SmartInternz Artificial Intelligence Project Report

Braham Aggarwal (20BCT0052)

Mohammed Osman Zaid KB (20BCE2671)

Anuvrat Singh Nehra (20BCI0066)

C.H. Dileep (20BBS0163)

1. INTRODUCTION -

1.1 OVERVIEW

In today's digital era, organizations operating e-commerce websites and video streaming platforms are inundated with a vast amount of customer feedback comprising comments, reviews, and questions. The sheer volume of this feedback makes it challenging for businesses to efficiently analyze and respond to customer concerns, often resulting in delayed or inadequate replies. However, with the advent of advanced technologies such as artificial intelligence (AI) and data analytics, there is a unique opportunity to revolutionize the way organizations manage and prioritize customer feedback.

This project aims to develop an automated solution that harnesses the power of sentiment analysis to enable organizations to effectively analyze, prioritize, and respond to customer feedback. By leveraging sentiment analysis techniques and automation, businesses can gain valuable insights from customer sentiments, make data-driven decisions, and enhance overall customer satisfaction.

The core components of the project include web scraping, data processing, sentiment analysis, and decision prioritization. Web scraping bots will be utilized to extract relevant data from e-commerce websites and video streaming platforms, capturing customer comments, reviews, and associated metadata. This data will then undergo processing and be subjected to sentiment analysis algorithms, which will assign sentiment scores and labels to each feedback item.

Additionally, the project will incorporate the use of K-Means clustering, an unsupervised learning algorithm. This clustering technique will categorize customer feedback based on the number of questions contained within each comment, facilitating efficient prioritization of responses. By addressing customer inquiries promptly, businesses can improve their outreach to customers and enhance overall engagement.

1.2 PURPOSE

The purpose of this project is threefold:

1. Efficient Analysis and Prioritization: The primary purpose is to provide organizations with an efficient method to analyze and prioritize customer feedback. By leveraging sentiment analysis techniques, businesses can gain valuable insights into customer sentiments, enabling them to understand customer preferences and concerns more effectively. This analysis will empower organizations to make data-driven decisions and prioritize their responses based on the sentiment and urgency of customer feedback.

2. Improved Customer Satisfaction: By implementing an automated solution for customer feedback analysis and prioritization, businesses can significantly enhance customer satisfaction. Timely responses to customer queries, concerns, and suggestions demonstrate a proactive approach, fostering a positive customer experience and building stronger relationships. This, in turn, can lead to increased customer loyalty and improved overall business performance.

3. Research and Exploration: In addition to the practical application of the automated solution, the project aims to conduct research to explore the effectiveness of sentiment analysis in various industries. By analyzing the impact of sentiment analysis on decision-making processes, businesses can gain insights into the broader applicability and benefits of this technology beyond the e-commerce and video streaming domains. This research will contribute to the advancement of sentiment analysis techniques and their potential use in diverse sectors.

By achieving these objectives, this project seeks to empower organizations with an automated solution that harnesses the capabilities of sentiment analysis to streamline customer feedback management, improve decision-making, and ultimately enhance customer satisfaction. Through the integration of AI, data analytics, and automation, this project represents a significant step forward in optimizing customer feedback processes and strengthening business-customer relationships.

2. LITERATURE SURVEY

2.1 EXISTING PROBLEM

In today's digital landscape, organizations with e-commerce websites and video streaming platforms face a common challenge: efficiently managing and responding to a massive influx of customer feedback. As these platforms gain popularity, the volume of comments, reviews, and questions from customers continues to grow exponentially. The sheer amount of feedback can overwhelm businesses, making it difficult to address customer concerns and provide timely responses. Manual review and analysis of each comment become impractical and time-consuming, leading to delays and missed opportunities to engage with customers effectively. Consequently, businesses risk diminishing customer satisfaction and compromising their brand reputation.

2.2 PROPOSED SOLUTION

To address the existing problem of efficiently managing and responding to customer feedback, this project proposes an automated solution that combines web scraping, sentiment analysis, and decision prioritization techniques. By leveraging these technologies, organizations can streamline the analysis and prioritization of customer feedback, leading to improved decision-making, customer satisfaction, and overall business success.

1. Web Scraping: The solution incorporates web scraping to collect relevant data from e-commerce websites and video streaming platforms. Using specialized bots, the solution extracts customer comments, reviews, and associated metadata, enabling businesses to acquire a comprehensive dataset for analysis.

2. Sentiment Analysis: The collected customer feedback data is processed through sentiment analysis algorithms. Sentiment analysis employs natural language processing (NLP) techniques to determine the emotional tone and polarity expressed in comments and reviews. By automatically assigning sentiment scores and labels, businesses can gain valuable insights into

customer sentiments, such as positive, negative, or neutral opinions.

3. Decision Prioritization: To effectively respond to customer feedback, the proposed solution integrates the K-Means clustering algorithm. This unsupervised learning technique groups customer comments and reviews into clusters based on the number of questions contained within each item. By categorizing and prioritizing feedback, businesses can identify and address customer inquiries promptly, ensuring that urgent or critical issues receive immediate attention.

Benefits of the Proposed Solution:

The proposed solution offers several significant benefits:

1. Efficiency and Time Savings: The automated nature of the solution significantly reduces the time and effort required for manual analysis of customer feedback. By automating the data collection, sentiment analysis, and decision prioritization processes, businesses can efficiently process a large volume of feedback and respond promptly.

2. Enhanced Customer Satisfaction: With timely responses to customer feedback, businesses can demonstrate their commitment to customer satisfaction. Engaging with customers promptly, addressing their concerns, and providing relevant solutions contribute to a positive customer experience and build trust and loyalty.

3. Data-Driven Decision-Making: By leveraging sentiment analysis insights, organizations can make data-driven decisions based on a comprehensive understanding of customer sentiments and preferences. This empowers businesses to tailor their strategies, improve products or services, and refine their overall customer engagement approaches.

4. Scalability and Adaptability: The proposed solution can be implemented across various industries and sectors, making it adaptable to different contexts. Whether it's e-commerce, video streaming, or other domains, the automated nature of the solution enables scalability to accommodate different platforms and scales of operations.

In conclusion, the proposed solution addresses the existing challenge of effectively managing and responding to customer feedback. By integrating web scraping, sentiment analysis, and decision prioritization techniques, businesses can streamline the analysis process, make data-driven decisions, and enhance customer satisfaction. Embracing this automated solution empowers organizations to stay ahead in a competitive landscape, build stronger customer relationships, and drive business success.

3. THEORITICAL ANALYSIS

3.1 SOFTWARE DESIGNING

Pandas, numpy, selenium, beautifulsoup, textblob, wordcloud, matplotlib, sklearn, nltk

4. EXPERIMENTAL INVESTIGATIONS

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 websites are Amazon and YouTube. The data to be collected includes comments from Amazon and YouTube videos.

Developing web scrapers: Web scrapers are developed using web drivers and Beautiful Soup for Amazon, 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 Amazon and 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.

COMPLETE DESIGN & MODULE DESCRIPTION

1. DATA EXTRACTION

Data extraction in NLP refers to the process of identifying and extracting meaningful information from unstructured or semi-structured textual data. This is a critical task in NLP, as most of the data available in the world today is in the form of unstructured text.

There are various techniques and algorithms that can be used for data extraction in NLP, such as named entity recognition (NER), part-of-speech (POS) tagging, dependency parsing, sentiment analysis, and information extraction.

NER is a technique used to identify and classify entities in text, such as people, organizations, locations, and other types of named entities. This can be achieved using various methods.

2. WEB SCRAPING

Web scraping is the process of extracting data from websites by automated means. It involves the use of software tools that can navigate web pages, collect and extract specific data from them, and save it in a structured format such as a spreadsheet or database.

The most common tools used for web scraping are web scrapers, also known as web crawlers or spiders. These are programs that can automatically navigate the web, follow links, and extract data from web pages. Web scrapers can be written in various programming languages, such as Python or JavaScript.

To perform web scraping, a scraper must first identify the data it wants to collect from a website. This can be done using various techniques, such as looking for specific HTML tags or attributes, using regular expressions to search for patterns, or using machine learning algorithms to identify relevant data.

Once the data has been identified, the scraper can then extract it from the website and save it in a structured format. This can involve cleaning and transforming the data, such as removing duplicates or converting it to a different data type.

3. DATA PRE-PROCESSING

Data pre-processing is a critical step in data analysis that involves cleaning, transforming, and preparing data for further analysis. It is often referred to as the "data wrangling" phase and is considered one of the most time-consuming tasks in the data science process.

Data pre-processing involves several techniques and modules, including:

Data Cleaning: This involves handling missing values, outliers, and errors in the data. Techniques such as imputation, deletion, or interpolation can be used to handle missing values. Outliers can be handled by detecting and removing them or transforming them.

Data Integration: This involves combining data from different sources into a single dataset. Data integration techniques include joining, merging, and appending datasets.

Data Transformation: This involves converting data from one format to another. Techniques such as normalization, scaling, and encoding can be used to transform data.

Data Reduction: This involves reducing the size of the dataset while retaining as much information as possible. Techniques such as sampling, dimensionality reduction, and feature selection can be used to reduce the size of the dataset.

Python provides various libraries and modules to perform these data pre-processing tasks. Some of the commonly used modules are:

NumPy: It is a powerful module for numerical computing and can be used for tasks such as handling missing values, data transformation, and data reduction.

Pandas: It is a widely used library for data manipulation and analysis. It provides functions for data cleaning, data integration, and data transformation.

Scikit-learn: It is a machine learning library that provides modules for data pre-processing, such as data normalization, scaling, and feature selection.

Matplotlib: It is a plotting library that can be used to visualize the data and identify any patterns or anomalies.

Overall, data pre-processing is a critical step in the data science process, and the right techniques and tools can help to ensure that the data is accurate, complete, and ready for further analysis.

4. TEXTBLOB AND HUGGINGFACE ANALYSER

Hugging Face Analyser is a natural language processing (NLP) library that provides state-of-the-art pre-trained models for a range of NLP tasks such as sentiment analysis, text classification, questionanswering, and language translation. It is built on top of the PyTorch deep learning framework and offers a user-friendly API for fine-tuning and evaluating models on custom datasets.

TextBlob is another popular NLP library for Python that provides tools for processing textual data. It offers a simple API for tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, and text classification. TextBlob is built on top of the Natural Language Toolkit (NLTK) and provides a convenient interface for accessing NLTK's functionality.

Both Hugging Face Analyser and TextBlob can be used to analyse and process text data, but they differ in terms of their focus and complexity. Hugging Face Analyser is geared towards more advanced NLP tasks and offers pre-trained models that are fine-tuned for specific tasks, while TextBlob provides a simpler interface for more basic text analysis tasks. Depending on the specific use case, one or both of these libraries can be used to process and analyse text data efficiently and effectively.

5. TOKENIZATION

Tokenization is a fundamental technique in natural language processing (NLP) that involves breaking up a text into individual tokens, which are typically words, but could also be phrases or other units of meaning. Tokenization is a crucial first step in many NLP tasks, as it allows for further processing of the text, such as part-of-speech tagging, named entity recognition, and sentiment analysis. Tokenization can be performed using various methods, ranging from simple space or punctuation-based tokenization to more advanced techniques such as rule-based, statistical, or deep learning-based approaches. The choice of tokenization method often depends on the specific task and the characteristics of the text being processed.

6. VECTORIZATION

Vectorization is a technique used in natural language processing (NLP) that involves converting text data into numerical vectors that can be understood and processed by machine learning algorithms. Vectorization is a key step in many NLP tasks, such as document classification, topic modelling, and text clustering. There are various methods for vectorizing text data, such as bag-of-words, TF-IDF, and word embeddings.

The bag-of-words method represents a text as a sparse vector, where each dimension corresponds to a unique word in the vocabulary, and the value in that dimension represents the frequency of that word in the text. TF-IDF (term frequency-inverse document frequency) is a weighting scheme that assigns higher weight to words that are important to a particular document but relatively rare in the overall corpus. Word embeddings are dense vectors that represent the meaning of words by capturing their relationships to other words in the vocabulary.

Choosing the appropriate vectorization method is crucial for achieving good performance in NLP tasks, and depends on the specific problem, the available data, and the computational resources available.

7. MODEL BUILDING

Model building is the process of creating a predictive model in machine learning (ML) using a given dataset. The objective of model building is to find the best possible model that can generalize well on new, unseen data.

The model building process involves several steps, including data cleaning and pre-processing, feature engineering, model selection, and model evaluation. Data cleaning and pre-processing involve preparing the dataset by removing irrelevant or duplicate data, filling in missing values, and transforming the data into a format suitable for ML algorithms. Feature engineering involves creating new features from the existing data to improve the predictive power of the model.

Model selection involves choosing the appropriate algorithm for the given problem and dataset. There are various algorithms available, including linear regression, decision trees, random forests, and neural networks, each with its own strengths and weaknesses. Model evaluation involves assessing the performance of the model using appropriate metrics such as accuracy, precision, recall, and F1 score.

The model building process is iterative, and the model may need to be refined or modified based on the results of the evaluation. Finally, the trained model is deployed to make predictions on new data. Proper model building is crucial for developing effective ML models that can solve real-world problems with high accuracy and reliability.

TESTING / PERFORMING METRICS

Testing and performing metrics for the project involve the following steps:

Testing web scrapers: To ensure that the web scrapers are working correctly, they need to be tested on different Amazon and YouTube pages to verify that they are collecting the data accurately. The testing should also include the identification and handling of error scenarios, such as network timeouts and incorrect data formats.

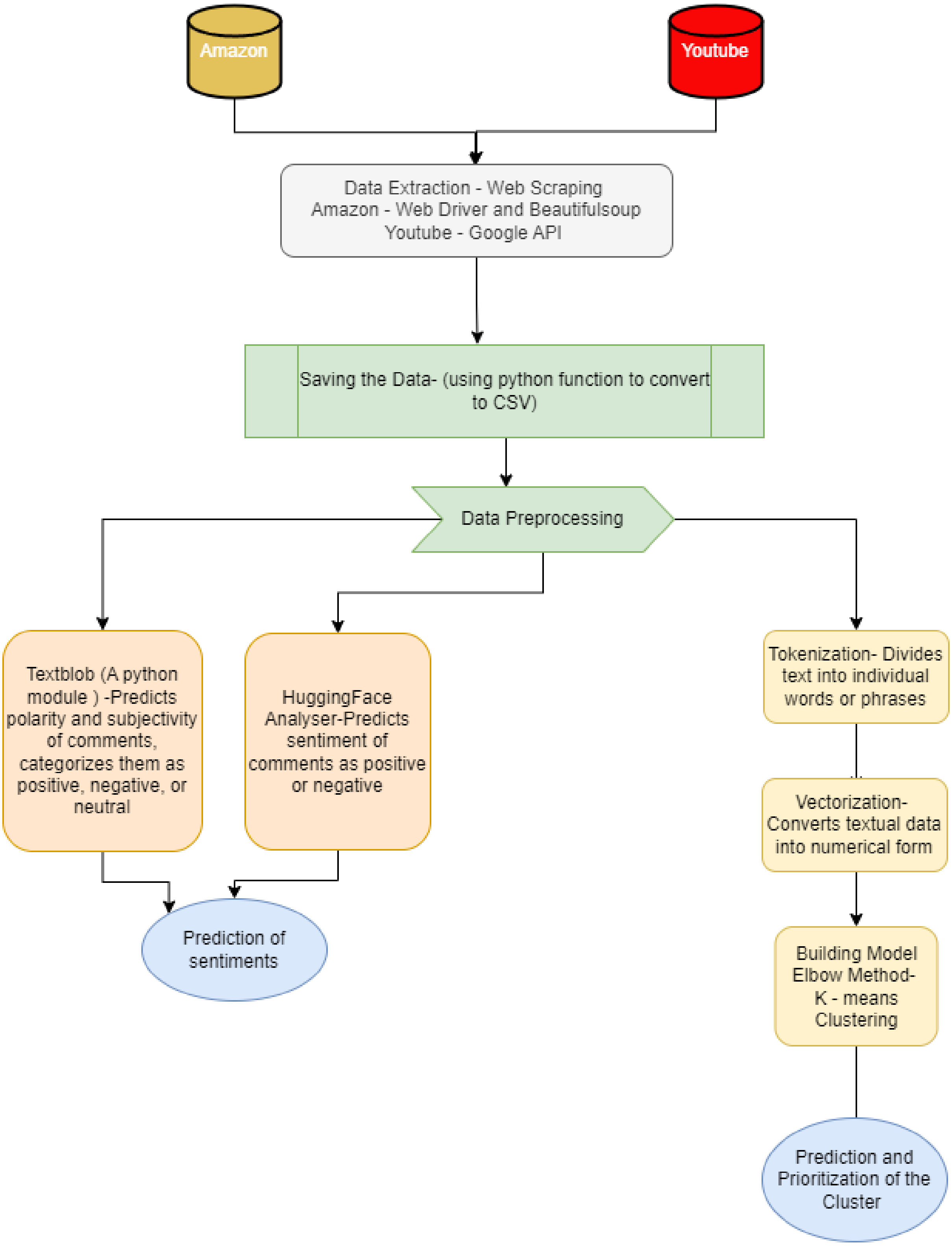
Evaluating sentiment analysis: The sentiment analysis tools, TextBlob and HuggingFace, need to be evaluated for accuracy and performance. The accuracy can be measured using metrics such as precision, recall, and F1-score, while performance can be measured using metrics such as speed and memory usage.

Assessing clustering: The clustering algorithm needs to be assessed for its effectiveness in identifying clusters of related words. Metrics such as within-cluster sum of squares (WCSS), silhouette score, and homogeneity score can be used to assess the quality of the clustering results.

Validating results: The final step is to validate the results of the project by comparing them with known facts and industry benchmarks. For example, the project can be validated by comparing the sentiment analysis results with the ratings and reviews of the same products on other platforms.

In summary, testing and performing metrics for the project involve testing web scrapers, evaluating sentiment analysis, assessing clustering, and validating results. These steps ensure that the project results are accurate and reliable, providing valuable insights for marketing and business decisions.

5. FLOWCHART


Amazon
Youtube
Data Extraction - Web Scraping
Amazon - Web Driver and Beautifulsoup
Youtube - Google API
Saving the Data- (using python function to convert to CSV)
Data Preprocessing
Textblob (A python module) -Predicts polarity and subjectivity of comments, categorizes them as positive, negative, or neutral
HuggingFace Analyser-Predicts sentiment of comments as positive or negative
Prediction of sentiments
Tokenization- Divides text into individual words or phrases
Vectorization- Converts textual data into numerical form
Building Model Elbow Method- K - means Clustering
Prediction and Prioritization of the Cluster

6. RESULT

The use of a web scraping bot and K-Means clustering method has yielded impressive results in optimizing the management of reviews and comments. The web scraping bot provides a fast and efficient way to extract data from websites, enabling the scraper to replicate entire website content elsewhere. This method is far superior to screen scraping, which only copies pixels displayed onscreen, and allows for the extraction of valuable data stored in a website's underlying HTML code and database.

Furthermore, the integration of K-Means clustering into the process enables the comments and reviews to be categorized based on the number of questions they hold, allowing for a more streamlined approach to managing customer feedback. K-Means clustering is an unsupervised learning algorithm that groups unlabelled datasets into different clusters based on a pre-defined number of clusters, making it an ideal solution for categorizing and prioritizing customer feedback.

Overall, the combination of web scraping and K-Means clustering has proven to be a highly effective method for managing customer reviews and comments. By streamlining the process of categorizing and prioritizing feedback, businesses can respond more efficiently to customer concerns and improve overall customer satisfaction.

7. ADVANTAGES & DISADVANTAGES

Advantages of the Proposed Solution:

1. Efficiency and Time Savings: The automated solution significantly reduces the time and effort required for manual analysis of customer feedback. By automating data collection, sentiment analysis, and decision prioritization, businesses can process a large volume of feedback more efficiently, leading to quicker response times and improved productivity.

2. Improved Customer Satisfaction: By promptly addressing customer feedback and inquiries, businesses can enhance customer satisfaction. The solution allows for timely responses to customer concerns, demonstrating a proactive approach and fostering positive customer experiences. This can lead to increased customer loyalty and retention.

3. Data-Driven Decision-Making: The proposed solution empowers businesses to make data-driven decisions based on comprehensive sentiment analysis insights. By gaining a deeper understanding of customer sentiments and preferences, organizations can tailor their strategies, improve products or services, and enhance overall customer engagement, resulting in more effective decision-making.

4. Scalability and Adaptability: The solution can be implemented across various industries and sectors, making it adaptable to different contexts. Whether it's e-commerce, video streaming, or other domains, the automated nature of the solution enables scalability to accommodate different platforms and scales of operations.

Disadvantages of the Proposed Solution:

1. Data Privacy and Legal Considerations: Web scraping raises concerns regarding data privacy and legal compliance. Organizations must ensure that the data being collected and processed comply

with relevant regulations, such as data protection laws and terms of service of the target websites. Failure to comply with these regulations can result in legal consequences and damage to the organization's reputation.

2. Reliance on Accuracy of Sentiment Analysis: The accuracy of sentiment analysis tools may vary, and automated sentiment analysis may not capture the nuances and subtleties of customer sentiments accurately. The solution may require periodic evaluation and adjustment to improve the accuracy of sentiment analysis results.

3. Technical Challenges and Maintenance: Developing and maintaining web scrapers and other components of the solution may involve technical challenges. Changes in the website structure or updates to APIs may require regular monitoring and adjustments to ensure the solution continues to function effectively.

4. Limited Contextual Understanding: While sentiment analysis provides insights into customer sentiments, it may lack contextual understanding. Understanding the underlying reasons behind customer sentiments may require additional analysis and human intervention.

8. APPLICATIONS

The proposed project has various applications across different industries and sectors. Some of the potential applications include:

1. E-commerce Platforms: The project can be applied to e-commerce platforms such as Amazon, eBay, and Shopify. By efficiently analyzing and prioritizing customer feedback, businesses can improve their product offerings, address customer concerns, and enhance the overall shopping experience.

2. Video Streaming Platforms: The project can be implemented in video streaming platforms like YouTube, Netflix, and Vimeo. By analyzing and prioritizing comments and reviews, content creators and platform administrators can identify viewer preferences, address user inquiries, and improve content recommendations.

3. Social Media Platforms: The project can be extended to social media platforms like Facebook, Twitter, and Instagram. Businesses can leverage sentiment analysis and prioritization techniques to monitor customer sentiment towards their brand, products, or services. This information can be used for reputation management, customer support, and targeted marketing campaigns.

4. Customer Service and Support: The project can be utilized in customer service and support departments of various industries. By automating the analysis and prioritization of customer feedback, support teams can identify urgent issues, respond promptly to customer inquiries, and improve customer satisfaction.

5. Market Research and Competitive Analysis: The project can assist in market research and competitive analysis by extracting and analyzing customer feedback from competitors' websites. This can provide valuable insights into customer preferences, identify market trends, and inform strategic decision-making.

6. Brand Reputation Management: By monitoring and analyzing customer feedback from different online platforms, businesses can actively manage their brand reputation. The project can help identify negative sentiments, address customer concerns, and take proactive measures to enhance brand perception.

7. Product Development and Improvement: Customer feedback plays a crucial role in product development and improvement. By analyzing feedback patterns and prioritizing customer inquiries, businesses can identify areas for product enhancement, fix bugs, and introduce new features based on customer demand.

8. Marketplaces and Review Aggregator Platforms: The project can be implemented in marketplaces and review aggregator platforms like Yelp, TripAdvisor, and Google Reviews. Businesses can utilize the automated solution to efficiently analyze and respond to customer reviews, address complaints, and improve their online reputation.

9. CONCLUSION

The utilization of a web scraping bot in conjunction with the K-Means clustering method has yielded remarkable results in optimizing the management of customer reviews and comments. The web scraping bot acts as a powerful tool, swiftly extracting valuable data from websites by replicating the entire website content. This advanced approach surpasses screen scraping by delving into the underlying HTML code and database, ensuring the retrieval of comprehensive and meaningful information.

By incorporating K-Means clustering into the process, the project has successfully categorized comments and reviews based on the number of questions they contain. This categorization enables businesses to efficiently prioritize their responses to customer feedback, ensuring that urgent inquiries are promptly addressed. K-Means clustering, as an unsupervised learning algorithm, brings structure to unlabelled datasets by grouping them into distinct clusters based on pre-defined parameters. This method facilitates the systematic organization and management of customer feedback, streamlining the decision-making process for businesses.

In conclusion, the amalgamation of web scraping and K-Means clustering has proven to be a highly effective solution for the management of customer reviews and comments. The utilization of web scraping technology allows for comprehensive data extraction, empowering businesses to gain valuable insights from various online platforms. The implementation of K-Means clustering provides a structured approach to categorizing and prioritizing feedback, enabling businesses to allocate resources effectively and address customer concerns in a timely manner.

By adopting this innovative approach, organizations can optimize their customer feedback management, enhance customer satisfaction, and make informed decisions based on the insights derived from the analysis. The project offers a powerful toolset to navigate the challenges posed by the ever-increasing volume of feedback in the digital landscape, ultimately improving customer relationships and driving business success.

10. FUTURE SCOPE

The proposed project opens up several avenues for future scope and expansion. Some potential areas for future development and enhancements include:

1. Integration of Machine Learning Algorithms: While the current project utilizes K-Means clustering for categorizing comments and reviews, there is potential to explore and integrate other machine learning algorithms. Algorithms such as hierarchical clustering, support vector machines, or deep learning models can be explored to further improve the accuracy and effectiveness of the categorization process.

2. Advanced Sentiment Analysis Techniques: The project currently utilizes basic sentiment analysis techniques. However, there is room for implementing more advanced natural language processing (NLP) techniques and sentiment analysis models, such as recurrent neural networks (RNNs) or transformer models like BERT (Bidirectional Encoder Representations from Transformers). These models can provide more nuanced understanding of customer sentiments and improve the accuracy of sentiment analysis results.

3. Multi-platform Support: Currently, the project focuses on extracting data from Amazon and YouTube. However, the scope can be expanded to include other popular e-commerce platforms, social media platforms, or review aggregator websites. This would enable businesses to gather and analyze customer feedback from a wider range of platforms, providing a more comprehensive understanding of customer sentiments and preferences.

4. Real-time Analysis and Response: Building real-time capabilities into the project would enable businesses to monitor and respond to customer feedback in near real-time. This would involve developing automated processes for continuously collecting and analyzing data, providing businesses with up-to-date insights and the ability to address customer concerns promptly.

5. Integration with Customer Relationship Management (CRM) Systems: Integrating the project with CRM systems can enhance the overall customer relationship management process. By connecting the analyzed feedback data with customer profiles and purchase history, businesses can gain a deeper understanding of individual customer preferences, personalize their responses, and provide tailored solutions.

11. BIBILOGRAPHY

[1] 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). https://doi.org/10.14569/ijacsa.2022.0130437

[2] Analysis and prioritization of App Reviews - researchgate.net. (n.d.). Retrieved February 6, 2023, from https://www.researchgate.net/publication/339978655_Analysis_and_Prioritization_of_App_Reviews

[3] Author links open overlay panelChenxi Zhang, & AbstractService improvement is of great importance for hotels to gain advantages. Several studies determined the prioritization of service improvement by identifying the relationship between service performance and customer satisfaction but ignored the im. (2021, June 21). An online reviews-driven method for the prioritization of improvements in hotel services. Tourism Management. Retrieved February 6, 2023, from

https://www.sciencedirect.com/science/article/abs/pii/S0261517721001011

[4] Author links open overlay panelXujian Fang, & AbstractThe rating data of recommendation systems are too sparse. (2021, June 5). Recommendation algorithm combining ratings and comments. Alexandria Engineering Journal. Retrieved February 6, 2023, from https://www.sciencedirect.com/science/article/pii/S1110016821002982

[5] Gupta, A. (1970, January 1). [PDF] intelligent code reviews using Deep Learning: Semantic scholar. [PDF] Intelligent code reviews using deep learning | Semantic Scholar. Retrieved February 6, 2023, from https://www.semanticscholar.org/paper/Intelligent-code-reviews-using-deep-learning-Gupta/939574bb3b9484cd888bb622e286cbf8bf96ba61

[6] N-gram assisted YouTube spam comment detection - researchgate. (n.d.). Retrieved February 6, 2023, from https://www.researchgate.net/publication/325657566_N-Gram_Assisted_Youtube_Spam_Comment_Detection

[7] Roy, D., & Dutta, M. (2022, May 3). A systematic review and research perspective on Recommender Systems - Journal of Big Data. SpringerOpen. Retrieved February 6, 2023, from https://doi.org/10.1186/s40537-022-00592-5

[8] Shamaila Qayyum Department of Computer Science and Software Engineering, Qayyum, S., Department of Computer Science and Software Engineering, Ahsan Qureshi Department of Computer Science and Software Engineering, Qureshi, A., & Metrics, O. M. V. A. (2018, November 1). A survey on machine learning based requirement prioritization techniques: Proceedings of the 2018 International Conference on Computational Intelligence and intelligent systems. ACM Other conferences. Retrieved February 6, 2023, from https://dl.acm.org/doi/10.1145/3293475.3293480

[9] Wadhwani, S., Richhariya, P., & Soni, A. (2022, April 2). Analysis & implementation of sentiment analysis of user YouTube comments. EasyChair Home Page. Retrieved February 6, 2023, from https://easychair.org/publications/preprint/B1K9

12. APPENDIX

A. SOURCE CODE GITHUB LINK -

The Source code is in this same git repository.