabel="Views", ylabel="Title Length")
```



#Now let's see how the variables in the dataset correlate to each other: for example, we would like to see how views and likes correlate, meaning that views and likes increase and decrease together:

```
h_labels = [x.replace('_', ' ').title() for x in
             list(df.select_dtypes(include=['number', 'bool']).columns.values)]
```

```
fig, ax = plt.subplots(figsize=(10,6))
_ = sns.heatmap(df.corr(), annot=True, xticklabels=h_labels, yticklabels=h_labels, cmap=sns.
cubehelix_palette(as_cmap=True), ax=ax)
```



CHAPTER-4

CONCLUSION

4.1 CONCLUSION

In conclusion, the project "YouTube Trending Video Analysis Using AI" is a valuable tool for content creators and marketers looking to optimize their videos for maximum performance on the platform. By applying machine learning algorithms to identify the factors that contribute to a video's success, the project provides insights into the characteristics of successful videos and can help to improve the reach and engagement of videos on YouTube.

The project involves several stages, including data collection, cleaning, feature extraction, model training, evaluation, and results presentation. These stages are essential for the successful implementation of the project and can be optimized to improve its performance over time.

Overall, the project has significant potential to provide benefits to content creators and marketers, helping them to understand the factors that contribute to success on YouTube and optimize their videos accordingly. With the continued growth of online video content, the project represents an important contribution to the field of digital marketing and content creation, and its insights can help to unlock the secrets of successful videos on the platform.

CHAPTER-5

REFERENCES

REFERENCES

1. R. A. Rifhano, "The Influence of Intensity of Accessing Youtube on Student Learning Activities of the Faculty of Business Economics Telkom University", *YouTube Introd*, vol. 4, no. 2, 2017.
2. E. David, M. Sondakh and S. Harilama, "The Influence of Vlog Content on Youtube on the Formation of Attitudes of Communication Science Students", *Acta Diurna*, vol. 6, no. 1, [online]