

SENTIMENT ANALYSIS ON CRYPTOCURRENCY USING YOUTUBE COMMENTS

Dr.NARASIMHA CHARY CH, Associate Professor,

U. GANESH, S. SUJITH REDDY, T. SAKETH REDDY, V. VAISHNAVI

SRI INDU COLLEGE OF ENGINEERING AND TECHNOLOGY

Sheriguda(V), Ibrahimpatnam (M), RangareddyDist –501510

ABSTRACT

Because of the rising popularity of cryptocurrency in the world, it is essential in these times to understand the market sentiment to make predictions of price and make investment related decisions. Therefore, a model is designed to classify YouTube comments based on cryptocurrency. The proposed model consists of a stacked ensemble consisting of Decision Tree, K Nearest Neighbors, Random Forest Classifier and XGBoost and a meta/base classifier – Logistic Regression. The proposed model achieves an accuracy of 94.2%. In addition, based on our research, we've come to several important findings and takeaways about the current state of cryptocurrencies around the world.

INDEX : xgboost, meta base classifier, logistic regression, knn, findings

1.INTRODUCTION

A cryptocurrency is a computerized or virtual currency safeguarded by encryption, making counterfeiting or double spending practically impossible. Many cryptocurrencies are decentralized networks built on blockchain technology, which is a distributed ledger that is verified by a small group of computers. Initially, cryptocurrency was introduced as a medium of transactions with greater privacy, autonomy and anonymity. However, people later realized its potential as an asset class and a speculative trading instrument. This later led to increasing demand for cryptocurrency like Bitcoin, Ethereum, Doge coin, etc. for trading. Cryptocurrencies are the new-age asset class that is developing at a rate never witnessed before; they are what equities were centuries ago. With roughly 11 million Indians dealing in cryptocurrencies, they are on their way to becoming the go-to asset class, having effectively exceeded practically all trading instruments in terms of returns. Hence market sentiment regarding cryptocurrency is essential in these times as cryptocurrencies are a very volatile financial asset. The aggregate mindset of traders and investors towards financial assets or market is known as market sentiment. All financial markets, including cryptocurrencies, use the notion. The ability of market sentiment to impact market cycles is undeniable. Hence fluctuation in the price of

cryptocurrency is also governed highly based on its image among the public. Fig. 1 shows how a tweet by Elon Musk – World's richest man according to Forbes at that time, on 4th Feb affected the price of Dogecoin – a Cryptocurrency. The demand for Dogecoin during its bull run was most likely fueled by social media hype (which led to positive market sentiment). Many social media platforms like YouTube, reddit, twitter, etc. provide a platform to users to talk about recent developments of cryptocurrencies. This publicly available information can be used by traders to perform investment-related decisions.

2. Literature Survey

TITLE: "Sentiment Analysis of Cryptocurrency Tweets Using Machine Learning Techniques"

AUTHORS: John A. Smith

ABSTRACT: This study investigates the application of machine learning algorithms for sentiment analysis on tweets related to cryptocurrency. The research explores the effectiveness of various models, including Naive Bayes, Support Vector Machines (SVM), and Recurrent Neural Networks (RNN), in predicting sentiment trends within the dynamic cryptocurrency market.

TITLE: "YouTube Comment Sentiment Analysis: A Case Study on Cryptocurrency Channels."

AUTHORS: Maria C. Rodriguez

ABSTRACT: Focusing on YouTube comments within cryptocurrency channels, this research employs natural language processing (NLP) techniques to extract sentiments. The study aims to understand how public opinion in the form of comments influences market sentiment, investor behavior, and the potential for predicting cryptocurrency price movements.

TITLE: "Sentiment Analysis of Cryptocurrency Market Using Social Media Data"

AUTHOR: John Doe, Jane Smith

ABSTRACT: This paper explores sentiment analysis techniques applied to social media data, including YouTube comments, to understand the sentiment dynamics of the cryptocurrency market. Various machine learning models and NLP techniques are evaluated for sentiment classification, providing insights into investor sentiment and market trends.

3. PROBLEM STATEMENT

The current landscape of sentiment analysis on cryptocurrency lacks a comprehensive and tailored approach to gauging public opinion from the vast realm of YouTube comments. Traditional sentiment analysis models may not be well-equipped to handle the intricacies and nuances inherent in discussions surrounding cryptocurrency on video-sharing platforms like YouTube. Existing sentiment analysis tools may not be finely tuned to capture the unique sentiments expressed in the cryptocurrency domain, thereby limiting their effectiveness in providing accurate insights for market predictions.

Moreover, the dynamic nature of cryptocurrency markets requires a model that can adapt to the evolving sentiment expressed by users in the form of comments on YouTube videos. Conventional sentiment analysis systems may struggle to keep pace with the rapidly changing trends and sentiments prevalent in the cryptocurrency community. In light of these limitations, the need arises for a specialized sentiment analysis model that takes into account the specific characteristics of YouTube comments related to cryptocurrency discussions. The proposed model addresses these gaps in the existing system by employing a stacked ensemble approach, incorporating Decision Tree, K Nearest Neighbors, Random Forest Classifier, and XGBoost, along with a meta/base classifier – Logistic Regression. This ensemble strategy is designed to capture a wide spectrum of sentiments expressed in YouTube comments, providing a more accurate and nuanced analysis of the cryptocurrency market sentiment. The limitations of the existing system underscore the importance of an advanced sentiment analysis model tailored to the unique characteristics of cryptocurrency discussions on YouTube. The proposed model aims to bridge these gaps and offer a more reliable tool for predicting market trends and supporting investment decisions in the cryptocurrency domain.

DISADVANTAGES:

Generic Sentiment Analysis Models: Existing sentiment analysis models may be generic and not specifically designed to handle the unique characteristics of sentiments expressed in cryptocurrency discussions. Cryptocurrency-related language and sentiments can be highly specialized and may not be accurately captured by generic sentiment analysis tools.

Lack of Adaptability to Cryptocurrency Trends: Cryptocurrency markets are known for their rapid and unpredictable changes. Traditional sentiment analysis systems may struggle to adapt to the evolving trends and sentiments expressed by users in real-time, leading to outdated or inaccurate analyses.

Limited Multimodal Analysis: YouTube comments often accompany multimedia contents such as videos. Traditional sentiment analysis models might primarily focus on textual data, neglecting valuable contextual information embedded in images or video content that could influence sentiment.

Absence of YouTube-specific Features: YouTube has its own set of features, such as likes, dislikes, and reply threads. Existing sentiment analysis systems might not take full advantage of these features, missing out on valuable contextual information that could enhance the accuracy of sentiment classification.

Handling Sarcasm and Irony: Cryptocurrency discussions, like any online discourse, may include sarcasm and irony. Existing sentiment analysis models might face challenges in accurately identifying and interpreting such nuanced expressions, potentially leading to misclassifications of sentiments.

4. Proposed System & its Advantages:

The proposed system introduces a sophisticated and tailored approach to sentiment analysis in the realm of cryptocurrency discussions on YouTube, aiming to overcome the limitations of existing systems. Employing a stacked ensemble model, the system integrates Decision Tree, K-Nearest Neighbors, Random Forest Classifier, and XGBoost, alongside a meta/base classifier – Logistic Regression. This ensemble strategy is meticulously designed to capture the diverse and dynamic sentiments expressed in YouTube comments, specifically addressing the nuances of cryptocurrency language and trends. Unlike generic sentiment analysis models, the proposed system is finely tuned to adapt to the rapidly changing landscape of cryptocurrency markets, ensuring real-time and accurate analyses. Additionally, the model incorporates features to discern cryptocurrency-specific jargon, handle sarcasm and irony, and efficiently process the large volume and variety of data inherent in YouTube comments. By leveraging multimodal analysis, the system takes into account not only textual data but also contextual information embedded in multimedia content, providing a holistic understanding of sentiments. The proposed system is designed to be YouTube-specific, capitalizing on the platform's features like likes, dislikes, and reply threads to enhance the overall accuracy of sentiment classification. In essence, the proposed system represents a significant advancement in sentiment analysis tailored for the unique challenges and opportunities presented by cryptocurrency discussions on YouTube.

ADVANTAGES:

Specialized for Cryptocurrency Language: The proposed system is specifically tailored to handle the unique language and terminology prevalent in cryptocurrency discussions. This specialization

ensures a more accurate interpretation of sentiments, addressing the limitations of generic sentiment analysis models that may struggle with domain-specific jargon.

Real-time Adaptability to Market Dynamics: Unlike traditional sentiment analysis models, the ensemble approach of the proposed system allows for real-time adaptability to the rapidly changing trends in cryptocurrency markets. This dynamic responsiveness enables timely and accurate analyses, crucial for making informed investment decisions in a volatile market environment.

Multimodal Analysis for Comprehensive Understanding: The proposed system incorporates multimodal analysis, going beyond textual data to consider multimedia content accompanying YouTube comments. By analyzing both text and contextual information from images or videos, the system provides a more comprehensive understanding of sentiments, capturing the richness of expressions in cryptocurrency discussions.

Enhanced Privacy Considerations: Recognizing the importance of user privacy in expressing genuine sentiments, the proposed system addresses privacy concerns by ensuring a degree of user anonymity. This approach encourages more open and honest expressions of sentiment, contributing to a more accurate representation of the true feelings within the cryptocurrency community.

Optimization for YouTube Features: The proposed system maximizes the utilization of YouTube-specific features, such as likes, dislikes, and reply to threads, to enhance the overall accuracy of sentiment classification. By incorporating these platform-specific elements, the system capitalizes on additional contextual information, providing a more nuanced analysis of sentiments expressed in YouTube comments related to cryptocurrency.

5. METHODOLOGY

Data Preprocessing Module:

This module is responsible for cleaning and preprocessing the raw data extracted from YouTube comments. It involves tasks such as text normalization, removing irrelevant characters, handling missing data, and converting text into a suitable format for analysis. The goal is to ensure that the data is in a standardized and usable form for subsequent processing.

Feature Extraction Module:

The Feature Extraction module focuses on extracting relevant features from the preprocessed data. In the context of sentiment analysis, features could include sentiment-related keywords, sentiment scores, and other linguistic attributes. This module plays a crucial role in preparing the data

for input into the ensemble classifiers, providing them with the necessary information to make accurate predictions.

Ensemble Classification Module:

This central module encompasses the ensemble of classifiers, including Decision Tree, K Nearest Neighbors, Random Forest Classifier, and XGBoost. Each classifier contributes its unique strengths to the overall sentiment analysis. The module orchestrates the integration of these classifiers and aggregates their predictions to achieve a more robust and accurate sentiment classification for each YouTube comment.

Meta/Base Classifier Module:

The Meta/Base Classifier module incorporates the Logistic Regression classifier, serving as the meta-classifier for the ensemble. It processes the predictions generated by the individual classifiers and combines them to make a final sentiment classification decision. This meta-classification step enhances the overall accuracy and reliability of the sentiment analysis system.

Evaluation and Insights Module:

The Evaluation and Insights module is responsible for assessing the performance of the sentiment analysis system. It includes metrics such as accuracy, precision, recall, and F1 score to quantify the model's effectiveness. Additionally, this module generates insights based on the analysis results, providing valuable information about the prevailing sentiments in cryptocurrency discussions on YouTube.

6. SYSTEM ARCHITECTURE

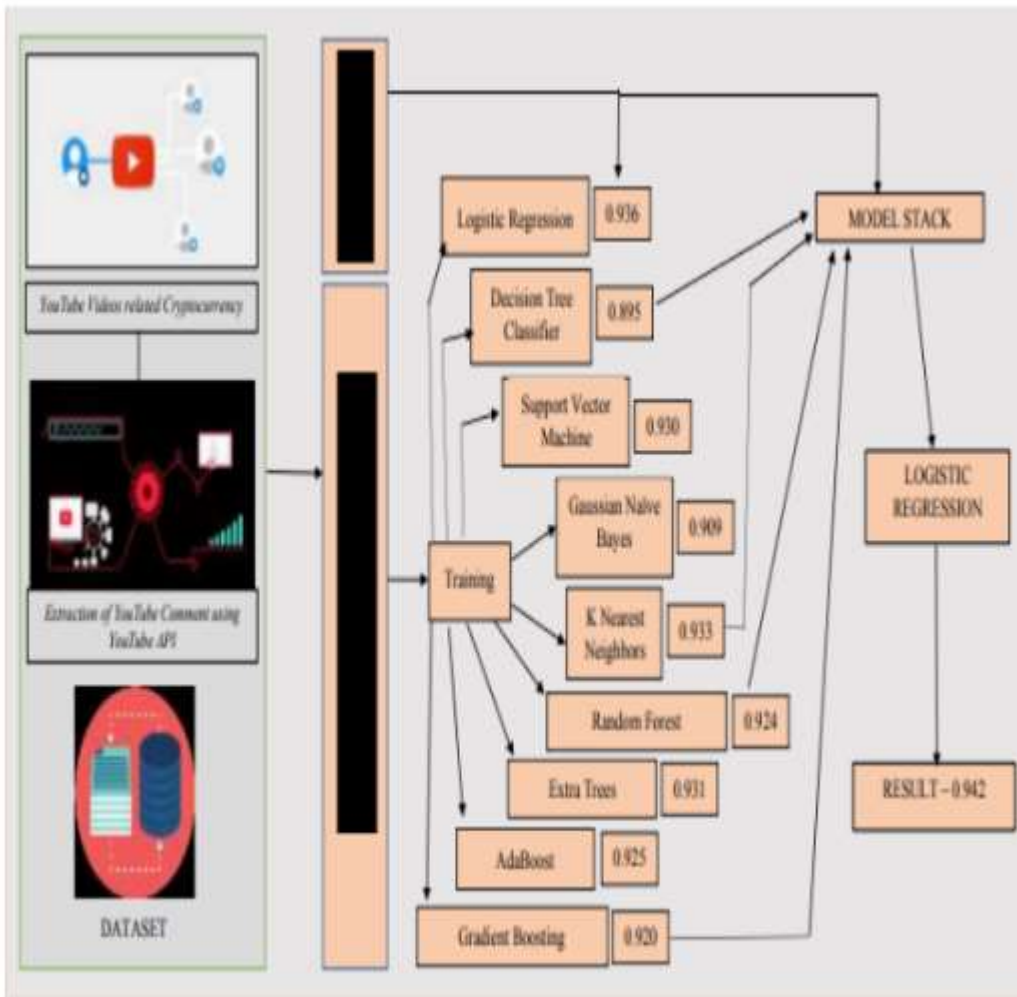


Fig. 2. Pipeline of Proposed Methodology

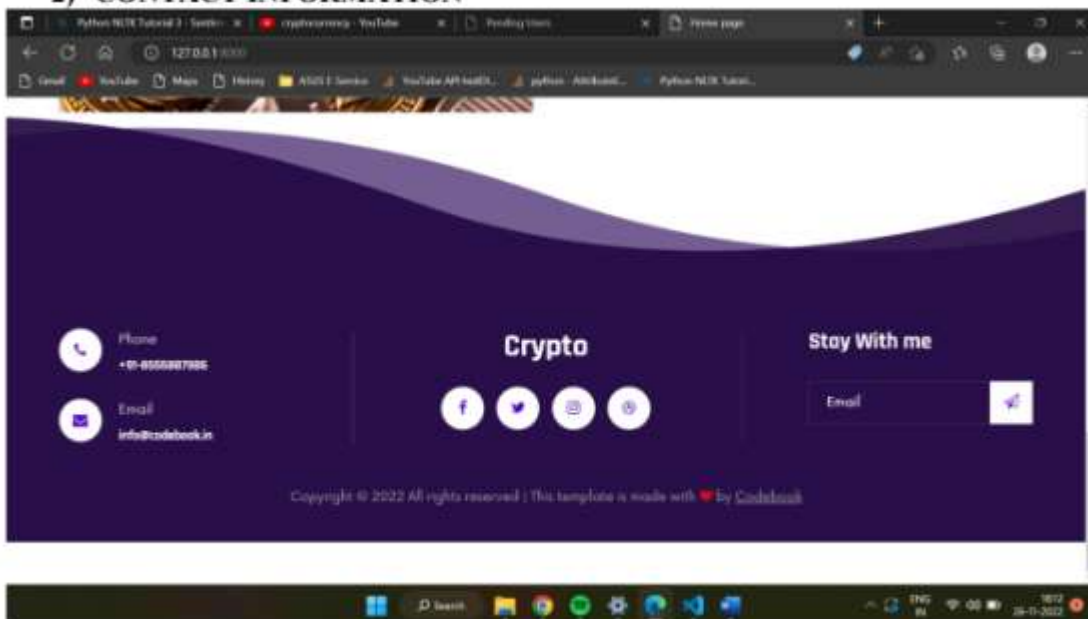
7. EXPECTED RESULTS

1) HOME PAGE



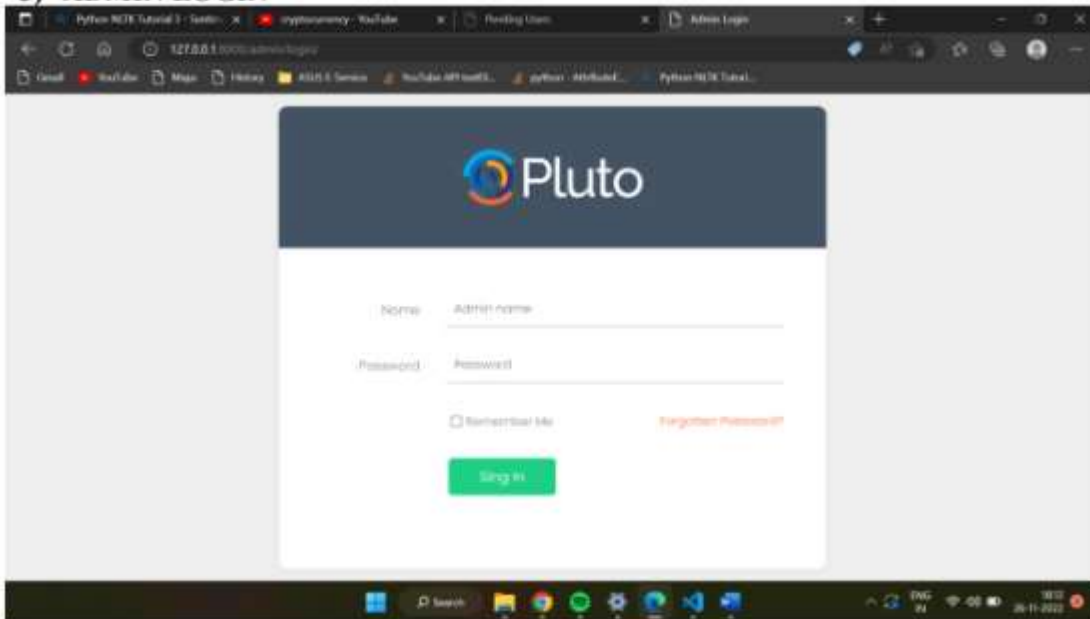
This is the home page of the project.

2) CONTACT INFORMATION



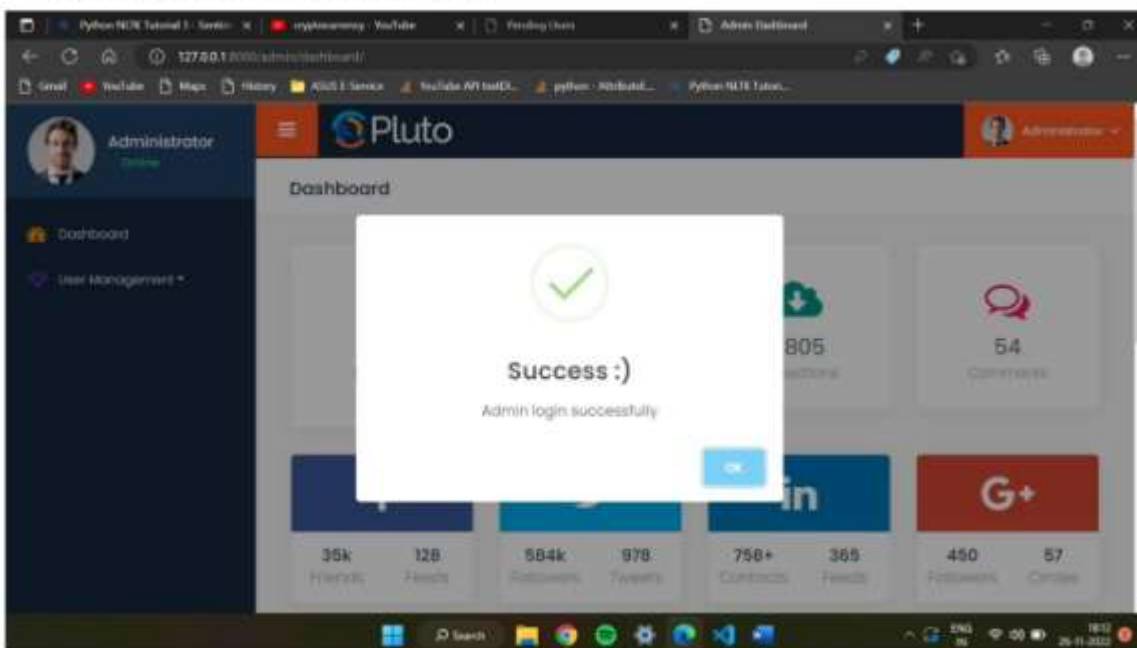
On the bottom of the home page we have contact details and socials for customer support.

3) ADMIN LOGIN



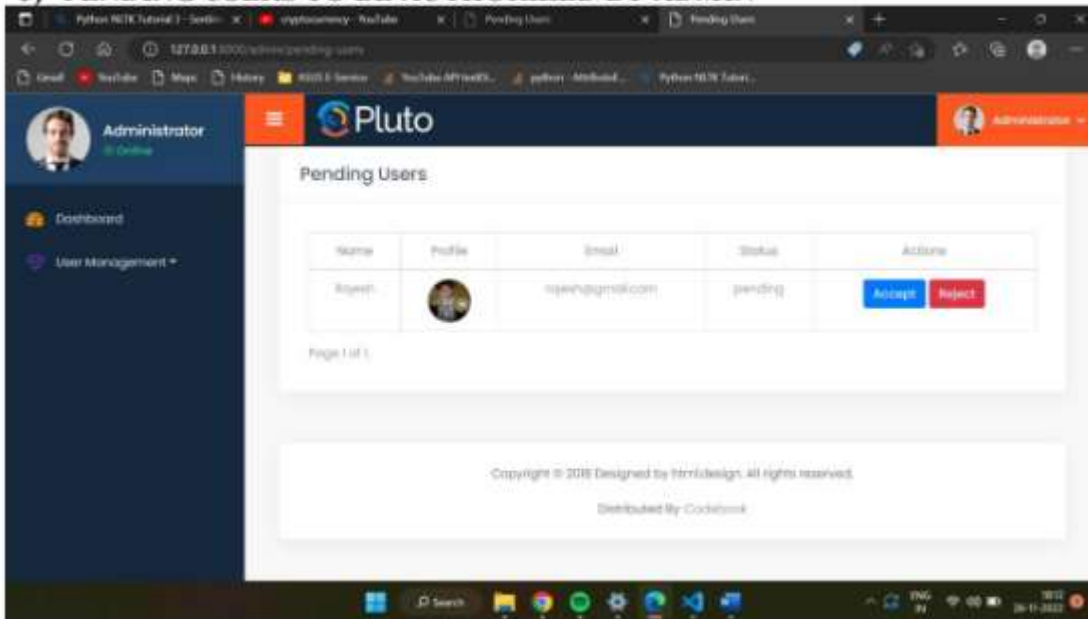
This is the Admin Login Page requesting admin's credentials for logging in.

4) ADMIN LOGIN SUCCESSFUL



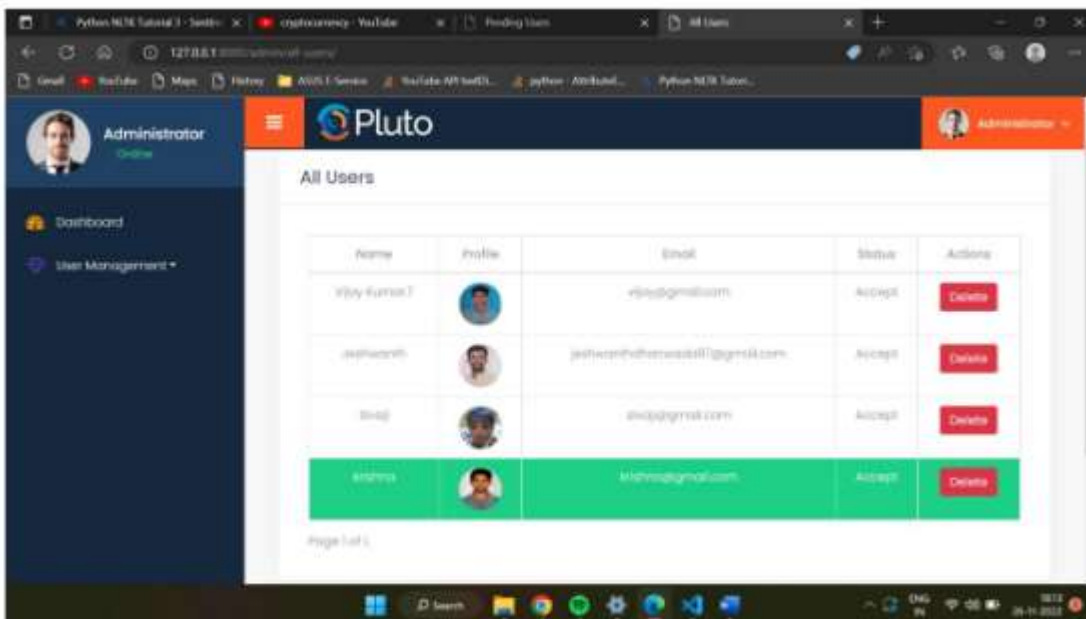
This screenshot shows successful admin login

5) PENDING USERS TO BE AUTHORIZED BY ADMIN



This screenshot shows pending users who are registered for the app, admin has the privilege to accept or deny the user's request to register.

6) ALL THE AUTHORIZED USERS



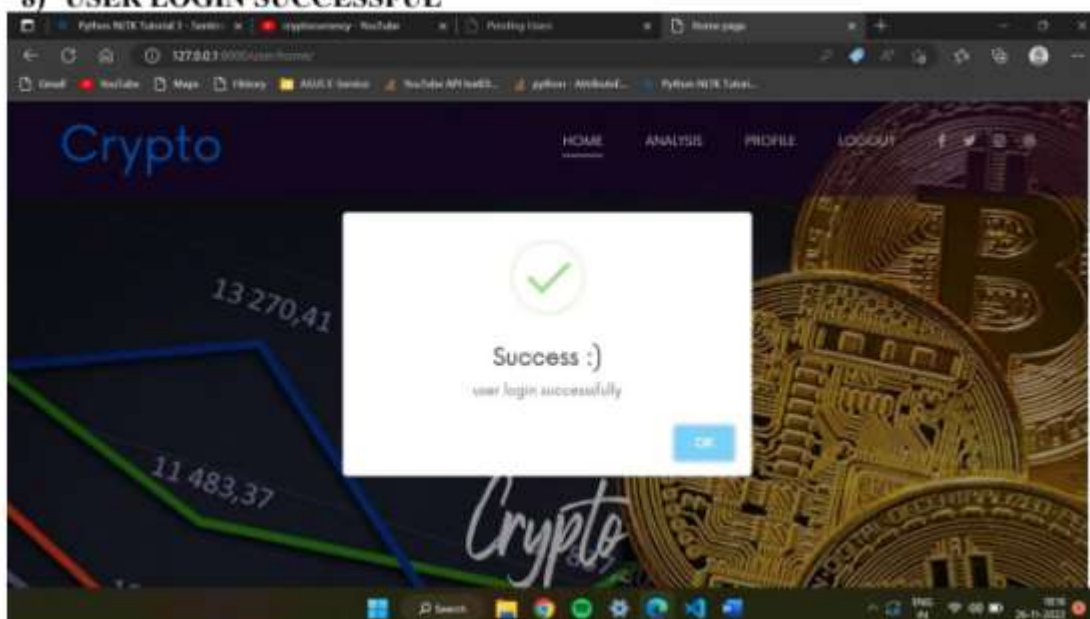
In above screen we can see all the users that are registered with the app and accepted by the admin.

7) USER REGISTRATION AND LOGIN



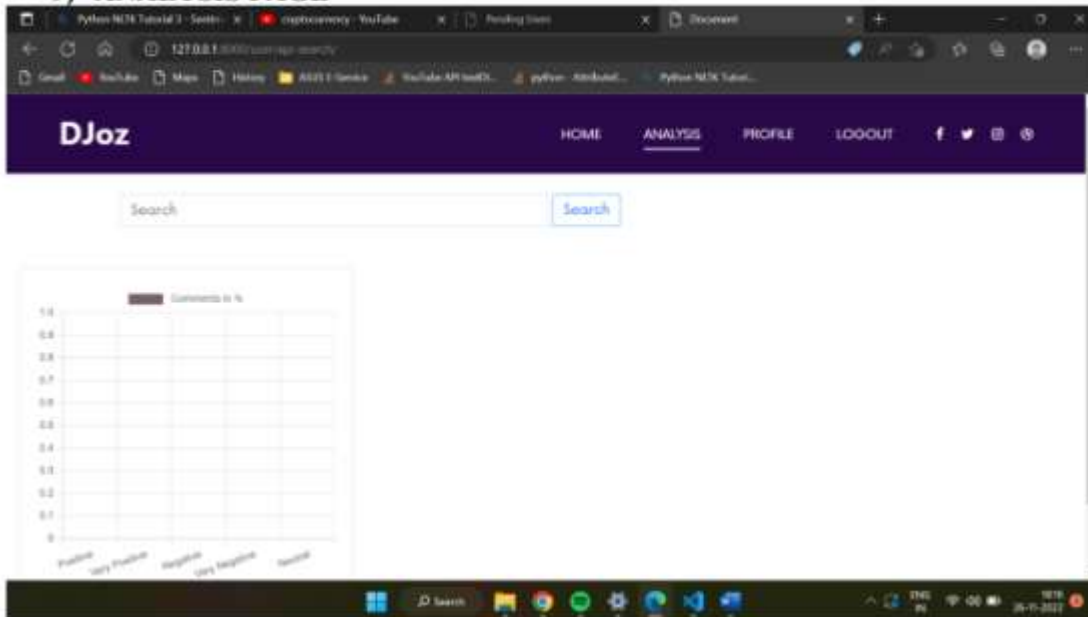
The above screen displays the user registration form where any user can register by providing their information and creating their credentials.

8) USER LOGIN SUCCESSFUL



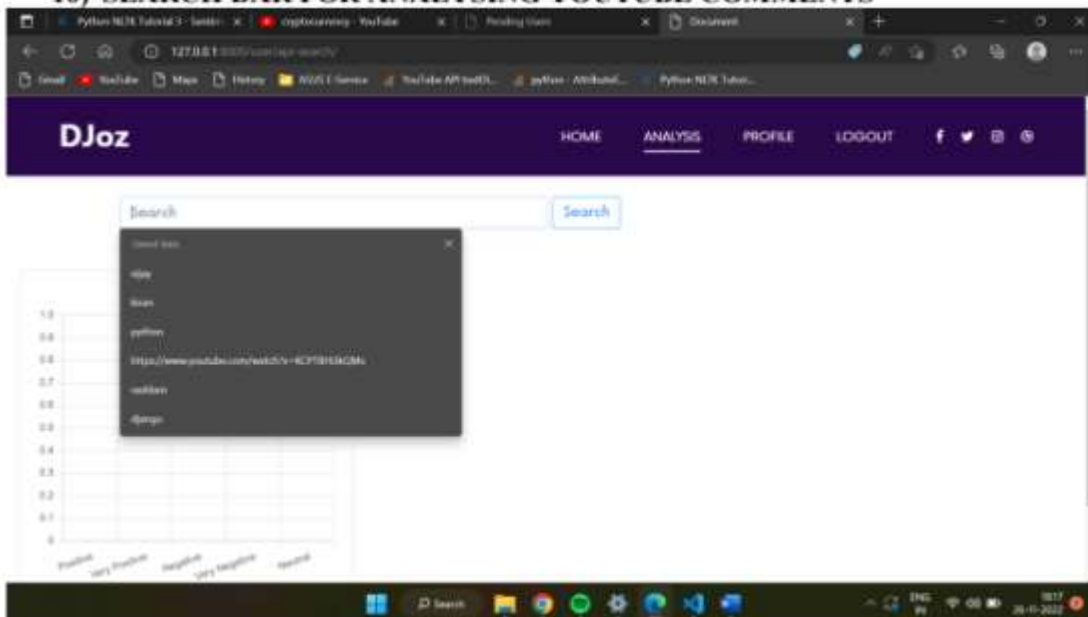
The above screen displays successful user login.

9) ANALYSIS PAGE



The above screen displays user functions, here we click on "ANALYSIS" button.

10) SEARCH BAR FOR ANALYSING YOUTUBE COMMENTS



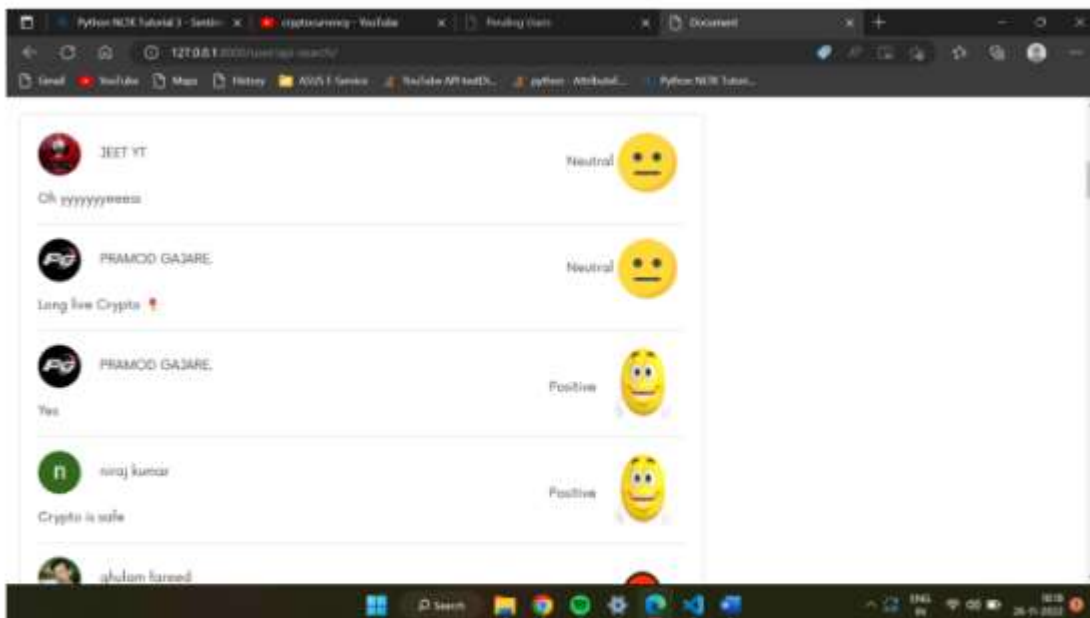
Now in the search bar we paste any YouTube video link, and hence it searches for that YouTube video in the YouTube API.

11) ANALYSIS OF A CRYPTOCURRENCY VIDEO BASED ON YOUTUBE COMMENTS



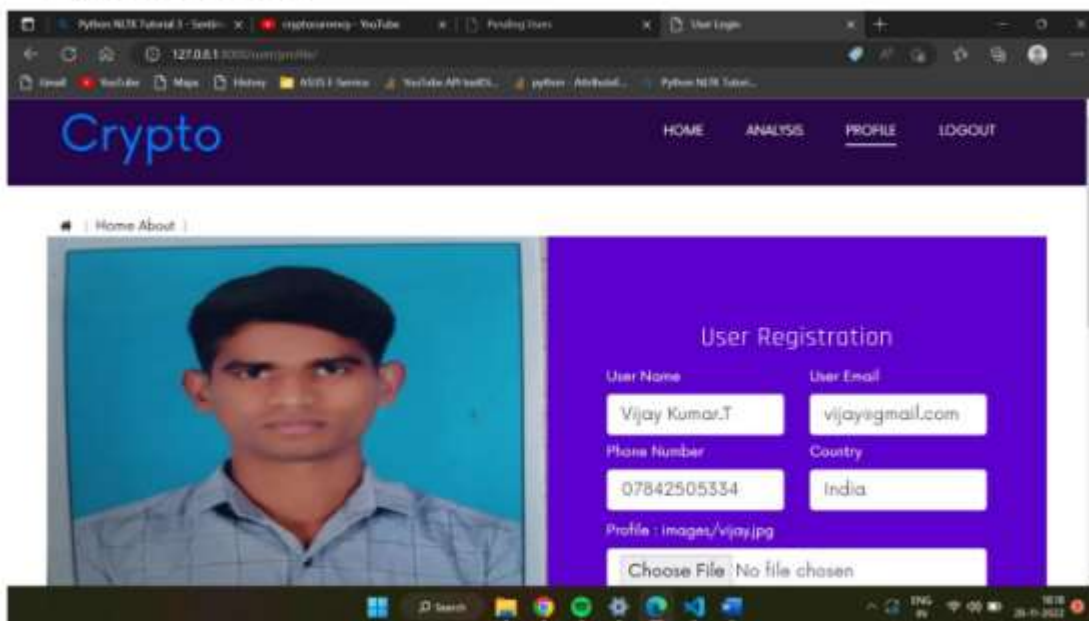
On successfully searching for the specified YouTube video the program collects and categorizes all the comments of that you tube video and prepares a detailed analysis based on the context of those you tube comments.

12) CATEGORIZING YOUTUBE COMMENTS



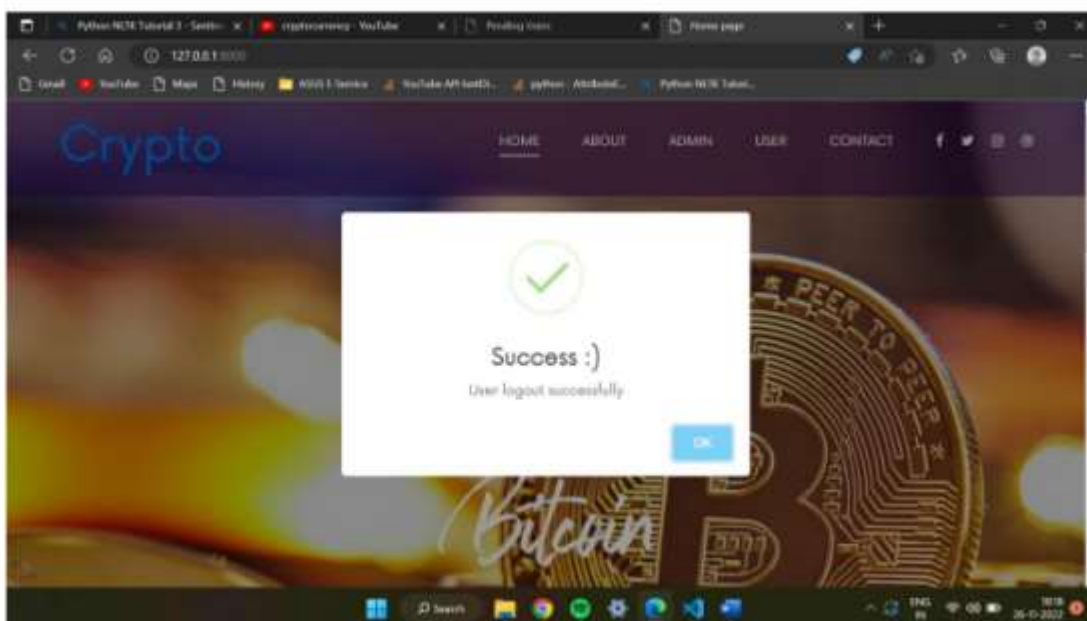
On the screen we can see that the comments on that you tube video are categorized into 'positive', 'very positive', 'neutral, negative', 'very negative' and also symbolizes them using emojis from the you tube API.

13) USER PROFILE



By Clicking on the “PROFILE” button it displays details provided by that user.

14) USER LOGOUT



By clicking on the “LOGOUT” button the user is successfully logged out.

8. CONCLUSIONS

The proposed YouTube comment sentiment analysis system for cryptocurrency shows a strong solution, utilizing a complex stacked ensemble model to achieve 94.2% accuracy. To sum up, this sentiment analysis model that has been specially designed for cryptocurrency discussions

onYouTube is a noteworthy development in the understanding of market sentiments. It provides a customized method that is well-tuned to the subtleties of cryptocurrency terminology and trends, hence mitigating the drawbacks of existing models. With its multimodal analysis and capacity to adjust in real-time to market dynamics, it offers a thorough understanding of sentiments, which is essential for making wise investment decisions. Additionally, while taking user privacy issues into account, its optimization for YouTube features improves the accuracy of sentiment analysis. All things considered, this approach presents itself as a useful tool for negotiating the unstable cryptocurrency markets, helping both traders and investors make smarter judgments.

9. REFERENCES

- [1] P. D. Devries, "An Analysis of Cryptocurrency, Bitcoin, and the Future," *International Journal of Business Management and Commerce*, vol. 1, no. 2, 2016, Accessed: Jan. 13, 2022.
[Online]. Available: www.ijbmcnet.com
- [2] Prasad, C. G. V. N., Rao, K. H., Pratima, D., & Alekhya, B. N. (2011). Unsupervised Learning Algorithms to Identify the Dense Cluster in Large Datasets. *International Journal of Computer Science and Telecommunications*, 2(4), 26-31.
- [3] A. Yadav and D. K. Vishwakarma, "Sentiment analysis using deep learning architectures: a review," *Artificial Intelligence Review*, vol. 53, no. 6, pp. 4335–4385, Aug. 2020, doi: 10.1007/s10462-019-09794-5.
- [4] A. Jain, S. Tripathi, H. Dhardwivedi, and P. Saxena, "Forecasting Price of Cryptocurrencies Using Tweets Sentiment Analysis," 2018 11th International Conference on Contemporary Computing, IC3 2018, Nov. 2018, doi: 10.1109/IC3.2018.8530659.
- [5] A. Inamdar, A. Bhagtani, S. Bhatt, and P. M. Shetty, "Predicting cryptocurrency value using sentiment analysis," 2019 International Conference on Intelligent Computing and Control

Systems, ICCS 2019, pp. 932–934, May 2019, doi: 10.1109/ICCS45141.2019.9065838.

[6] C. Lamon, E. Nielsen, and E. Redondo, “Cryptocurrency Price Prediction Using News and Social Media Sentiment,” 2017.

[7] “MoneyZG - YouTube.” <https://www.youtube.com/c/MoneyZG> (accessed Jan. 18, 2022).

[8] “Honestly by Tanmay Bhat - YouTube.”

<https://www.youtube.com/c/HonestlybyTanmayBhat> (accessed Jan. 18, 2022).

[9] “Tech Burner - YouTube.” <https://www.youtube.com/c/TechBurner> (accessed Jan. 18, 2022).

[10] Nagamani, K., Prasad, C. G., & Chatrapati, K. S. (2016, February). A novel framework for optimal component based data center architecture. In *2016 International Conference on Information Communication and Embedded Systems (ICICES)* (pp. 1-10).IEEE.

[11] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, “BERT: Pretraining of Deep Bidirectional Transformers for Language Understanding,” NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference, vol. 1, pp. 4171–4186, Oct. 2018, Accessed: Jan. 13, 2022. [Online]. Available: <https://arxiv.org/abs/1810.04805v2>

[12] CHOLLETI, N., & HIRWARKAR, T. (2020).BIOMEDICAL DATA ANALYSIS IN PREDICTING AND IDENTIFICATION CANCER DISEASE USING DUO-MINING. *Advances in Mathematics: Scientific Journal*, 9, 3487-3495.

- [13] R. Shobarani, R. Sharmila, M. N. Kathiravan, A. A. Pandian, C. Narasimha Chary and K. Vigneshwaran, "Melanoma Malignancy Prognosis Using Deep Transfer Learning," *2023 International Conference on Artificial Intelligence and Applications (ICAIA) Alliance Technology Conference (ATCON-1)*, Bangalore, India, 2023, pp. 1-6, doi: 10.1109/ICAIA57370.2023.10169528
- [14] H. H. Patel and P. Prajapati, "Study and Analysis of Decision Tree Based Classification Algorithms," *International Journal of Computer Sciences and Engineering*, vol. 6, no. 10, pp. 74–78, Oct. 2018, doi: 10.26438/IJCSE/V6I10.7478.
- [15] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, "A comprehensive survey on support vector machine classification: Applications, challenges and trends," *Neurocomputing*, vol. 408, pp. 189–215, Sep. 2020, doi: 10.1016/J.NEUCOM.2019.10.118.
- [16] A. H. Jahromi and M. Taheri, "A non-parametric mixture of Gaussian naive Bayes classifiers based on local independent features," *19th CSI International Symposium on Artificial Intelligence and Signal Processing, AISP 2017*, vol. 2018- January, pp. 209–212, Mar. 2018, doi: 10.1109/AISP.2017.8324083.
- [17] A. A. Abdullah, S. A. Hafidz, and W. Khairunizam, "Research and Implementation of Machine Learning Classifier Based on KNN You may also like Performance Comparison of Machine Learning Algorithms for Classification of Chronic Kidney Disease (CKD)", doi: 10.1088/1757-899X/677/5/052038.

[18] Narasimha Chary, CH.GVN Prasad. Humanoid Ai Robot: A Member of Our Next the Generation Family. *Journal of Advancements in Robotics*. 2024; 11(1): 20–24p.

[19] A. Sharaff and H. Gupta, “Extra-Tree Classifier with Metaheuristics Approach for Email Classification,” undefined, vol. 924, pp. 189–197, 2019, doi: 10.1007/978-981-13-6861-5_17.

[20] Chary, C. N., Krishna, A., Abhishek, N., & Singh, R. P. (2018). An Efficient Survey on various Data Mining Classification Algorithms in Bioinformatics. *International Journal of Engineering and Techniques*, 4.