

# CSE4022 NATURAL LANGUAGE PROCESSING

**J-COMPONENT PROJECT**

**FALL SEMESTER 2023-2024**

# SOCIAL MEDIA COMMENTS CATEGORIZATION, PRIORITIZATION AND SENTIMENT ANALYSIS

# FACULTY NAME – PROF. PRIYA G

**SUBMITTED BY –**

### SWETA CHANDRASEKHAR - 20BCE2625

### KIRTHIKA V – 20BCE2628

### MYLIE MUDALIYAR – 20BCE2661

**TABLE OF CONTENTS**

|  |
| --- |
| Acknowledgement 3 |
| Abstract 4 |
| Problem Statement 4 |
| Objective 5 |
| Hardware and Software  Requirements 5 |
| Introduction 6 |
| Related Work 7 |
| Proposed Methodology 17 |
| Results & Discussions 18 |
| Conclusion 18 |
| Future Work 19 |
| References 20 |

**ACKNOWLEDGEMENT**

This is to state on the record that a report of this stature on the topic “Social Media Comments Categorization, Prioritization and Sentiment Analysis” was made possible because of a concerted effort put in by the group members who succored and motivated throughout and due to the availability of resources that helped us in forming the contents. In this regard, we would like to thank our Professor Dr. Priya G, an eminent faculty member of this reverent institution who has incidentally also been our professor for the course Natural Language Processing (CSE4022) in the Fall Semester 2023. It is her way of teaching and guidance that helped us to work on our project. Her suggestions made us improve our project in each phase.

All the group members have helped whenever someone encountered difficulties and this acknowledgement would be incomplete without them.

Finally, this institution provided us with sufficient resources and allocating the course Natural Language Processing.

**ABSTRACT:**

The proliferation of customer comments on social media platforms presents an engagement challenge for content creators and organizations. Manual re-view methods are inefficient for systematically extracting, analyzing, and re-sponding to large volumes of feedback. This paper proposes a methodology integrating web scraping and K-means clustering to systematically categorize and prioritize responses. Custom web scrapers extract comments from YouTube videos, overcoming limitations of traditional screen scraping. Sen-timent analysis using TextBlob and HuggingFace provides insight into overall video sentiment. The unsupervised K-means algorithm clusters comments based on certain factors, enabling data-driven prioritization. This research highlights the potential of integrating web scraping, machine learning, and clustering techniques to elevate customer satisfaction via prompt, tailored re-sponses to feedback at scale. This methodology can be integrated with an us-er interface, with the content creators uploading the desired video ids and getting relevant comment categories or overall user sentiment. This integrated methodology promises to optimize understanding and response to diverse customer feedback at scale.

**PROBLEM STATEMENT:**

Managing a burgeoning YouTube channel or any social media platform becomes a formidable challenge when inundated with an overwhelming volume of comments and reviews, many of which include numerous questions. The current manual process of sifting through this deluge is inefficient, hindering the ability to identify and respond promptly to customer queries.

To address this, the proposed project aims to revolutionize review and comment management through the integration of a web scraping bot and the application of the K-Means clustering method. This dynamic duo will not only streamline the extraction of valuable information from website content but also categorize feedback based on the number of questions posed. The goal is to enhance the efficiency of response prioritization.

The broader problem is the lack of an effective system to analyze and prioritize customer feedback across e-commerce and video streaming platforms. By incorporating sentiment analysis, organizations can make informed, data-driven decisions to bolster customer satisfaction. The challenge lies in seamlessly integrating these technologies into various platforms for real-time analysis, thus improving outreach to audiences and customers.

Furthermore, there is a need for comprehensive research to gauge the effectiveness of sentiment analysis across diverse industries. Exploring the impact of sentiment analysis on decision-making processes will contribute valuable insights for organizations seeking to optimize customer engagement strategies. The overarching problem statement revolves around devising a solution that not only addresses the current inefficiencies in handling customer feedback but also explores the untapped potential of sentiment analysis in driving strategic decision-making.

**OBJECTIVE:**

• To design a tool that will function as a bot to help you determine which question to answer first by exploiting the properties of web scraping tools.

• To develop a software bot that would contain the scripts.

• To develop a user-friendly and cost-effective way to get work done faster with much less effort of going through and reading all comments.

**HARDWARE AND SOFTWARE REQUIREMENTS:**

**Hardware Requirements –**

* Computer or Server
* CPU
* RAM
* Storage
* Internet Connection

**Software Requirements –**

* Operating Systems
* Python Programming Language
* Web Scrapping Libraries
* Database
* Jupyter Notebook
* Visual Studio Code
* K-Means Clustering Libraries
* Data Analysis and Visualization Tools
* Text Preprocessing Libraries
* Version Control

**INTRODUCTION:**

Crafted for swift response to reviews and comments, a web scraping bot accelerates engagement by swiftly extracting content and data from websites. Diverging from screen scraping, which merely captures on-screen pixels, web scraping delves into the website's underlying HTML code and database, enabling the replication of entire website content elsewhere.

Beyond web scraping, the project harnesses the power of the K-Means clustering method to cluster and categorize comments and reviews. This method prioritizes responses by evaluating the number of questions in each feedback. K-Means Clustering, an Unsupervised

Learning algorithm, organizes unlabeled datasets into distinct clusters. The parameter "K" dictates the number of pre-defined clusters, shaping the clustering process K=2 yields two clusters, K=3 results in three clusters, and so forth.

**RELATED WORK:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Paper** | **Proposed Work** | **Aim** | **Problem Statement** | **Limitations** | **Dataset** | **Results / Accuracy & Evaluation Matrices** |
| Analysis and Prioritization of App Reviews | Data analysis, visualization and prioritization process is carried out by Text Preprocessing, Frequency and Word cloud, Association mining based on correlation, Sentimental analysis based on customers reviews and ratings, And labelling the priority level using Multinomial Naive Bayes | To analyze the crucial feedback present in the app and prioritize the reviews | Since smartphone applications market is highly recognized and is growing rapidly, the app developers has to know what changes they have to make in their application to make it more efficient. this paper helps the developers to get help from the customer's reviews by prioritizing the frequent review's by the end users | Subjectivity: App reviews can be subjective, and the sentiment expressed in a review may not always align with the actual experience of the user.Limited data. biased reviews, privacy concerns etc are the factors which affect the current paper.This incorporates only signle app reviews it should be applicable to other domains such as commerce and hotels | MyTrack | the gaps pertaining in the statements was identified and then the analysis and prioritization of app reviews was resolved. They evaluated usability of tool through app developers using System Usability Scale approach based on the SUS and MyTracks app developers found the tool useful |
| An online review driven method for the prioritization of improvements in hotel services | This paper uses Penalty reward contrast analysis, sentimental analysis, LDA to compute the Prioritization | An online reviews driven method to determine the Prioritization of improvements in hotel services | In every other research papers previously the impact of attribute  types on service improvement but ignore the impact of consumer ex- pectations on customer satisfaction, so in this paper they have improvised the solution for this problem | Since the goal of this article is to provide a way for service improvement, the internet reviews are only superficially screened. Reviews might one day be screened.  by making use of the literature that catalogues false reviews. Second, this study conducts its research using the entire customer population. In actuality, various consumer categories frequently have varying expectations for service characteristics. The impact of various customer types should be taken into account in the future. Thirdly, this essay simply takes its own limited resources into account, not the resources of the market. When market resources are few, resource allocation must take the competitive environment's impact on service qualities into account. | TripAdvisor | The prioritization of improvement can be  obtained by comparing the final values of NEG(xi) and POS(xi). the prioritization of im-  provements in services should follow the order of Rooms, Location, Sleep,  Value, Food, Cleanliness, Facility, Service. |
| A Survey on Machine Learning Based Requirement Prioritization Techniques | In this study Priogov, CBRank,GDRank which are currently used as prioritization requirements using machine learning were throughly observed these techniques were studied and analysed then further compared along with their strength and limiations with respect to the data given in the paper | To identify and analyse the existing techniques of requirements prioritization that are based on machine learning. | Building software is not merely an art of Software Engineering is a systematic approach of applying effectively the engineering principals to build a software that fulfills its required purpose. the software complying stakeholders needs, the requirements . Software requirement prioritization means, identifying which requirements hold more importance from stakeholder`s It is important to understand stakeholders | Although Machine learning method is a good approach for the tasks proposed by the literature there are certain drawbacks of using the method as it is very time consuming, not always accurate enough and evenethough it is accurate sometimes the large data are not easy to handle.Hence it needs more features to overcome the mentioned problems and decrease the reliability of requirements. |  | Pirogov - Less Accurate CBRank - Less accurate GD Rank -Most Accurate Machine learning-based techniques have been applied to some requirements prioritisation techniques in an effort to simplify the task.Machine learning-based methods for prioritising needs are quite rare. Each method has its own advantages and disadvantages. When very accurate findings are necessary and there are few constraints, GDRank can be employed. When we can compromise on accuracy to some extent, CBRank can be employed. |
| Analysis & Implementation of Sentiment Analysis of User YouTube Comments | The work includes acquisitions of raw data extracted from you tube and its conversion into standard datasets,analysis of contents retrieved from you tube, development of Sentiment Analysis Tool,Prediction of content from you tube and classification of content. | Performance of sentiment analysis on YouTube comments on the most popular topics nowadays by using Classifier technique | Social media usage is rapidly increasing due to its ease of use and simplicity in creating and sharing pictures from everyone, including those who are technically unaware of social media. Non-textual content is also shared by various platforms, such as animations, videos, images, and audios, which allow users to provide feedback through comments. YouTube is widely regarded as the most popular of the various web platforms. The rich content is analysed and exploited by sharing it on YouTube based on the interests of the research community. According to some studies, the retrieval potential of the video is investigated, and comments as well as Meta data must be used. In comparison to inappropriate videos, YouTube is an appropriate way that receives a large number of positive comments. As a result, it can be stated that comments are one of | 1. Challenges in present sentiment dictionaries. 2. Users are using informal language. 3. On the basis of community-created terms, sentiments are estimated. 4. Proper labels must be assigned to the events. 5. To attain satisfactory classification performances  6. Challenges involved in “social media sentiment analysis”. | Youtube | Accuracy performance measure 89.3%. In comparison with the statistical model, the results showed that this model achievedbetteraccuracy. The range of classification accuracy is in between 70% to 89% |
| Intelligent code reviews using deep learning | Development of automatic code  analysis system called DeepCodeReviewer to automatically review code for commonly faced issues and developer can overcome it by proactively correcting the code and d reviewers can  perform code reviews focused towards finding defects. | Evaluation of proposed deep learning model, and results from user study to evaluate overall performance of automatic, flexible, and adaptive code analysis system called DeepCodeReviewer. | The manual review of source code by one or more peers of the code author is a crucial step in the software development lifecycle. the origin  The source code is examined for errors, to make sure it adheres to best practises, and to look for vulnerabilities including race situations, malware, memory leaks, buffer overflows, and format string exploits. Code review is done to identify these potential issues. | This system can provide code reviews automatically for common issues.The DCR algorithm does not works continuously when the pull request is made this is planned to make happen in future. It does not highlights any information related to code semantics or long range context. DCR tends to fail learn new issues, personalize itself to a  team/repository, and learn complex issues. | Internal corporate git repositries (github) | Evaluated the code review model using automatic metrics  and evaluated the overall DCR system using a user study. The  multi-encoder deep learning achieved 136% higher MRR than  logistic regression using TF-IDF. The model was able to  achieve good MRR because of its capability to learn complex  code patterns by analyzing sequence of code and review  tokens- the capability not present in TF-IDF representation |
| Sentiment Analysis on Amazon Product Reviews using the Recurrent Neural Network (RNN) | Vivid deep learing models have been implemented,in this research, the proposed model (RNN) will be  used. Also, other types of neural networks will be applied as  comparison experiments to the proposed model. | To perform the sentiment analysis using Arabic dataset and Recurrent Neural Network model | The rise of social media, including chat rooms, blogs, microblogs, and Twitter, and the growing significance of sentiment analysis are related . For the first time in human history, a sizable number of opinions are available for study in a digital format. Internet users comment on products on Amazon.com. These opinions, which differ from product to product, are used to enhance items and notify businesses when customers have unpleasant things to say about them | Accuracy of the sentiment analysis performed turned out to be around 85% but a major drawback in this lietrature was that the analysis was only based on Arabic dataset and also the accurate results were obtained only when a large data set was used to perform analysis procedure. | Amazon reviews Dataset | Comparing with the deep learning models the current algorithm or model has obained an accuracy of 96% which is 2% efficient from the best deep learning model. It has got recall value as 85% |
| N-Grams Assisted Youtube Spam Comment Detection | Identification of spam comments through implementation of conventional machine  learning  algorithms such as Random Forest, Support Vector Machine, and Naive Bayes along with certain custom heuristics such as Count Vectorizer that proved to be very effective in detection and to subsequently combat spam commentary. | Identification of the Machine Learning algorithms and application of specific heuristics for detection of  spam with accuracy. | One of the most utilised features of Youtube is its commenting system where users can comment on  videos uploaded to other channels. Commenting on the video allows the users to interact with each other and share their feelings, ideas etc. Yet, this has also turned as a chance for malicious users to share divulgatory content also known as spam. Resluting in a lot of spam comments on a video | Modern machine learning methods like Random Forest and Naive Bayes classifier are used to  detect spam comments. However, it can be further enhanced by utilising the spam detection Deep Learning model, which primarily focuses on: embedding words  GRU plus a bi-directional deep learning model (Gated Recurrent Units). When utilising deep learning, the model recognises the features while being trained, and we may design the network using Keras where the aforementioned elements are present. Therefore, it could be quite beneficial and helpful when these kinds of methods are thoroughly investigated. | Youtube comments | The average detection accuracy was as high as 96% Such a result provides insight on the usefulness of  the proposed video spammer feature set. In order to investigate more, additional experiments need to be performed, comparing the results against existing feature set for spammer detection |
| Sentiment Analysis of Consumer Reviews Using Deep Learning | This paper uses vivid methodolgies for the consumer reviews, from python programming environment to deep learning based classification methods as well as  extracting the benchmark dataset. There are other methods involed such as data preprocessing, erase puncutation, convert the text to lowercase,tokenization,vectorization , removal of hyperlinks , padding, POS Tagging, word enmbeding,feature encoding for numerical representation of textual data. | To perform sentiment analysis of customer reviews with the use of various Deep Learning models | The increased usage of social media and e-commerce websites is constantly generating a massive volume of data about image/video, sound, text, etc. The text  among these is the most significant type of unstructured data, requiring special attention from researchers to acquire meaningful information. | The proposed results were better than or comparable to previous approaches. In the future, this work can involve other deep learning architectures, such as transformers.Transformer  model is based on the self-attention mechanism, which allows it to weigh the importance of different parts of the input sequence when making predictions. | IMDB, Yelp and Amazon reviews dataset , Amazon fine food reviews | Model 1 shows that the performance of all measures, Accuracy, Precision, Recall, and F-Measure, is 83%, 79%, 59%, and 61%, respectively. Model 2 displays that the performance  of all measures of Accuracy, Precision, Recall, and F-Measure is 82%, 71%, 62%, and 64%,  respectively Model 3 indicates that the performance of all measures, Accuracy, Precision,  Recall, and F-Measure, is 83%, 73%, 61%, and 64%, respectively |
| Recommendation algorithm combining ratings and comments | The importance of each and every review is completely considered in Local-Global Awareness Attention Model(LGAA) to model the  comment information | To analyse which method is more efficient in recommendation systems using LGAA based on text views,deep  learning algorithms | The matrix factorization-based recommendation model has achieved good performance. However, it is limited by the dot product operation, the inability to fit the nonlinear relationship between features, and the inability to incorporate other relevant information. | This paper creates a rating based feature but before that the rating matrix can be pre-processed to expand the learnable samples to obtain better feature vectors. expand the learnable samples to obtain better feature vectors; the fused final feature vectors can be extended to make them more expressive; the local attention mechanism can be added to the window mechanism, which can model the adjacent reviews to further improve the computational efficiency. | Digital Music (DM) . Toys and Games(TG) , Kindle Store ( KS) of Amazon | LR-LGAA: Final prediction scores are given using logistic regression which is simple and effective. It is a good benchmark algorithm.. FM-LGAA: Final prediction score using factorization machine. DNN-LGAA: The final prediction score is given using fully connected network. If considering the improvement scale there has been 0.6% to 5.7% imporvent using LGAA model on different exisitng deep learning models |
| Online critical review classification in response strategy and service provider rating: Algorithms from heuristic processing, sentiment analysis to deep learning | The classification of online critical reviews with a specific emphasis on response strategies employed by service providers and their impact on service provider ratings. | To make a comprehensive investigation of various algorithms and approaches for effectively categorizing critical reviews and understanding their implications for service providers. | The challenge lies in developing methods that can accurately assess the sentiment of reviews, capture nuances in response strategies, and provide insights into how these factors influence the perception of service providers among consumers. This paper proposes and evaluates algorithmic solutions that enhance the understanding of online critical reviews' role in shaping consumer opinions and service provider reputations. | The effectiveness of heuristic processing and sentiment analysis heavily rely on the quality of preprocessing and the choice of sentiment lexicons. | Online reviews received by hotels in a London hotel chain | This paper proposed twelve algorithms to classify critical online reviews in the service industry using three major approaches: Heuristic Processing, Linguistic Feature analysis and Deep Learning-based Natural Language Processing (NLP) and developed criteria to examine the effectiveness of each algorithm empirically. That is, a good (bad) classification algorithm produces more (less) accurate identification on true critical reviews. Therefore, the percentage of classified critical reviews responded to would have a strong (weak or no) positive relationship with the evaluation of service providers. |
| Approaches for prioritizing feature improvements extracted from app reviews | This paper focuses on developing and evaluating different approaches to effectively identify and prioritize feature enhancement requests and issues raised by users in app reviews. | To enhance the software development process by providing app developers with systematic methods to analyze user feedback, identify key enhancement requests, and prioritize them based on various factors, such as frequency, sentiment, and context. | In the landscape of mobile application development, user reviews play a pivotal role in providing valuable feedback and suggestions for improving app features and functionalities. However, the sheer volume and diversity of user reviews pose a challenge for app developers in systematically extracting, analyzing, and prioritizing feature improvements that arise from user feedback. | The proposed approaches are not suitable for real-time analysis of user reviews, this limits their practicality in dynamic app development environments. | Reviews from the MyTracks app | Identified four attributes (frequency, rating, negative emotions and deontics) that serve as the base constructs for prioritization. Thereafter, used these four constructs, to develop three approaches (individual attribute-based approach, weighted approach and regression-based approach) that may help developers to prioritize features for improvements. |
| Customer preferences extraction for air purifiers based on fine-grained sentiment analysis of online reviews | This paper proposes an approach based on fine-grained sentiment analysis and the Kano model to extract consumer demands for product attributes from online reviews. | To propose an approach based on fine-grained sentiment analysis to extract consumer preferences for product attributes from online reviews | It is possible to improve product design and enhance competitiveness by mining consumer demands for product functions from online reviews. However, the existing aspect-level sentiment analysis methods fail to deal with the over-segmentation of multi-word aspects, which is likely to cause the omission of aspects. | Due to the large amount of professional  knowledge in the field of air purifier, part of the work in this  work is still manually accomplished in the process of attribute  recognition.  Additionally, only a qualitative analysis of  consumer demands is conducted in this study. | Online reviews of air purifiers in Chinese market crawled from T-mall.com | Aspect-level sentiment analysis  is used to identify the sentiment orientation of the consumers  towards the product attributes, and the Kano model-based extraction rules are designed to extract consumer demands from  the sentiment orientation. In this regard,  mining consumers’ demand and preferences from consumer generated online reviews is increasingly highlighted. |

**PROPOSED METHODOLOGY:**

The proposed methodology for the project involves the following steps:

**Identifying the target websites and data to be collected:** The first step is to identify the websites from which data needs to be collected. In this case, the target website is YouTube. The data to be collected includes comments from YouTube videos.

**Developing web scrapers:** Web scrapers are developed using web drivers and the Google API for YouTube. The web scrapers are designed to navigate through each element of a given page, extract comments, and save them as a DataFrame and CSV file.

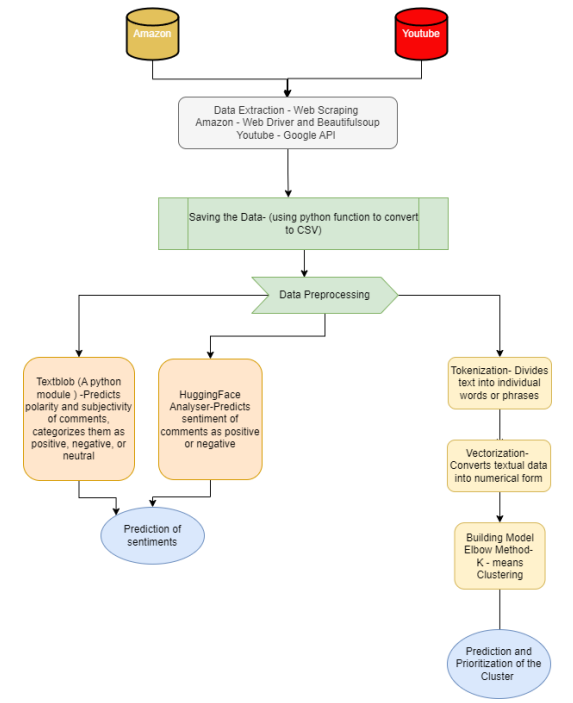
**Analyzing sentiment:** To analyze the sentiment of the comments, two different tools are used - TextBlob analyzer and HuggingFace analyzer. The comments are preprocessed using techniques such as vectorization, tokenization, and punctuation removal.

**Identifying clusters of related words:** The elbow method is used to determine the optimal number of clusters for K-means clustering. The words are then clustered into the identified number of clusters, and the cluster with the most number of questions is prioritized.

**Interpretation and reporting**: The final step is to interpret the results and report on the findings. The results can be used to gain insights into customer opinions and preferences, as well as inform marketing and business decisions.

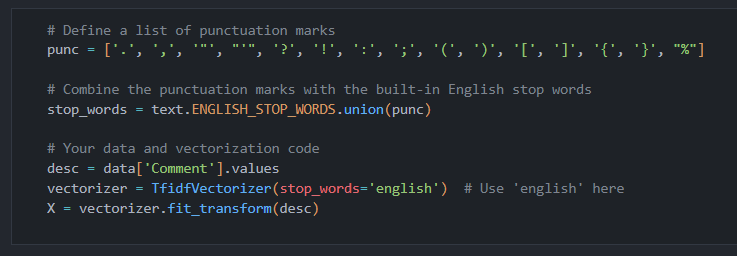
In summary, the methodology involves collecting data from YouTube using web scrapers, analyzing sentiment using natural language processing tools, identifying clusters of related words using K-means clustering, and interpreting and reporting on the findings.

**ARCHITECTURE DIAGRAM:**

****

**IMPLEMENTATION:**

**Text Classification**



Removing Stop Words

A screenshot of a computer program

Description automatically generated

Tokenization

A screen shot of a computer program

Description automatically generated

K Means Clustering

A screenshot of a computer program

Description automatically generated

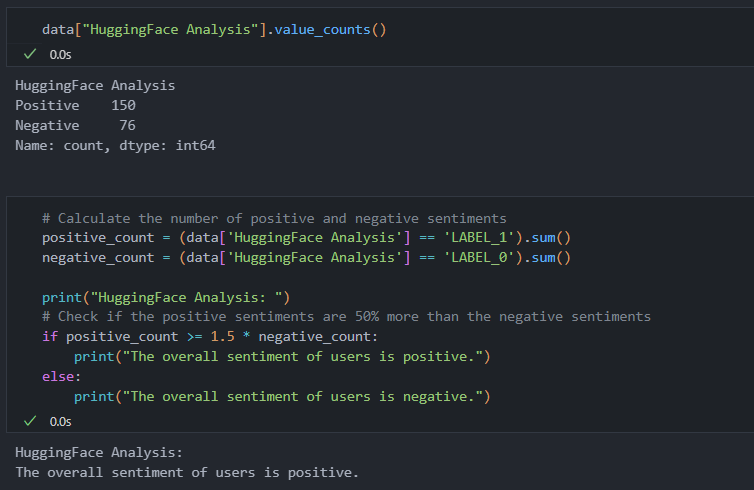
Question Based Grouping

**Sentiment Analysis**

**A screenshot of a computer program

Description automatically generated**

TextBlob Analysis

****HuggingFace Analysis

**RESULTS AND DISCUSSIONS:**

The implementation of a web scraping bot coupled with the K-Means clustering method has demonstrated noteworthy outcomes in enhancing the management of reviews and comments. The web scraping bot has proven its efficacy by swiftly and effectively extracting data from websites, surpassing the limitations of screen scraping, which merely captures on-screen pixels. This advanced method allows for the comprehensive extraction of valuable data stored in a website's underlying HTML code and database, facilitating the replication of entire website content elsewhere.

The incorporation of the K-Means clustering method further refines the management process by categorizing comments and reviews based on the number of questions they contain. This strategic approach streamlines the handling of customer feedback, providing a more organized framework for response prioritization. The unsupervised nature of K-Means clustering, grouping unlabeled datasets into clusters based on a pre-defined number, makes it an optimal solution for efficiently categorizing and prioritizing the diverse spectrum of customer feedback.

In summary, the synergy between web scraping and K-Means clustering has emerged as a highly effective method for managing customer reviews and comments. The streamlined process of categorization and prioritization enables businesses to respond more promptly and efficiently to customer concerns, ultimately contributing to an enhanced level of overall customer satisfaction. This approach not only saves time but also empowers businesses to make data-driven decisions for continuous improvement.

**CONCLUSION:**

In conclusion, the integration of a web scraping bot and the K-Means clustering method has revolutionized the landscape of review and comment management. The web scraping bot's efficiency in extracting comprehensive data, surpassing the limitations of screen scraping, offers a rapid and superior approach to replicating website content. Complementing this, the K-Means clustering method categorizes feedback based on question count, enhancing the systematic handling of customer input. As an unsupervised learning algorithm, K-Means clustering excels in organizing unlabeled datasets, providing an optimal solution for prioritizing and addressing diverse customer concerns. This combined approach not only streamlines the management process but also enables businesses to respond

promptly and effectively to customer feedback, ultimately elevating overall customer satisfaction. The synergy between web scraping and K-Means clustering emerges as a potent strategy for businesses seeking to enhance their engagement and responsiveness in the dynamic digital landscape.

**FUTURE WORK :**

In future iterations, expanding the project's scope involves refining and diversifying the methodology for more comprehensive insights. Firstly, enhancing web scrapers to adapt to dynamic content changes on YouTube will ensure sustained data accuracy. Introducing machine learning models for sentiment analysis can elevate accuracy, considering the nuanced nature of language. Exploring advanced clustering techniques beyond K-means, such as hierarchical or density-based clustering, may uncover subtler patterns in customer feedback.

Additionally, integrating user feedback into model training processes will enhance the adaptability of sentiment analysis tools. Implementing real-time analysis for more immediate insights and leveraging emerging natural language processing advancements will keep the methodology at the forefront of technological capabilities. Exploring the integration of social media platforms beyond YouTube for a holistic view of customer sentiment can provide a more comprehensive understanding.

Furthermore, incorporating user demographics for tailored insights and delving into causality analysis to understand the impact of specific factors on sentiments represent promising directions for future research. Continuous refinement and evolution of the methodology will ensure its sustained relevance in the ever-evolving landscape of online customer feedback analysis.

**REFERENCES :**

[1] Hirave, T., Malgaonkar, S., Alwash, M., Cheriyan, J. and Surve, S., 2019, December. Analysis and Prioritization of App Reviews. In *2019 International Conference on Advances in Computing, Communication and Control (ICAC3)* (pp. 1-8). IEEE.

[2] Zhang, C., Xu, Z., Gou, X. and Chen, S., 2021. An online reviews-driven method for the prioritization of improvements in hotel services. *Tourism Management*, *87*, p.104382.

[3] Qayyum, S. and Qureshi, A., 2018, November. A survey on machine learning based requirement prioritization techniques. In *Proceedings of the 2018 International Conference on Computational Intelligence and Intelligent Systems* (pp. 51-55).

[4] Wadhwani, S., Richhariya, P. and Soni, A., 2022. *Analysis and implementation of sentiment analysis of user YouTube comments* (No. 7703). EasyChair.

[5] Gupta, A. and Sundaresan, N., 2018, August. Intelligent code reviews using deep learning. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD’18) Deep Learning Day*.

[6] Alroobaea, R., 2022. Sentiment analysis on amazon product reviews using the recurrent neural network (rnn). *International Journal of Advanced Computer Science and Applications*, *13*(4).

[7] Aiyar, S. and Shetty, N.P., 2018. N-gram assisted youtube spam comment detection. *Procedia computer science*, *132*, pp.174-182.

[8] Iqbal, A., Amin, R., Iqbal, J., Alroobaea, R., Binmahfoudh, A. and Hussain, M., 2022. Sentiment analysis of consumer reviews using deep learning. *Sustainability*, *14*(17), p.10844.

[9] Fang, X., Wang, J., Seng, D., Li, B., Lai, C. and Chen, X., 2021. Recommendation algorithm combining ratings and comments. *Alexandria Engineering Journal*, *60*(6), pp.5009-5018.