**AIML Project**

Logo, company name

Description automatically generated

**Team:**

**Shreya (19csu299)**

**Siddharth (19csu307)**

**Tuhini (19csu416)**

**Topic: YouTube Sentiment Analysis**

Logo, company name

Description automatically generated

**ACKNOWLEDGEMENT**

It is our pleasure to be indebted to various people, who directly or indirectly contributed to the development of this work and who influenced our thinking, behavior, and acts during study.

We express our sincere gratitude to **Miss Divya Jain** (Faculty), for providing an opportunity to learn and implement the Concepts of Machine Learning in different ways and on different Topics. Thanks for helping us on each step. We are thankful for her support, cooperation, and motivation provided to us during the learning lessons and theoretical parts of the machine learning and artificial intelligence.

We would also like to thank the almighty and our friends with whom we shared our day-to-day experience and received lots of suggestions that improved our quality of work.

**Shreya (19csu299)**

**Siddharth (19csu307)**

**Tuhini (19csu416)**

**CONTENT**

[ACKNOWLEDGEMENT 2](file:///G:\NCU\SEMESTER%20V\Summer%20Internship\Summer%20Internship%20Report%20final.docx#_Toc49638108)

[INTRODUCTION 4](file:///G:\NCU\SEMESTER%20V\Summer%20Internship\Summer%20Internship%20Report%20final.docx#_Toc49638110)

[ABOUT PROJECT 5](file:///G:\NCU\SEMESTER%20V\Summer%20Internship\Summer%20Internship%20Report%20final.docx#_Toc49638111)

[ABOUT DATASET 5](file:///G:\NCU\SEMESTER%20V\Summer%20Internship\Summer%20Internship%20Report%20final.docx#_Toc49638112)

QUESTIONS and [OUTCOMES 6-](file:///G:\NCU\SEMESTER%20V\Summer%20Internship\Summer%20Internship%20Report%20final.docx#_Toc49638113)13

[ANALYTICS 14-15](file:///G:\NCU\SEMESTER%20V\Summer%20Internship\Summer%20Internship%20Report%20final.docx#_Toc49638123)

[ANALYSIS 16](file:///G:\NCU\SEMESTER%20V\Summer%20Internship\Summer%20Internship%20Report%20final.docx#_Toc49638131)

[CONCLUSION 16](file:///G:\NCU\SEMESTER%20V\Summer%20Internship\Summer%20Internship%20Report%20final.docx#_Toc49638132)

[BIBLIOGRAPHY 16](file:///G:\NCU\SEMESTER%20V\Summer%20Internship\Summer%20Internship%20Report%20final.docx#_Toc49638133)

**INTRODUCTION**

[Sentiment Analysis](https://www.analyticssteps.com/blogs/7-natural-language-processing-techniques-extracting-information) is one of the [Natural Language Processing](https://www.analyticssteps.com/blogs/introduction-natural-language-processing-text-cleaning-preprocessing)[techniques](https://www.analyticssteps.com/blogs/7-natural-language-processing-techniques-extracting-information), which can be used to determine the sensibility behind the texts, i.e. tweets, movie reviews, YouTube comments, any incoming message, etc.  It’s in use on a large scale as many big firms are using it to examine customer reviews about their product/services, on social media, their websites which helps them to maintain their brand values. Some of the [big corporations](https://theappsolutions.com/blog/development/sentiment-analysis-for-business/) using this technique to invigorate customer engagement with their service areas:

* Trip Advisor
* Google
* Apple
* KFC

Other application of sentiment analysis includes:

* Twitter sentiment analysis,
* IMDB movie ratings,
* [Amazon customer reviews,](https://www.datasciencecentral.com/profiles/blogs/sentiment-analysis-of-amazon-customer-reviews-with-visualizations#:~:text=The%20sentiment%20analysis%20of%20customer,comments%20and%20improve%20their%20products.&text=The%20sentiment%20analyzer%20such%20as,as%20shown%20in%20figure%201.)
* YouTube videos comments

Current YouTube usage statistics indicate the approximate scale of the site: at the time of this writing there are quite 1 billion unique users viewing video content, watching over 6 billion hours of video each month. Also, YouTube accounts for 20% of web traffic and 10% of total internet traffic. YouTube provides many social mechanisms to gauge user opinion and views a few videos by means of voting, rating, favourites, sharing and negative comments, etc. It’s important to notice that YouTube provides more than just video sharing

**OUR PROJECT**

we have done

sentimental analysis of public comments by using

we have done

sentimental analysis of public comments by using

we have done

sentimental analysis of public comments by using

Our Project is based on Natural language processing technique We have done sentimental analysis of public comments by using python library called "NLTK", a platform for building python programs to work with Human language data. More specifically, what I have used is called VADER (Valence Aware Dictionary and Sentiment Reasoner) which is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed on social media. Calculating the sentiment scores of positive, Negative, and neutral comments, analysing YouTube channels with 100% positive comments and 0% positive comments and forming Wordcloud of most frequent , positive negative and neutral comments.

**ABOUT DATASET**

Our dataset contains 2 csv files named as USvideos and UScomments both containing string and int datatypes and are totally different from one another. Dataset has been collected / downloaded from Kaggle.

More about dataset:

1. USvideos:

Contains data in 11 sections i.e.

1. Video\_id
2. Title
3. Channel\_title
4. category\_id
5. tags
6. views
7. likes
8. dislikes
9. comments
10. thumbnail
11. date

Graphical user interface, application, table, Excel

Description automatically generated

1. USvideos:

Contains data in 4 sections i.e.

1. Video\_id
2. Comment\_text
3. Likes
4. replies

Graphical user interface, application, table, Excel

Description automatically generated

**BEHIND THE PROJECT WORK**

Text, letter

Description automatically generated

Conducting multiple teams meeting and exploring various sections and possibilities, we came to conclusion of a total of 19 possible step our project holds.

**OUTCOMES**

**Importing The Libraries**

**Graphical user interface, text, application, Word

Description automatically generated**

**Loading the Dataset**

**Graphical user interface, text, application, email

Description automatically generated**

## Exploratory Analysis and Data Cleaning

## Graphical user interface Description automatically generated

## Graphical user interface, text, application, email Description automatically generated

## Graphical user interface, text, application Description automatically generated

## Graphical user interface, text Description automatically generated

## Graphical user interface, text, application, email Description automatically generated

## Graphical user interface, text, application Description automatically generated

## Graphical user interface, text, application, email Description automatically generated

## Graphical user interface, text, application, email Description automatically generated

## Tokenization

## Graphical user interface, text, application Description automatically generated

## Lemmatization

## Graphical user interface, text, application, email Description automatically generated

## 

**ANALYSIS**

**Sentiment Analysis**

**Graphical user interface, text, application, email

Description automatically generated**

## Setting The Sentiment Scores

## Table Description automatically generated

## Classifying the Sentiment scores as Positive, Negative and Neutral

## Table Description automatically generated

## 

## Calculating Positive Percentage

## Text, application Description automatically generated

## Making a dataframe of the videos with their Positive Percentages

## Graphical user interface, text, application, email Description automatically generated

## Adding channel name to dataframe

## Graphical user interface, text, application, email Description automatically generated

### Getting videos and channels with 100% positive comments

## Graphical user interface, text, application, email Description automatically generated

### Getting channels with 0 % positive percentage i.e either negetive or neutral comments

## Table Description automatically generated

## Wordcloud of most frequent comments

## Graphical user interface, text Description automatically generated with medium confidence A picture containing text, newspaper Description automatically generated

## Wordcloud of Positive Comments

## Graphical user interface, text, application Description automatically generated

## Text Description automatically generated

## Wordcloud of Negative Comments

## Graphical user interface, text, application Description automatically generated

## A picture containing text, newspaper Description automatically generated

**CONCLUSION**

**This project predicts sentiment analysis system for you-tube comments. The basic idea behind sentiment analysis using Vader lexicon. We have analysed 305358 positive sentiments, 260986 neutral and 25030 negative sentiments. Channels with their positive percentage number. We have also scrutinized the channels having 100% positive sentiments and channels with 0% positive comments. Word cloud of most frequent, positive,**

**neutral and negative comments**

**BIBLIOGRAPHY**

* [www.youtube.com](http://www.youtube.com)
* [www.kaggle.com](http://www.kaggle.com)
* [www.google.com](http://www.google.com)
* www.github.com
* jupyter
* Faculties