International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 04 | Apr 2022 www.irjet.net p-ISSN: 2395-0072

# Social Media Content Analyser

**Bhushan Pradhan[1], Sreelakshmi Nair[2], Satyajeet Suryawanshi[3], Mihir Tillu[4], Prof. Sagar Kulkarni[5]**

*[1,2,3,4]UG Student, Dept. of Computer Engineering, Pillai College of Engineering, New Panvel, India*
*[5]Assistant Professor, Dept. of Computer Engineering, Pillai College of Engineering, New Panvel, India*

---

**Abstract -** *Social media analysis is the process of collection of information from social media platforms and analyzing patterns present in those data which would help businesses to make effective decisions for their growth. In this project, we have implemented the concepts of Machine Learning, Natural language processing, BDA and Data Science. Analysis of social media data like followers, likes, comments etc from platforms like Facebook, Instagram, Twitter, Youtube. The project will be a web application and the end users can be business users, influencers and personal users who want to enhance their profile. The features of the project are post scheduling which is a system that will automate the process of posting on a particular date and time. Campaign generation is a feature where users can create a series of posts to post it simultaneously. Data is analyzed using big data analytics concepts such as Machine Learning, NLP. Input data from Social media API includes likes, followers/subscribers, number of comments, views. To understand consumer's satisfaction and the feedback it is important to understand what consumers are thinking about the company. Consumer's sentiments about the business posts is detected using sentiment analysis of comments. Using APIs Hashtags are extracted from social media according to input tags given by the consumer for the selected post.*

***Key Words*: APIs, Natural language processing, sentiment analysis, Hashtags,SVM, Naive Bayes**

## 1. INTRODUCTION

For effective business decisions it is necessary to gather and find meaning in data gathered from various social media channels. This project mostly focuses on collecting information from social media API and analyzing patterns from the dataset which helps businesses to make effective conclusions. It involves figuring out users' social media activities that are influencing your business results. In this project, analysis of social media data like followers, Likes, Comments from platforms like Instagram, Youtube, etc. will be done.

## 2. LITERATURE REVIEW

### A. Regionalization of social interactions

This paper takes into consideration various dimensions of social media data like geographical information and combines it with geo self organizing maps to identify similar regions where social interaction in cities takes place along with identifying locations for new points of interest from unstructured data generated in social media.

### B. Towards deep learning prospects: Insights for Social Media Analytics

This paper provides an overview of existing deep learning methods which are used to retrieve useful information for social media analytics. Some application domains of social media such as user behavior analysis, sentiment analysis are also discussed where deep learning methods are incorporated. It highlights the pros and cons of existing techniques and also research challenges and future trends.

### C. Comparative study of Twitter Sentiment On COVID - 19 Tweet

This paper is about the analysis of twitter comments. The comments are extracted and sentiment analysis is performed on it. After preprocessing of the text three models are used to train and test the data and provide the classification. They have used BERT, Logistic regression and VADER. They used a data set of about 25000 lines. They achieved accuracy of about 83% using logistic regression, 92% using BERT model and 88% using VADER model.

### D. Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler: Twitter

This paper is based on comparative study of three algorithms Naive Bayes, SVM and KNN. Naive Bayes gave 75.58% accuracy, SVM gave 63.99% accuracy and KNN gave 73.34% accuracy. Preprocessing is done using NLTK python libraries.

### E. Survey on data analysis in social media: A practical application aspect

This survey paper is a study of previous papers and the applications on practical grounds. A commonly used pipeline was considered for building social media apps and studied analysis techniques such as topic analysis, time series analysis, sentiment analysis, and network analysis. Impact analysis was done on three different areas including disaster management, healthcare, and business.

Existing challenges were studied and solutions were given in terms of data privacy, 5G wireless network, and multilingual support. Various practical applications were discussed related to business value, healthcare and disaster management.

### 2.1 SUMMARY OF RELATED WORK

The summary of methods used in literature is given in Table 1.

Table 1 Summary of literature review

| Literature | Information |
|---|---|
| Achilleas Psyllidis and Jie Yang et al. 2018[1] | Multiple dimensions like spatial, temporal, topical, and demographic information from social media data |
| Malik Khizar Hayat and Ali Daud,et al. 2019 [2] | Behavior analysis, business analytics, sentiment analysis which helps to understand the parameters to take into consideration. |
| Anu J Nair Aadithya Vinayak Veena G,et al. 2021 [3] | Research on sentiment analysis using BERT,Logistic regression and VADER. |
| Meylan WongkarApriandy Angdresey , et al. 2019[4] | Research on Sentiment analysis using Naive Bayes,SVM and KNN. |
| Qixuan Hou, Meng Han, Zhipeng Cai, et al. 2020 [5] | Detailed analysis on data visualization and study on social media analysis |

## 3. PROPOSED WORK

We have used social media (instagram and youtube) api as a source of data and SVM and NB for sentiment analysis.

### 3.1 SYSTEM ARCHITECTURE

The system architecture is given in Figure 1. Each block is described in this Section.

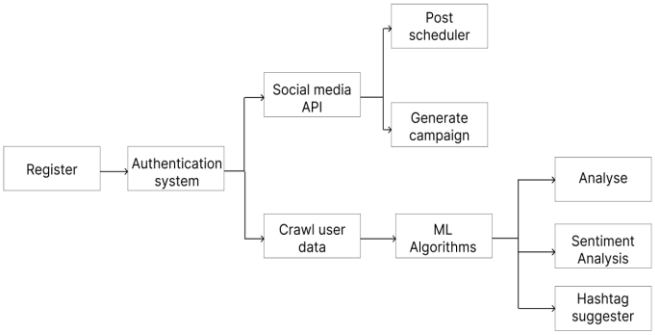


Fig. 1 Proposed system architecture

***A. Analyze Description:*** This part is used to analyze the followers/likes/impressions of each user posts of social media accounts. This will be done with the help of data visualization tools & techniques including statistics. This will be the main module of the project as users can monitor their social media growth in a single page of their multiple social media pages.

***B. Post Scheduler Description:*** This is a functionality which allows users to schedule a post on a certain day and time automatically. This functionality is implemented by using python schedule library. This library simply allows us to run certain pieces of code at a particular time. By taking input from users for this time and simply using api queries which will shoot according to that time will post it into the social media account.

***C. Generate Campaign Description:*** This module will focus on the business aspect of the project where generating campaigns for company products or events is an integral part. A campaign is a planned sequence of activities and processes which promote an individual product, service or resources. This will be done with the help of Social Media API endpoints that will be used to help advertise on social media platforms. Campaigns generation and advertising on multiple social media platforms will be monitored in a single page with the help of this module.

***D. Sentiment analysis Description:*** Sentiment analysis is an important aspect of social media. Consumers' sentiments about posts from business accounts will be analyzed. Sentiments are obtained using the dataset we get from the API. Comments and posts will be considered for the analysis. This analysis is possible using NLP and ML Algorithms like Naive Bayes or SVM to classify data and visualize that data into useful information.

***E. Hashtags Suggester Description:*** Clustering algorithms like k-means clustering will be used to cluster popular Hashtags according to input tags given by the consumer for the selected post. The system will ask the user for the input, the suggester will suggest the most trendy hashtag using the social media API. Another way to generate

hashtags is by simply asking the user to describe the image in key words and the application will generate hashtags.This feature is implemented with the help of natural language processing. The NLTK python library is used for text preprocessing and the hashtag is generated using a simple python script.To generate hashtag using image file, image will be taken as input file and using google vision api and all the words related to image will be obtained. By using those words hashtags are generated.

## 4. CONCLUSION

A web application where analysis of social media data like youtube, instagram data is done and data visualization is performed. Along with that, post creation is implemented due to which posts can be created through our web app. Campaign generator is implemented where pre defined templates are available. Hashtag generator is implemented which will be helpful during creation of post and lastly sentiment analysis of users' comments in a particular post.

## REFERENCES

[1] Achilleas Psyllidis, Jie Yang, "Regionalization of Social Interactions and Points-of-Interest Location Prediction With Geosocial Data", June 2018.

[2] Malik Khizar Hayat, Ali Daud, "Towards Deep Learning Prospects: Insights for Social Media Analytics", Mar 2019.

[3] Meylan Wongkar, Apriandy Angdresey, "Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler: Twitter", Feb 2020.

[4] Qixuan Hou, Meng Han, Zhipeng Cai, "Survey on data analysis in social media: A practical application aspect", Nov 2020.

[5] Anu J Nair, Aadithya Vinayak, Veena G, "Comparative study of Twitter Sentiment On COVID - 19 Tweet", May 2021.

[6] Nargiza Bekmamedova, Graeme Shanks. "Social Media Analytics and Business Value: A Theoretical Framework and Case Study."2014 47th Hawaii International Conference on System Sciences. IEEE, 2014.

[7] Alexander Semenov, Jari Veijalainen. "A generic architecture for a social network monitoring and analysis system."2011 14th International Conference on Network-Based Information Systems. IEEE, 2011.

## BIOGRAPHIES

**Bhushan Pradhan** is an undergraduate student of mumbai university. His area of interests are NLP and software development.

**Sreelakshmi Nair** is an undergraduate student of mumbai university. Her area of interests is data science and NLP.

**Satyajeet Suryawanshi** is an undergraduate student of mumbai university. His area of interests are data mining and NLP .

**Mihir Tillu** is an undergraduate student of mumbai university. His area of interests are UI/UX and NLP.