

2025年臺灣國際科學展覽會

優勝作品專輯

作品編號	190033
參展科別	電腦科學與資訊工程
作品名稱	AI-Based Customer Sentiments Dashboard
得獎獎項	四等獎

就讀學校	Bryanston High School
指導教師	Mfanasibili Innocent Sibitane
作者姓名	Aradhya Kaushik

關鍵詞	<u>Sentiment Analysis、Natural Language Processing (NLP)、Power BI、Lexicon、Stop words</u>
-----	---

Topic: AI-Based Customer Sentiments Dashboard

Aradhya Kaushik

Project Report

Defining Terms

Sentiment Analysis - Sentiment analysis is a method used to determine the emotional tone behind a piece of text, categorizing it as positive, negative, or neutral to understand opinions and attitudes.

Natural Language Processing (NLP) - NLP is a field of artificial intelligence focused on the interaction between computers and human language, enabling machines to understand, interpret, and generate human language effectively.

Power BI - Power BI is a business analytics tool by Microsoft that enables users to visualize and share insights from their data through interactive reports and dashboards, facilitating informed decision-making.

Lexicon - A lexicon is a collection of words, and their meanings used in NLP(Natural Language Processing) to analyze text, where each word or phrase is associated with specific attributes like sentiment, helping in various linguistic tasks.

Stop words - Stop words are common words (such as "and" "the," and "is") that are filtered out during text processing in NLP(Natural Language Processing) because they carry minimal meaning and do not contribute to the overall analysis.

Introduction

In this fast-paced digital economy, customers' judgment is based on their experience with a company's products and services. Customer reviews become a vital source of information for companies because this information can be used to enhance their products, understand customer wants and needs, improve brand reputation, and provide a competitor's advantage.

A company can understand customer needs and wants by going through reviews. Customers are encouraged to leave not only their opinion but also their ideas for the development of the product or service. By understanding these reviews, a company can actively respond and engage with a reviewer or problem.

Failure of companies who don't answer customer queries may negatively impact customer loyalty. Customers will feel neglected by these companies and will choose competing companies to handle their needs. Additionally, customers may speak negatively about a company that does not respond to reviews.

The AI-based customer sentiment dashboard can help gain a company's competitive advantage by identifying weaknesses in themselves and others. Companies will be enabled to understand where they succeed and where improvement is needed compared to their competitors, leveraging businesses to address strengths and weaknesses before competitors do.

Through AI-based customer sentiment dashboards, a company can analyze its competitor's reviews and use that information as leverage to make improvements to its products and services.

Customers are increasingly leaving reviews on popular apps like Google Play, Stamped.io, Yapto, and Judge.me, Loox, Qualaroo, and Yelp. The reviews are rich in customer sentiments offering valuable insights into user satisfaction and pointing out the areas for improvement that are crucial to every company no matter how big or small.

Despite their value, manually processing these reviews is a challenging task due to the large volume of unstructured data. Manual processing is also vulnerable to bias and human error, leading to inaccurate information. Traditional methods such as surveys have been proven to be ineffective if the main focus is targeted feedback and have low responses compared to reviews.

The advances in artificial intelligence like Natural Language Processing (NLP) help us interpret and analyze human language and generate outputs like predicting what type of sentiments are in reviews.

This project proposes developing an AI-based sentiment analysis model to evaluate customer feedback on two widely used taxi applications. Natural Language Processing libraries, such as the Valence Aware Dictionary and Sentiment Reasoner (. The model aims to categorize customer reviews into positive, negative, and neutral sentiments.

Literature Review

In today's fast-paced online commercial environment, customer feedback is crucial for improving products and services(Ula Kamburov-Niepewna, 2024). For example, a review says "Your car arrived 5 hours later than when I booked it "a business is communicated about the problem and now can be informed decisions on it. Businesses use this feedback to identify issues with their products or services. Continuous feedback also helps to enhance quality, allowing opportunities for improvement and innovation for the business company. (Flori Needle, 2023) Service providers can rely on these reviews to analyze and improve their services and products. However, with thousands of reviews pouring in on various platforms, it becomes a challenge for service providers to manually analyze such large volumes of information and draw value from them.(Ula Kamburov-Niepewna, 2024)

Service providers have several ways to seek customer feedback. They can seek feedback through several digital channels based on customer reviews of their experiences with the products and services provided. For example, platforms for downloading and reviewing apps have become popular for customers to share their experiences by writing reviews and giving star ratings about their applications, products, and services. Companies also use a common method of online surveys that are sent via websites or email. Companies create customer feedback forms on their websites or apps to collect customer reviews. They may also use Net Promoter Score (NPS) surveys, where they ask customers to rate how likely they are to recommend the product or service to someone else on a scale of 1 to 10. Companies have employed marketing firms and have PR teams that monitor public reviews of their companies on other public platforms like social media platforms. They may also host in-person surveys where they ask questions to their consumers to discuss their product or service.

Taxi apps are an ideal choice to explore in discussions about machine learning models, such as sentiment analysis, due to their widespread usage, real-time data, and reliance on customer feedback for improvement. (Sarker, 2021) Taxi apps depend on large data sets generated from customer reviews, driver ratings, and trip details, making them perfect for testing how models like the Valence Aware Dictionary and Sentiment Reasoner (VADER) handle real-world feedback. The diversity of user emotions and the urgency of responses create ideal conditions for evaluating the performance of sentiment analysis.(Malikal Adilah ,et al., 2020)

Taxi apps are mobile applications designed to connect passengers with drivers for convenient and efficient transportation services. This is a digitization of the process of booking rides, tracking vehicles, and handling payments. Many Taxi Apps include rating systems where passengers and drivers can rate each other, ensuring accountability and quality service.

These apps operate in highly competitive environments, and customer satisfaction is crucial. Sentiment analysis helps understand customer experiences, enabling companies to adjust their services accordingly.(Sarker, 2021) Analyzing user feedback, driver behavior, and trip satisfaction

offers valuable insights for improving overall service, customer retention, and competitive advantage.

The Valence Aware Dictionary and Sentiment Reasoner (VADER) is a rule-based sentiment analysis tool developed in 2014 by a team of researchers led by C.J. Hutto and Eric Gilbert. It was created to analyze the sentiment of social media texts and other forms of informal communication.

VADER is a tool designed for analyzing large volumes of data, such as reviews, and assessing sentiments. Being a pre-trained model, VADER does not require extensive training data, making it accessible for companies that may not have experience with artificial intelligence.

VADER effectively combines a human language dictionary with a straightforward scoring algorithm. VADER features a Human-Centric Lexicon, which includes a dictionary of words rated by human evaluators based on their sentiment intensity. The tool accounts for punctuation, capitalization, and degree modifiers that influence sentiment scores. VADER is specifically designed for short, informal texts like tweets, comments, and reviews that often use emojis and slang. (Hutto and Gilbert, 2014)

VADER has significantly impacted the customer service and sentiment analysis sectors by offering accessible, interpretable, and reliable sentiment scoring through its various features. This ability influences the interaction between companies and their customers, making VADER a crucial component in the customer service industry.

Artificial Intelligence (AI) is transforming the customer service sector in several impactful ways. One significant application is sentiment analysis, where AI analyzes customer feedback to provide valuable insights into overall satisfaction and identify trends in complaints or preferences. Additionally, AI-powered chatbots play a crucial role in handling customer inquiries and resolving issues in real-time, ensuring that support is available around the clock. Another key application is recommendation systems, which leverage historical data to personalize services and suggest improvements tailored to individual customer preferences. Furthermore, predictive analysis enables AI to identify potential churn risks, allowing businesses to implement preemptive strategies aimed at retaining customers. Together, these applications illustrate how AI is enhancing customer service, making it more efficient and responsive to user needs.

Taxi apps must prioritize safety and security to ensure reliable services for users. In countries like South Africa and others with high crime levels, safety concerns have become a significant issue. Incidents such as robberies during rides or crimes linked to taxi apps are frequently reported, raising concerns about passenger security. These incidents not only endanger users but also harm the reputation of these apps, as negative reviews citing safety concerns discourage potential customers, leading to a loss of business.

This project addresses these safety challenges through two data-driven approaches. Firstly, by analyzing user reviews, the project identifies recurring mentions of crime-related terms. Words

like "robbed" or "unsafe" frequently appearing in reviews are highlighted in a word cloud analysis, allowing businesses to detect and act on safety concerns in specific areas. Secondly, the project includes a visualization dashboard with a QA section, enabling businesses to search for and examine negative reviews linked to safety issues. This feature provides actionable insights into the nature and location of safety concerns, helping companies take targeted measures to improve security.

In regions with high crime rates, such as South Africa, enhancing the safety of taxi apps is critical for restoring public trust and ensuring the sustainability of the industry. By leveraging user feedback and implementing solutions like those offered in this project, taxi companies can create safer and more reliable services for their customers (World Bank, 2023).

The model will be built to perform real-time analysis. This means that when the dashboard is refreshed the code will automatically run itself giving an analysis of the most recently posted 500 reviews. This is helpful as it helps the companies make decisions based on the most recent problems.

Survey platforms like Microsoft Teams Form, SurveyMonkey, and Google Forms use a similar approach to the visualization of data as Power BI.

Microsoft Teams Forms fully integrates with Microsoft Teams, allowing users to create, distribute, and analyze forms seamlessly within the Teams interface while collaborating effectively with other Microsoft 365 tools like Excel and Power BI for comprehensive data analysis. In contrast, Google Forms integrates effortlessly with Google Workspace, making it ideal for users entrenched in the Google ecosystem.

Microsoft Teams Forms is often preferred in professional and academic environments for quick polls, quizzes, and feedback collection during meetings and collaborative projects. Google Forms enjoys widespread use in casual and professional arenas, including education, small businesses, and personal projects. Microsoft Teams Forms provides basic capabilities with simple analytics and limited customization. Google Forms, meanwhile, offers robust design options, including themes, branching logic, and plugins for further personalization.

Microsoft Teams Forms is a powerful tool that can be used to collect and analyze customer review sentiments. By creating a form within Teams, companies can gather feedback from customers in a structured format. Forms can be made to ask various types of questions, such as rating scales, multiple-choice, and open-ended responses, which allow customers to express their sentiments about products or services. Once the data is collected through Teams Forms, it can be exported to Excel or directly integrated with Power BI for advanced visualization and analysis. (JunyuB , 2023)

Power BI can transform the raw data into interactive dashboards that display sentiment trends, key metrics, and visual representations such as pie charts, bar graphs, and word clouds. These visualizations provide an intuitive and comprehensive view of customer feedback, helping service providers quickly identify patterns and insights. While a team form is user-friendly and offers real-time insights, it is not the best for handling large volumes of data as stated on the website a team's form can only handle 4000 answers at a time and making the form can be time-consuming as you have to do a lot of steps such as exporting data to Excel and integrating with Power BI. Microsoft Teams Forms doesn't have built-in functionality to extract review data from external sources like Google Play Store. There is one solution to the problem and that is to do **integration with** Microsoft Power Automate can be used to create workflows that pull data from external APIs and insert it into a Teams Form or directly into an Excel sheet or Power BI dataset (Microsoft, 2023).

It is possible to get Google Play reviews by utilizing a Google Play Store API or a custom scraping script while ensuring compliance with Google's terms. The extracted data can then be stored in a structured format, such as a CSV file or a database. If necessary, the data can be manually inputted into a Teams Form. Subsequently, Power BI can be employed to import the stored review data for analysis and to generate interactive dashboards featuring sentiment analysis, trends, and key insights from the review data.

There is a way to get Google Play reviews but following these steps can be quite challenging. The model uses a short and less time-consuming method for extraction of Google Play data for analysis providing a better solution.

AI models, especially those utilizing machine learning and natural language processing, could be transforming for sentiment analysis across various industries not just taxi apps. Their scalability allows for the rapid analysis of large datasets, allowing them to process thousands of customer reviews and social media posts in real time, which are tasks difficult and inefficient for human analysts.

Furthermore, AI's context understanding enabled a deeper understanding of words and phrases, being crucial for accurately interpreting customer feedback. With the help of real-time analysis, businesses can monitor brand reputation and customer satisfaction effectively, allowing them to address and act on complaints quickly. The AI is cost-efficient with automated sentiment analysis tools to reduce the financial struggles associated with manual methods. Additionally, AI is adaptable and allows evolution by understanding new data and capturing changing trends and slang. Finally, AI's ability to extract upcoming trends in words helps businesses to make informed, strategic decisions, enhancing proactive customer engagement and placing them ahead in the marketplace.

Despite significant advancements, AI models still face challenges in accurately understanding content, such as sarcasm and irony, leading to inaccurate understandings, particularly within informal texts such as social media posts.

One of the biggest challenges of using AI for sentiment analysis is that it highly depends on quality data training for it to perform its functions accurately. Additionally, privacy concerns arise when companies choose not to anonymize users' identities during the sentiment analysis process and use. AI also has a lack of emotional intelligence meaning it cannot fully with human emotional context, therefore, resulting in inaccurate outputs.

There have been numerous projects that utilized a similar approach of using AI models for sentiment analysis to the one employed by this model for sentiment analysis. Analyzing reviews is a common practice. For example, a project titled "Sentiment Analysis of Product-Based Reviews Using Machine Learning Approaches" was completed by a team of students at the RCC Institute of Information Technology. They utilized Python and various NLP (Natural Language Processing) libraries for this project. The tools used included data preprocessing and sentiment classification using algorithms such as Naive Bayes, Logistic Regression, Random Forest, and Support Vector Machine.

Logistic regression is also a good choice for sentiment analysis (Liu, 2012). Logistic regression is a commonly used machine learning algorithm in natural language processing, which classifies texts into positive, neutral, and negative categories. Unlike VADER, Logic Regression does not assign a numerical value but instead produces an output with a probability ranging from 0 to 1, indicating the positivity, negativity, or neutrality of a text it assigns the category with the highest probability. Logistic regression is simple to implement, works fast and efficiently, and performs well in binary classification, efficiently categorizing statements as positive or negative (Liu, 2012)..

Logistic regression may struggle with handling large volumes of data and diverse views. While, VADER has been proven to be better for sentiment analysis in social media language and works well with intensive words, such as "very," to emphasize sentiment. . (Kinza Yasar, et al., 2024)

Random Forest is a complex machine learning algorithm often used in sentiment analysis for classifying text into positive, negative, or neutral. It builds multiple decision trees from random subsets of data and features and then averages their predictions to reduce overfitting and improve accuracy. Random forests handle larger volumes of data and non-linear relationships in text performing robustly in high-dimensional spaces like n-grams or TF-IDF features. (Breiman, 2001)

Random forests can be intensive and less interpretable than simpler models like logistic regression and VADER. The machine learning model is harder to understand and interpret, especially for a company that doesn't have expertise in these complex models.

The report used Amazon reviews from amazon.com. It then employed stop words to eliminate extraneous words. (Tony Yiu 2019)

The evaluation of classification models in machine learning relies on several key metrics that assess performance based on predictions compared to actual labels. These metrics include True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). Accuracy, defined as the proportion of correct predictions out of total predictions, provides a general sense

of model performance but can be misleading in imbalanced datasets. Recall, or sensitivity, measures the ability to identify positive instances, making it crucial in contexts like disease detection where missing positives can have severe consequences. Precision, on the other hand, gauges how many predicted positives are truly positive, which is particularly relevant in scenarios where false positives are detrimental, such as spam filtering. Specificity focuses on correctly identifying negative instances, important in contexts where true negatives are significant. The F1 Score incorporates both precision and recall into a single metric, offering a balanced view when there's a tradeoff between the two, especially useful in uneven class distributions. Collectively, these metrics provide valuable insights into the strengths and weaknesses of classification models, guiding decisions in model evaluation and selection.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{60+100}{60+100+20+20} = \frac{160}{200} = 0.8$$

$$Recall = \frac{TP}{TP+FN} = \frac{60}{60+20} = \frac{60}{80} = 0.75$$

$$Precision = \frac{TP}{TP+FP} = \frac{60}{60+20} = \frac{60}{80} = 0.75$$

$$Specificity = \frac{TN}{TN+FP} = \frac{100}{100+20} = \frac{100}{120} \approx 0.833$$

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} = 2 \times \frac{0.75 \times 0.75}{0.75 + 0.75} = 0.75$$

Figure 1: evaluation of classification models in machine learning. Picture taken by Aradhya Kaushik of the confusion matrix formulas.

	Predicted 0	Predicted 1
Actual 0	TN	FP
Actual 1	FN	TP

Figure 2: Image displaying the structure of the confusion matrix used to calculate the accuracy of classification models .Screenshot taken by Aradhya Kaushik.

The image displays the structure of the **confusion matrix** used to calculate the accuracy of classification models. The matrix shows four quadrants representing the counts of:

1. **True Negatives (TN)**: Correctly predicted negative instances (blue).
2. **False Positives (FP)**: Incorrectly predicted positive instances (cyan).
3. **False Negatives (FN)**: Incorrectly predicted negative instances (green).
4. **True Positives (TP)**: Correctly predicted positive instances (light green).

The accuracy of the different models was calculated using the confusion matrix, which follows a certain format.

While sentiment analysis is a common way of evaluating customer feedback, this project provides a solution that will stand out due to its unique combination of AI-based models using Python-based NLTK libraries like VADER, mathematical formulas, and Power BI-based for visualization. By using alternate NLP(Natural Language Processing) based machine learning models, more accurate results can be achieved. Companies often use readily available review platforms to view customer reviews, but this can be inconvenient as there are typically thousands of reviews to sort through.

Companies also conduct surveys to gather customer feedback, but typically only a maximum of 1000 people participate, which means they cannot get reviews that cover the majority of their customer base (Sabrina Fox, 2024). Real-time dashboard analytics can be made available for action proactive analysis by service providers. This is another different approach from where servicing providers download reviews, analyze, and compile them, and then make improvement decisions months after reviews. The model provides the latest reviews of a company or service and provides an analysis by stating the percentage and number of reviews that are neutral, positive, and negative in each Taxi App. The model will be programmed to find specific words and see how many times they were used in the reviews. One more benefit the model provides is the ability to do competitor comparisons: By analyzing reviews of competitor apps (e.g., the top 2 taxi apps), companies can understand how they compare against their rivals. Insights from competitor reviews can highlight areas where a company excels or needs improvement compared to competitors. This can be used as a major advantage.

The taxi app usage has grown over the years. The taxi app industry is currently facing many challenges leading to customer dissatisfaction. On a bigger more global scale taxi apps have inconsistent service quality and slow response times to complaints, (Vinay Jain, 2024)

Problem Statement

The fast growth of taxi apps has led to overwhelming large volumes of reviews pouring into the customer review systems making it very challenging for companies to extract meaningful and important sentiments from those reviews. Reviews of taxi apps are sentiment rich containing valuable feedback on user experience, and service quality and pointing out areas for improvement. Manually analyzing is not as efficient does reviewing data due to the large volumes

of unstructured data nature reviews. The analysis process is also prone to human error. Other methods of getting customers like surveys have been proven to be inefficient as they only capture a small group of the customer base.

To address these challenges, an AI-driven sentiment analysis model is needed that can identify these reviews into positive negative, and neutral sentiments and visualize the real-time outputs to make it user-friendly for service providers.

Without the adaptation of an automated solution, companies risk losing their customer base by not addressing customer feedback. This decreases user satisfaction, service quality, and a loss of competitive advantage. This is crucial for highly competitive industries like taxi services. The company can maintain customer trust and improve its services by addressing issues in the reviews by using an AI-based customer sentiment analysis model.

Objectives

- To develop a method to assist companies and service providers in analyzing large volume customer sentiment data provided through reviews. The project also aims to reduce to time needed to analyze large product and service data.
- To develop an AI-based model that uses sentiment analysis for evaluating customer feedback on the commonly used 2 taxi apps (App 1 and App 2) to predict sentiments as negative, positive, and neutral as output.
- To create an interactive dashboard in Power BI to visualize and summarize the category-wise sentiments output for customer reviews.
- To evaluate the top 20 words in each category (negative positive and neutral) to gain deeper insights into customer feedback.

Engineering Goal

The aim is to develop an AI-based customer sentiment analysis model that performs sentiment analysis on taxi app reviews with the use of natural language processing techniques. The model will use a pre-trained model called VADER for performing sentiment classifying the reviews as neutral, positive, and negative at 98% accuracy and visualize real-time results in an interactive power BI dashboard. The dashboard will give a summarized view of the reviews that were identified as neutral, positive, and negative and give a summarized score analysis of the ratings given in the reviews. The model will make a word cloud analysis that will include the most words used in each category and also do a top 20-word analysis of words used in each category.

Research Questions

1. How accurately can an AI-based sentiment analysis model classify customer reviews into positive, negative, and neutral sentiments for taxi apps?

Python will be used as a programming language. Google API will be used to extract the review data from the Play Store from Taxi App One and Taxi App 2. Extracted data is organized in tabular format, cleansed, and prepared for modeling. Then NLP(Natural Language Processing) based model is built and trained on the extracted data. Various NLTK and other Python libraries are used to build the model. The data will then be stored, and the model will perform sentiment analysis. Initially, a pre-trained Vader model library is used for modeling. The model will then evaluate and predict the sentiments as negative, positive, and neutral for the reviews.

2. What are the most frequent words used in positive, negative, and neutral reviews for the selected taxi apps, and how do these words reflect customer sentiment

To implement sentiment analysis in a real-life business, and integrate it into customer feedback systems to gauge public sentiment from reviews, social media, and surveys. Use natural language processing (NLP) tools to analyze and categorize sentiments, such as positive, negative, or neutral. This analysis can provide actionable insights into customer satisfaction, product performance, and brand perception. Incorporate these insights into business strategies by addressing negative feedback, enhancing customer service, and tailoring marketing campaigns. Tools like Power BI can visualize these insights, helping to make data-driven decisions and improve overall business performance.

3. What are the potential limitations of using sentiment analysis models, such as VADER, in categorizing customer reviews for taxi apps, and how do these limitations impact the results?

To test for accuracy, the data is split into training and testing data with a 70% and 30% split respectively. The model produces the sentiment prediction using VADER (Value Aware Dictionary and Sentiment Reasoner) on all the data. The outputs are compared between the training data model and the testing data model using rating numbers as the baseline and finding the matching count of reviews identified.

4. How do sentiment trends in customer reviews for the two taxi apps compare, and what actionable insights can be drawn to improve service quality and customer satisfaction?

The accuracy of the different models was done using the confusion matrix that follows the certain format shown above. The accuracy of all these models was in the 80s which is still a good accuracy the model Vader which is what our model uses is much higher in the 90s proving to be better. One of the major reasons to choose VADER was because it has an average accuracy of 96 % and its function to work in a social media context.

Method

Platforms/Tools

- Commercially Available Application Programming Interfaces
- Open source customer review data on selected Taxi App 1 and Taxi App 2
- highh-Level general purpose programming language
- Notebook Authorising Application Jupyter Notebook (7.0.8)
- Anaconda 3
- Database (DB) tools provided in Anaconda (open-source base)
- Pandas
- NumPy
- (Application Data Extraction Tool)Google Play scraper Google-play-scraper
- Import CSVs
- Textblob
- NumPy as np
- pandas as pd
- Visualization packages:
 - matplotlib.pyplot as plt
 - Seaborn
 - word cloud
- Nltk packages –
 - import nltk
 - from nltk. Sentiment.Vader import Sentiment Intensity Analyzer
 - nltk.download("vader_lexicon")
 - nltk.download('stopwords')
 - nltk.download('wordnet')
 - nltk.download('omw-1.4')
 - from nltk.corpus import stopwords
 - from text blob import TextBlob
 - from text blob import Word
- Power BI

Procedure:

The model will be constructed to do as follows :

- Perform data extraction by extracting updated customer reviews, about the experience taxi app users have, from the Google Play Store for each taxi app using Python.
- Perform data pre-processing and clean the data for convenient use using Python and DB tools.

- Perform text-processing and text mining to extract necessary and meaningful words from customer reviews using relevant libraries and algorithms using Python.
- Categorize the reviews into positive, negative, and neutral using Python.
- Generate overall feedback representation based on customer reviews for each taxi app using Python, DB tool, and Power BI.
- Implement an effective dashboard with detailed sentiment analysis and customer scores using Python, DB tool, and Power BI.

Steps:

- **Step 1: Data collection**

- Use Google API to extract data from the open-source data repository from a well-known App Gallery website
- Provide real-time input if needed for analysis.
- Collect data by itself in real-time to ensure accuracy and updated reviews.
- Make use of an API that works on the web.

- **Step 2: Data Understanding & Preparation**

- Format collected data in a clear and understandable structure using Python.
- Data cleaning and analysis of extracted data with Python
 - Organize the data in tabular form.
 - Find out missing data rows.
- Performing text-mining and using lexicon features such as stemming, removing stop words, etc. should be effectively implemented to ensure the relevance and reliability of customer reviews using Python and VADER.
- Anonymize Reviews
- Store data in a panda's data frame to maintain structure and for convenient Python analysis.
- Prepare Data and transform to prepare data for modeling using Python.

- **Step 3: Data Modelling & Evaluation**

- Use a pre-trained Vader library based pre-trained model for sentiment analysis
- The data for each Taxi App 1 and 2 should be summarized in tables and graphs to access summarized information for each Taxi App.
- Create a Confusion matrix for each model.
- Choose the most accurate and low-complex model for output.

- **Step 4: Data storage**
 - Store data in any DB tool for real-time analysis.
- **Step 5: Sentiment Analysis model programming**
 - Data modeling based on NLP(Natural Language Processing) which is a type of text classification.
 - Categorize reviews into negative, positive, and neutral sentiments.
 - Show word cloud with the most frequently used words
- **Step 6: Visualization and Analytics using Power BI and Python**
 - Present data in a dashboard format
 - Identify the word the company seeks to find in the review and visualize it into a graph.
 - Summarize sentiments analysis outcome.
- **Step 7: Test data for modeling**
 - 466 known data values would be used to test the models for accuracy in taxi app 1 and 426 data values for taxi app 2. The data values are the taxi app reviews of the customers.
 - 3500 unknown review data is used to train the model
 - Testing of the model using predetermined or known sentiments.
 - Compare the model results to the actual results.

Model Diagram

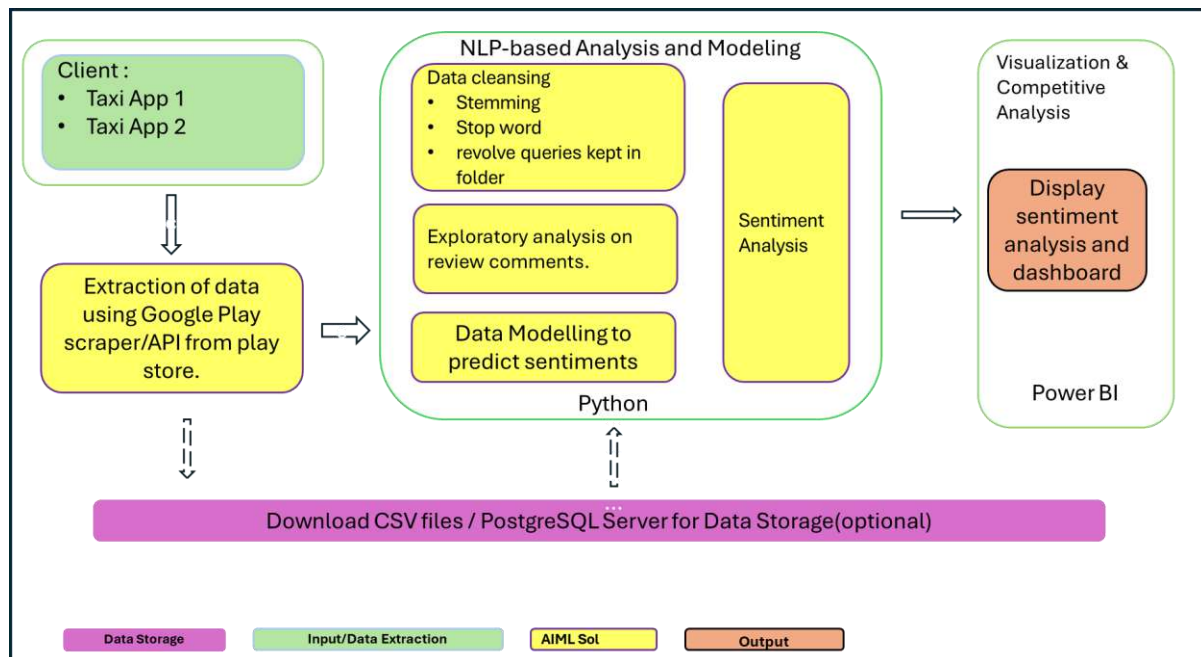


Figure 3: The workflow for sentiment analysis of reviews from two taxi apps involves four primary stages. Flow Diagram showing made by Kaushik 2024

The workflow for sentiment analysis of reviews from two taxi apps involves four primary stages. First, data extraction occurs, where information is gathered from the Google Play Store using a scraper or API and then stored in CSV files or a PostgreSQL database. The second stage encompasses NLP-based analysis and modeling, which includes data cleansing techniques such as stemming, removing stop words, and processing queries in a designated folder. Following this, exploratory analysis is conducted to identify patterns in the reviews, and Python is employed to build predictive models for sentiment analysis. The visualization stage showcases the results of this analysis using Power BI, offering competitive insights. Finally, there's an optional storage and output section, integrating the data storage (in CSV or database form) with the entire pipeline. Overall, this comprehensive setup combines data engineering, machine learning, and visualization to analyze user reviews and deliver actionable insights.

Execution

Data Extraction

- Data for Taxi App 1 and Taxi App 2 as listed below will be used to extract customer reviews from the Google Play Store by installing Google Play Scraper and generating output as *.csv for Taxi App 1 and Taxi App 2 in real-time :
 - Taxi App 1
 - Taxi App 2

```
from google_play_scraper import Sort, reviews_all

#Scraping required reviews
sa_reviews = reviews_all(
    'ee.mtakso.client',
    sleep_milliseconds=0, # defaults to 0
    lang='en', # defaults to 'en'
    country='sa', # defaults to 'sa'
    sort=Sort.NEWEST, # defaults to Sort.MOST_RELEVANT
)
```

Figure 4: Screenshot showing the data extraction for taxi app 1 from the code of the model from Jupyter Notebook taken by Aradhya Kaushik.

```
from google_play_scraper import Sort, reviews

sa_reviews, continuation_token = reviews(
    'com.ubercab',
    lang='en', # defaults to 'en'
    country='sa', # defaults to 'us'
    sort=Sort.NEWEST, # defaults to Sort.NEWEST
    count=500, # defaults to 100
    #filter_score_with=5 # defaults to None(means all score)
```

Figure 5: Screenshot showing data extraction for taxi app 1 from the code of the model from Jupyter Notebook taken by Aradhya Kaushik.

	reviewId	userName	userImage	content	score	thumbsUpCount	reviewCreatedVersion	at	replyContent	repliedAt	appVers
0	1c732234-556c-4380-8b7c-78b121d8449d	Raphael Anthony	lh.googleusercontent.com/a-/ALV-U...	nice ride	5	0	CA.108.0	2024-04-20 20:17:07	None	NaT	CA.10
1	c15a34f5-2156-4634-a383-2fd7851dacce	Kilenga Naftal	lh.googleusercontent.com/a/ACg8oc...	I enjoyed the service.	5	0	CA.60.1	2024-04-20 20:10:27	None	NaT	CA.6

This driver

Figure 6: Screenshot showing data samples in the code of the model from Jupyter Notebook taken by Aradhya Kaushik.

Data Understanding, Preparation, Data Modelling and evaluation

- This is a key step to understanding data structure, statistics, and quality of data and to figuring out useful portions of data for further processing, analysis, and modeling
- This data is then further **cleaned up** to remove unrequired columns.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 500 entries, 0 to 499
Data columns (total 11 columns):
#   Column                Non-Null Count  Dtype
---  -
0   reviewId              500 non-null   object
1   userName              500 non-null   object
2   userImage             500 non-null   object
3   content               500 non-null   object
4   score                 500 non-null   int64
5   thumbsUpCount         500 non-null   int64
6   reviewCreatedVersion  468 non-null   object
7   at                    500 non-null   datetime64[ns]
8   replyContent          2 non-null     object
9   repliedAt             2 non-null     datetime64[ns]
10  appVersion            468 non-null   object
dtypes: datetime64[ns](2), int64(2), object(7)
memory usage: 43.1+ KB
```

Figure 7: Screenshot showing data understanding in the code of the model from Jupyter Notebook taken by Aradhya Kaushik.

```
#Dropping unnecessary columns
df = df.drop(['replyContent', 'repliedAt', 'reviewCreatedVersion', 'thumbsUpCount
```

Figure 8: Screenshot showing data understanding in the code of the model from Jupyter Notebook taken by Aradhya Kaushik.

- For exploratory data analysis (EDA) it is critical to efficiently summarize the analyzed data. The number of comments in each positive, negative, or neutral category is obtained and represented graphically.

Data pre-processing

This section focuses on the content of the user comments which is the most critical aspect of our data. The Natural Language Toolkit (NLTK) and Textblob are used to process the text within the customer reviews. Stop words are removed from the reviews using relevant features from the NLTK libraries like. The top 3500 comments for each Taxi App are used to perform sentiment analysis.

```
new_df=df[df['Sentiment']=='Negative']
words = ' '.join(df['review'])

wordcloud = WordCloud(stopwords=STOPWORDS,
                       background_color='black',
                       colormap='Paired',
                       width=3000,
                       height=2500
                       ).generate(words)
plt.figure(1,figsize=(12, 12))
plt.imshow(wordcloud)
plt.axis('off')
plt.title('Reviews with Negative Sentiments',fontsize=25)
plt.show()
```

Figure 9: Screenshot data pre-processing in the code of the model from Jupyter Notebook taken by Aradhya Kaushik.

Sentiment analysis (conceptual & logical definition)

- This section focuses on analyzing the comments of each user. This is done using the NLP (Natural Language Processing) **Sentiment Intensity Analyzer**.

```
: # Calculating all polarity scores

sentiments = SentimentIntensityAnalyzer()
df["Positive"] = [sentiments.polarity_scores(i)["pos"] for i in df["review"]]
df["Negative"] = [sentiments.polarity_scores(i)["neg"] for i in df["review"]]
df["Neutral"] = [sentiments.polarity_scores(i)["neu"] for i in df["review"]]
df['Compound'] = [sentiments.polarity_scores(i)["compound"] for i in df["review"]]
```

Figure 10: Screenshot showing sentiment analysis in the code of the model from Jupyter Notebook taken by Aradhya Kaushik.

```
'''Adding Labels to categorize the scores calculated above
Industry Standard says "if compound score is >0.05 then Positive, if <-0.05 then
Knowing this i'm adding a new column called Sentiment to hold the sentiments'''

score = df["Compound"].values
sentiment = []
for i in score:
    if i >= 0.05 :
        sentiment.append('Positive')
    elif i <= -0.05 :
        sentiment.append('Negative')
    else:
        sentiment.append('Neutral')
df["Sentiment"] = sentiment
```

CCC
If compound_score > 0.05 then sentiment_label
= positive
If compound_score < -0.05 then sentiment_label
= negative
else sentiment_label = neutral

Figure 11: Screenshot showing sentiment analysis in the code of the model from Jupyter Notebook taken by Aradhya Kaushik.

- Thus, each comment is categorized with a positive, negative, or neutral sentiment label.

	reviewId	userName	userImage	review	score	createdat	appVersion	Positive	Negative	Neutral	Compound	Sentiment
0	1c732234-556c-4380-8b7c-78b121d8449d	Raphael Anthony	lh.googleusercontent.com/a-/ALV-U...	nice ride	5	2024-04-20 20:17:07	CA.108.0	0.737	0.000	0.263	0.4215	Positive
1	c15a34f5-2156-4634-a383-2fd7851dacce	Kilenga Naftal	lh.googleusercontent.com/a/ACg8oc...	enjoyed service.	5	2024-04-20 20:10:27	CA.60.1	0.767	0.000	0.233	0.5106	Positive
2	eb771783-a526-4abb-bc5e-f55f16ded5bd	Gloria Asuquo	lh.googleusercontent.com/a/ACg8oc...	driver came pick hotel john chika odiri peugeo...	1	2024-04-20 20:09:05	CA.112.0	0.228	0.000	0.772	0.7579	Positive

Figure 12: Screenshot showing sentiment analysis in the code of the model from Jupyter Notebook taken by Aradhya Kaushik.

Dashboard and Visualization

- For visualization Python library Marplot lib was used.
- Power BI was used for user-friendly dashboards.
- The Power BI dashboard included
 - A bar graph showing the rating of the reviews
 - A pie chart showing the amount of positive, negative, and neutral reviews
 - A question-and-answer section (word search)
- The dashboard is being represented in Power BI so it is very interactive, and the user can choose their choice of visual they like.
- The report can be published to the service providers if they don't have Power BI.

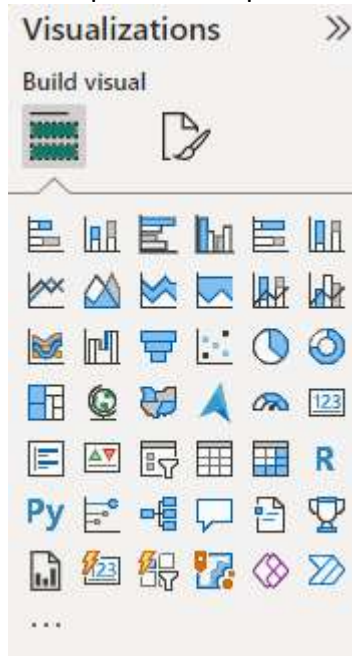


Figure 13: Screenshot of the Power BI visualization options taken by Aradhya

Results

Sentiments for Taxi App 1 and Taxi App 2

The model was successfully constructed as per the method. 500 reviews were used to build the model, and 466 reviews were used to test the model in Taxi App 1 and 426 in Taxi App 2.

For taxi app 1, 71.8% were classified as positive compared with 7% for taxi app 2. For taxi app 1, 10.3 % were classified as negative compared with 14.5 % for taxi app 2. For taxi app 1, 17.9% were classified as neutral compared with 14.5 % for taxi app 2.

Table 1 Table showing the number and percentage of reviews classified by the model into positive, negative, or neutral for Taxi apps 1 and 2.

Taxi App	Positive	Negative	Neutral
Taxi App 1	333(71,46%)	46(9,87%)	87(18,70%)
Taxi App 2	293(68,80%)	61(14,32%)	72(16,90%)

Taxi App 1 – Power BI Dashboard



Figure 14: Screenshot of the dashboard for Taxi App 1 from Power BI taken by Aradhya Kaushik.

Taxi App 2 – Power BI Dashboard

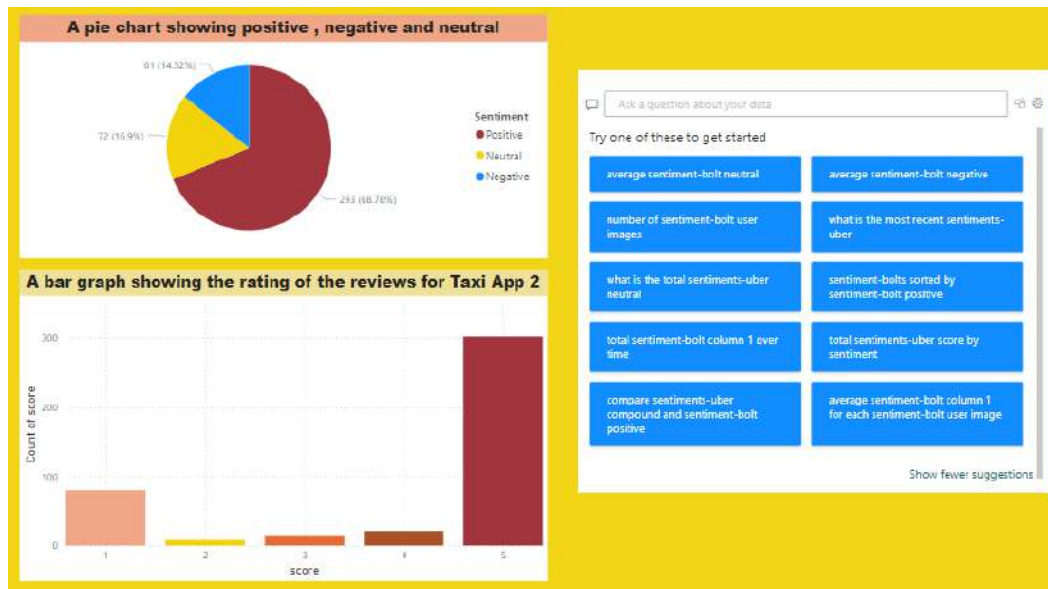


Figure 15: Screenshot of the dashboard for Taxi App 2 from Power BI taken by Aradhya Kaushik

Word cloud analysis:

Some words were commonly identified in the positive, neutral, and negative groups for both taxi apps such as nice, good, great, problem, etc.

Taxi App 1 :



Figure 16: Screenshot of Word Cloud Analysis for negative words in Taxi App 1 from Jupyter Notebook taken by Aradhya Kaushik.



Figure 17: Screenshot of Word Cloud Analysis for Neutral Words in Taxi App 1 from Jupyter Notebook taken by Aradhya Kaushik.



Figure 18: Screenshot of Word Cloud Analysis for positive words in Taxi App 1 from Jupyter Notebook taken by Aradhya Kaushik.

Tabl 2: Table showing some of the words found in the positive, negative, and neutral for Taxi app 1's word cloud analysis.

POSITIVE	NEGATIVE	NEUTRAL
Nice	Cancelling	Friendly
Good	Wrong	Nice
Great	Tired	Useful
Safe	Rude	Rude
Best	Bad	Time
Perfect	Lost	Happy
Useful	Expensive	Always
Cheap	Cancel	Wait
Affordable	Disrespect	Respect

Taxi App 2 :



Figure 19: . Screenshot of Word Cloud Analysis for negative words in Taxi App 2 from Jupyter Notebook taken by Aradhya Kaushik.



Figure 20: . Screenshot of Word Cloud Analysis for positive words in Taxi App 2 from Jupyter Notebook taken by Aradhya Kaushik.



Figure 21: . Screenshot of Word Cloud Analysis for neutral words in Taxi App 1 from Jupyter Notebook taken by Aradhya Kaushik.

Table 3: Table showing some of the words found in the positive, negative, and neutral for Taxi app 2's word cloud analysis.

Positive	Negative	Neutral
driver	Good	nice
service	Bad	Bad
time	Long	Driver
app	Wait	Time
nice	Driver	Wait
great	Expensive	Good
Friendly	Costly	Behavior
best	Nice	Safe
price	Improve	Quick
fast	Driver	Friendly
reliable	Limit	Limited
professional	Experience	Space
rude	Cancel	Cancel
Cancel	App	Fast
Affordable	great	Car

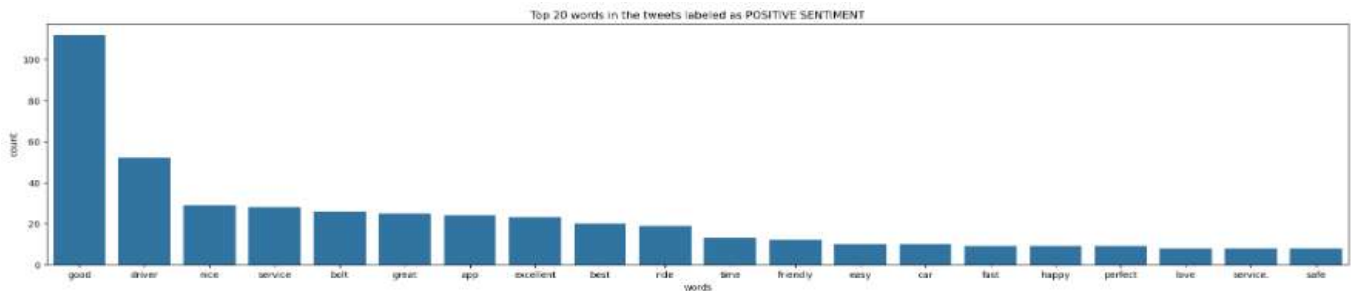


Figure 22: . Screenshot of Top 20 for positive words in Taxi App 1 from Jupyter Notebook taken by Aradhya Kaushik.

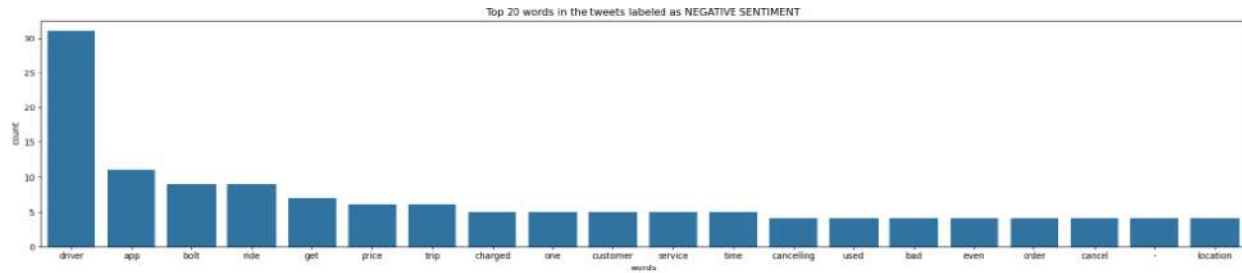


Figure 23: . Screenshot of Top 20 for negative words in Taxi App 1 from Jupyter Notebook taken by Aradhya Kaushik.

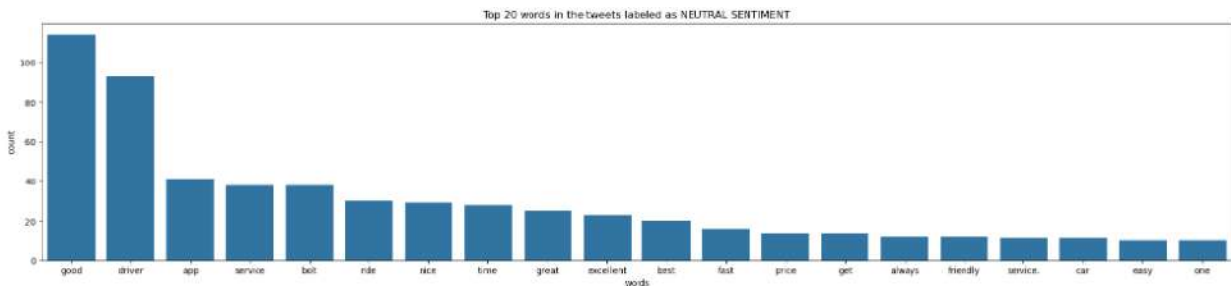


Figure 24: . Screenshot of Top 20 for neutral words in Taxi App 1 from Jupyter Notebook taken by Aradhya Kaushik.

Taxi App 2 :

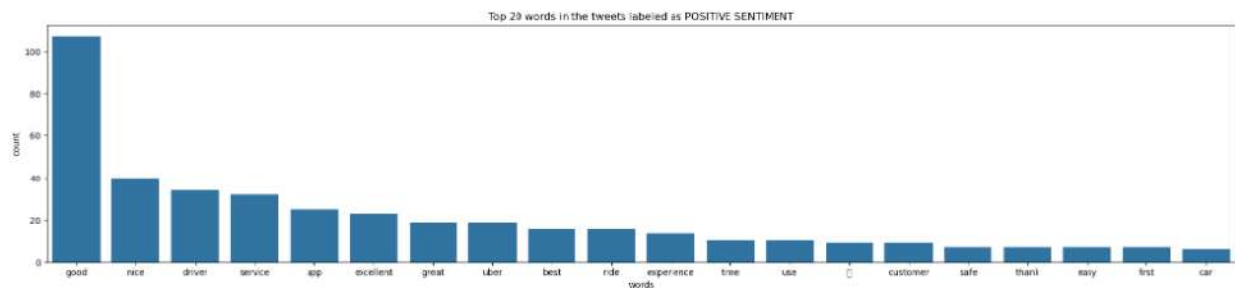


Figure 25: . Screenshot of Top 20 for positive words in Taxi App 2 from Jupyter Notebook taken by Aradhya Kaushik.

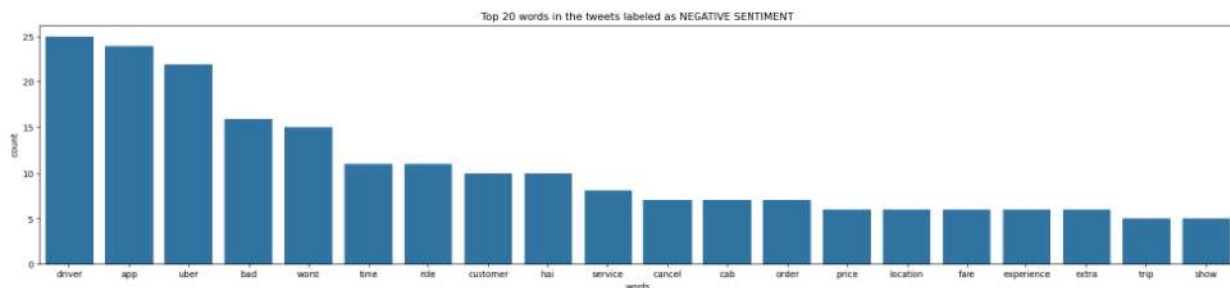


Figure 26: . Screenshot of Top 20 for negative words in Taxi App 2 from Jupyter Notebook taken by Aradhya Kaushik.

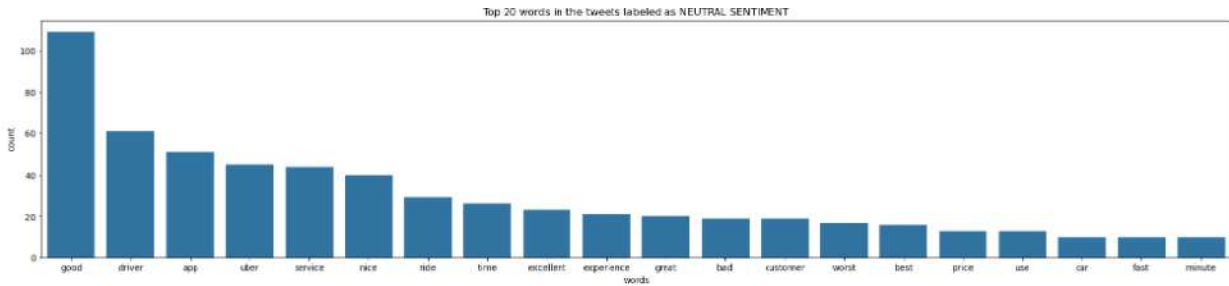


Figure 27: . Screenshot of Top 20 for neutral words in Taxi App 2 from Jupyter Notebook taken by Aradhya Kaushik.

Result Accuracy

The project involved taking a sample data set of 466 for Taxi App 1 and 426 for Taxi App 2 review, where the rating scores of those reviews were assumed to determine their base sentiment classification. Reviews with a score of one and two were classified as negative, reviews with a score of three were classified as neutral, and reviews with a score of four and five were classified as positive sentiments. The output data from the sentiment analysis model was then compared to these scoring-based sentiment classifications to determine the accuracy percentage of the model overall for each sentiment category. The accuracy was based on the matching number of reviews. If the model identified more positive reviews than the baseline (rating numbers) then the model had false positives and was not accurate.

Taxi App 1 :

Table 4: Table showing the accuracy calculation for Taxi App 1

Taxi App 1	Prediction	SCORES			Match count	Accuracy (% Match)
	Count	rating	count	sum		
Positive	333	4	42	351	333	94.9%
		5	309			
Neutral	87	3	18	18	18	20.7%
Negative	46	2	20	97	46	47.4%
		1	77			
Total	466			466	397	85.2%

Taxi app 2 :

Table 5: Table showing the accuracy calculation for Taxi App 2

Taxi App 2	Prediction	SCORES			Match count	Accuracy (% Match)
	Count	rating	count	sum		
Positive	293	4	21	323	293	90.7%
		5	302			
Neutral	72	3	15	15	15	20.8%
Negative	61	2	8	88	61	69.3%
		1	80			
Total	426			426	369	86.6%

The results showed sentiments, with Taxi App 1 having 333 positive, 46 negative, and 87 neutral reviews, while Taxi App 2 had 293 positive, 61 negative, and 72 neutral reviews. The model for taxi app 1 had an accuracy of 85.2% and 86.6% for taxi app 2.

Discussion

The results came to be accurate, but the model failed to classify all the reviews as positive negative, and neutral accurately. The engineering goal was for a 98% accuracy in classifying the sentiments.

The main reason why many of the reviews weren't classified accurately was because the content in them would have been sarcastic or indirect. The model wasn't trained to understand so it is listed as one of the limitations of the model.

A concern that was raised was words like "good" and "affordable" were showing up in negative and neutral word clouds. The model isn't trained to handle indirect and sarcastic comments therefore it takes the statement literally and therefore the words show up in the word cloud analysis. However, many people say comments like "the car ride experience was good, but I struggled with using the app" can be taken as a neutral and negative sentiment. This is the reason why some words like good and affordable may show up in negative and neutral sentiments.

The model was successfully able to handle large volumes of data. The model provides an advantage where it can provide real-time analysis. The model had an 85,2% accuracy in Taxi App 1 which when you compare it to our engineering goal, is 98%. This means there was a 13% difference. The model's accuracy can be improved if a more accurate NLP (Natural Language Processing) toolkit can be found instead of VADER.

Some reviews that might have affected the accuracy such as false negatives and positives have been identified. Models like VADER provide valuable insights into customers' sentiments, but they come without their faults.

Sentiment analysis models often struggle with false negatives and positives. A false positive is when the model incorrectly identifies a negative or neutral statement as positive. And false negative is when a model doesn't recognize a positive sentiment and puts it as a negative or neutral statement.

Vader still can make mistakes with analyzing as he has a limited understanding of context. This means that it cannot analyze the sarcastic or indirect meaning in text which can lead to inaccurate results as there may be many false negatives and false positives. VADER is better for more straightforward short phrases. A very long text may confuse the model leading to incorrect sentiment analysis.

The word cloud analysis and TOP 20 words are directly influenced by these results therefore if a positive word shows up in the negative word cloud it means there was a mistake or many of those were genuinely used.

False positives occur when reviews are incorrectly classified as positive due to factors such as sarcasm, mixed sentiment, or polite language with negative contexts. For example, a sarcastic comment like, "Oh great, the app crashed for the fifth time today. Fantastic work!" might be misinterpreted as positive because of the words "great" and "fantastic." Similarly, a statement like, "The app interface is nice, but the rides are always late, and customer service is unhelpful," may get a positive classification due to the favorable mention of the interface, despite expressing significant complaints. In another case, phrases such as, "I appreciate the effort, but the app doesn't meet my needs," can mislead models to categorize them as positive because of the term "appreciate." On the other hand, false negatives occur when positive sentiments are incorrectly classified as negative. An example is when someone states, "It took some time to figure out, but this is the best taxi app I've used," where the initial criticism might overshadow the final positive evaluation. Likewise, understated praise such as, "Not bad, I guess it works okay," might be perceived as negative due to its seemingly lukewarm wording. Furthermore, informal language, as seen in phrases like "Gr8 app, luv it!" may confuse models that are primarily trained on formal text.

VADER can be inaccurate in making a review that is positive or negative to a human reader and saying it's neutral. However, monitoring the model with other feedback systems, an AI-based sentiment dashboard that uses tools like VADER offers businesses the insights they need to maintain a good reputation.

This will provide a way for service providers to make a business decision based on data insights. Service providers can take a glance at the consolidated and summarized view of customer reviews and their sentiments. Categorized data can be used for further root cause analysis, and the same will help service providers make an action plan to improve their application for a better user

experience. This way service providers can save time by not having to read through every review online. Instead, they can have an AI-based model do it for them, providing a more efficient and accurate way to analyze customer feedback. The model has many advantages over conventional surveys including that the model deals with a larger number of reviews whereas the survey only has a maximum of 1000 responses.

The sentiment analyzer evaluates customer reviews by assigning a percentage score for positive, negative, and neutral sentiments based on the words used. These scores collectively determine the compound score, representing an overall sentiment. For instance, a review might have scores like Positive: 0.196, Negative: 0.399, Neutral: 0.405, with a compound score of -0.250. The final sentiment label is determined by this compound score: if it is greater than 0.05, the review is labeled as positive; if less than -0.05, it's labeled as negative. Otherwise, the sentiment is considered neutral. This approach helps categorize each review as positive, negative, or neutral.

The model was trained on the 500 most recent reviews for each taxi app. However, there was a difference in the number of reviews used during sentiment analysis. This discrepancy occurred because some reviews contained only emojis and no words, which were not considered valid during the scraping process. Consequently, during pre-text processing, these reviews were removed.

For Taxi App 1, the model started with 500 reviews but ended up with 466, resulting in a difference of 34 reviews. In contrast, Taxi App 2 also began with 500 reviews but concluded with 426, leading to a difference of 74 reviews.

One of the most prominent words found in the word cloud analysis was “driver”. There are various reasons for this as a lot of reviews that were given for the taxi apps are based on the experience the customers had in the car driven by the driver. Customers may have felt a positive negative or neutral sentiment towards the drivers resulting in drivers as a prominent word in the word cloud analysis.

Limitations and Errors

Lack of contextual awareness: The sentiment analyzer depends on the exact words used by a user in the review. This ignores the possibility of backhanded compliments or the use of figurative language. Any ambiguity within a review can also go unnoticed. This could make the classified sentiment of feedback incorrect. The chances of casual language usage and ambiguity should be considered since the user base for a taxi app is very wide.

Language limitations and Emoticons: The language for this model is set to English however some users tend to insert words from other local languages or include slang words that might not be included in the sentiment analyzer. This would reduce the accuracy of the result of the analyzer.

Recommendations for future development

Real-Time Sentiment Analysis App: Make an app for both mobile and web that lets users log in and see real-time sentiment analysis of recent customer reviews. This makes sure users can log in securely with their usernames and passwords to see personalized dashboards. The app should automatically pull the latest reviews from platforms like the Google Play Store, App Store, and social media, updating the sentiment analysis in real-time. Users should be able to customize their dashboards to show the most relevant metrics and insights, such as sentiment trends, word clouds, and top keywords. A feature to send notifications or alerts to users when there is a significant change in sentiment can be added so they can take immediate action.

High-Accuracy Sentiment Identification System: This idea is to develop or integrate more advanced Natural Language Processing (NLP) models that can better understand sentiments in reviews. Implement a system that continuously trains and updates the sentiment analysis model with new data to maintain and improve accuracy over time. Improve the system's ability to understand context, sarcasm, and nuanced language to better identify the true sentiment of reviews. Ensure that the system can accurately analyze reviews from different platforms and formats, considering platform-specific language usage and rating systems.

Integration with ChatGPT for Problem Solving: Adding a ChatGPT feature to the model where service providers can explore insights and issues in the reviews. Like asking specific questions about the data, like "What are the most common complaints?" or "How can we improve customer satisfaction?" and get detailed, AI-generated responses.

Feedback Loop: Create a system where the chatbot suggests potential actions or improvements, and users can give feedback on the effectiveness of these suggestions to improve the system's recommendations.

Conclusion

This project aimed to develop an AI-based model using sentiment analysis for evaluating customer feedback on two taxi apps and to visualize the results using Power BI. The model successfully categorizes customer reviews into positive, negative, and neutral sentiments. It further analyzes the data to extract the most frequently used words within each sentiment category and visualizes them through word clouds and graphs. The integration of Power BI provides a user-friendly, interactive dashboard that allows service providers to gain insights quickly and efficiently.

The model successfully utilized the Google Play Store API to collect customer reviews for Taxi App 1 and Taxi App 2. The data extraction process was followed by cleaning and structuring the data using Python, ensuring it was ready for analysis. For sentiment analysis, the model used the VADER model, which effectively categorized the reviews into positive, negative, and neutral sentiments. Each review was assigned a compound score, which was then used to accurately label the sentiment. Power BI was used to visualize the data. These dashboards provided a summarized view of the sentiment analysis and included visualizations such as word clouds. The model performed a random word search of specific words within reviews, offering targeted insights for service providers. The model's reliability and accuracy were tested by using the sentiments of the known reviews and seeing if the model successfully identified them. The model successfully identified 93.8% of the reviews accurately in Taxi App 1 and 89.4% in Taxi App 2.

This model provides a significant solution in the way service providers can analyze customer feedback. Traditional methods of manually sorting and analyzing reviews are time-consuming and often impractical due to the large volume of data. This AI-based model provides a scalable, efficient, and accurate solution for sentiment analysis and data visualization. With the power of machine learning and natural language processing, service providers can gain deeper insights into customer sentiments and make informed decisions to enhance user experience and satisfaction.

The development of this model also highlights the importance of integrating advanced technologies like AI and data visualization tools in modern business practices. It demonstrates how technology can be used to transform raw data into meaningful insights, driving better business outcomes and fostering a customer-centric approach.

In conclusion, the AI-based sentiment analysis model, combined with Power BI visualization, offers a robust tool for service providers to understand and act upon customer feedback, ensuring continuous improvement and competitive advantage in the fast-paced online commercial environment.

Acknowledgments

Here I would like to take the opportunity to thank and express my gratitude to all the people who supported me in the completion of this project.

I would like to express my deepest gratitude to my dad, Nitin Kaushik for his support and guidance throughout the development of this model. His extensive knowledge and passion for data science have been invaluable in shaping my understanding and approach to this project. I am grateful for

the time he dedicated to teaching me the concepts of data science and for his constant encouragement.

A special thanks to Mrs. Meg Lester, whose expertise in report writing and critical thinking has greatly enhanced the quality of my work. Her insightful feedback and assistance in making crucial decisions about my model have been instrumental in my progress. I appreciate her patience and dedication to helping me refine my reports and overall project.

I also extend my heartfelt thanks to Dr. Adele Cheddie for her continuous support and guidance. Her contributions have significantly improved my writing skills and provided me with clear direction throughout the project. Her encouragement and valuable advice were essential to the successful completion of this work.

Also special thanks to Kavya Kaushik who helped me with my code and provided me with guidance in many areas of my work.

Thank you all for your invaluable support and contributions

References

1. Alex The Analyst (no date) How to create and manage relationships in Power BI | Microsoft Power BI for beginners. YouTube. Available at: <https://www.youtube.com/watch?v=sW5LoDA1ssM> (Accessed: September 10, 2024).
2. Breiman, L., 2001. Random forests. *Machine Learning*, 45(1), pp.5-32.
3. Chandrika, G. (2024) "Natural Language Processing (NLP)," *International Journal for Research in Applied Science and Engineering Technology*. doi: 10.22214/ijraset.2024.63281.
4. Coursesteach (2024) Machine learning (Part 33)-Regularized Logistic Regression, *Medium*. Available at: <https://medium.com/@Coursesteach/machine-learning-part-33-regularized-logistic-regression-> (Accessed: September 11, 2024).
5. DHARA, A., SAHA, A., SENGUPTA, and S., BOSE, P. (2014). Sentiment Analysis of Pro
6. Form, question, response, and character limits in Microsoft Forms. (n.d.). *Microsoft.com*. Retrieved July 2, 2024, from <https://support.microsoft.com/en-us/office/form-question-response-and-character-limits-in-microsoft-forms-ec15323d-92a4-4c33-bf88-3fdb9e5b5fea> (Accessed: April 27, 2024).
7. Fox, S. (2024). How to determine survey sample size: A guide (no date). *Survicate.com*. Available at: <https://survicate.com/blog/survey-sample-size/> (Accessed: April 30, 2024).
8. Guthrie, L. et al. (1996) "The role of lexicons in natural language processing," *Communications of the ACM*, 39(1), pp. 63–72. (Accessed: September 11, 2024).
9. Hutto, C. and Gilbert, E. (2014) "VADER: A parsimonious rule-based model for sentiment analysis of social media text," *Proceedings of the International AAAI Conference on Web and Social Media*, 8(1), pp. 216–225. Available at: <https://ojs.aaai.org/index.php/ICWSM/article/view/14550> (Accessed: September 12, 2024).

10. Improve, G. (2017) Removing stop words with NLTK in Python. *GeeksforGeeks*. Available at: <https://www.geeksforgeeks.org/removing-stop-words-nltk-python/> (Accessed: September 13, 2024).
11. Kaabar, S. (2024) A-Z machine learning: Random Forest in time series analysis, *Medium*. Available at: <https://medium.com/@kaabar-sofien/a-z-machine-learning-random-forest-in-time-series-analysis-28f3ac185666> (Accessed: September 12, 2024).
12. Liu, B. (2012) *Sentiment Analysis and Opinion Mining*. 1st ed. San Rafael, CA: Morgan & Claypool Publishers.
13. Microsoft support (no date) *Microsoft.com*. Available at: <https://support.microsoft.com/en-US> (Accessed: September 10, 2024).
14. Podolsky, M. (2023) What is review management, and how can businesses do it effectively? *Forbes*. Available at: <https://www.forbes.com/sites/forbesbusinesscouncil/2023/04/13/what-is-review-management-and-how-can-businesses-do-it-effectively/?sh=60a536f461c2> (Accessed: April 30, 2024).
15. RandomForestClassifier (no date) *scikit-learn*. Available at: <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html> (Accessed: September 15, 2024).
16. Shahid, K. (2024) Review management: What it is and how it benefits brands. *Sprout Social*. Available at: <https://sproutsocial.com/insights/review-management/> (Accessed: April 30, 2024).
17. Simplilearn (2022) Sentiment analysis of Amazon reviews | sentiment analysis Python tutorial | ML tutorial. *YouTube*. Available at: <https://www.youtube.com/watch?v=SMecj9AIDF8> (Accessed: April 30, 2024).
18. van Luijtelaaar, M. (2023) Sentiment analysis for Google Business Profile reviews. *Gmbapi.com*. Available at: <https://gmbapi.com/news/get-to-know-your-customers-better-sentiment-analysis-for-google-business-profile-reviews/> (Accessed: April 30, 2024).

19. What is logistic regression? (2021) *Ibm.com*. Available at: <https://www.ibm.com/topics/logistic-regression> (Accessed: September 14, 2024).
20. What is natural language processing? (2024) *Ibm.com*. Available at: <https://www.ibm.com/topics/natural-language-processing> (Accessed: April 30, 2024).
21. What is Python used for? A beginner's guide (2021) *Coursera*. Available at: <https://www.coursera.org/articles/what-is-python-used-for-a-beginners-guide-to-using-python> (Accessed: April 30, 2024).
22. Wikipedia contributors (2024a) Positive and negative predictive values, *Wikipedia, The Free Encyclopedia*. Available at: https://en.wikipedia.org/w/index.php?title=Positive_and_negative_predictive_values&oldid=1239496898 (Accessed: September 11, 2024).
23. Wikipedia contributors (2024b) Precision and recall, *Wikipedia, The Free Encyclopedia*. Available at: https://en.wikipedia.org/w/index.php?title=Precision_and_recall&oldid=1242064846 (Accessed: September 11, 2024).
24. Yiu, T. (2019) Understanding random forest, *Towards Data Science*. Available at: <https://towardsdatascience.com/understanding-random-forest-58381e0602d2> (Accessed: September 10, 2024).
25. (No date a) *Machinelearningmastery.com*. Available at: <https://machinelearningmastery.com/logistic-regression-for-machine-learning/> (Accessed: September 12, 2024).
26. (No date b) *Datacamp.com*. Available at: <https://www.datacamp.com/tutorial/random-forests-classifier-python> (Accessed: September 13, 2024).
27. 1.1. Linear models (no date) *scikit-learn*. Available at: https://scikit-learn.org/stable/modules/linear_model.html (Accessed: September 2, 2024).
28. Kamburov-Niepewna, U. (2021) *7 reasons why customer feedback is important to your business, Startquestion - create online survey and forms*. Available at:

<https://www.startquestion.com/blog/7-reasons-why-customer-feedback-is-important-to-your-business/> (Accessed: December 1, 2024).

29. Needle, F. (2023) *The benefits of customer feedback, according to experts, HubSpot*. Available at: <https://blog.hubspot.com/service/benefits-of-customer-feedback> (Accessed: December 1, 2024).
30. Sarker, I.H. (2021) "Machine learning: Algorithms, real-world applications, and research directions," *SN computer science*, 2(3), pp. 1–21. Available at: <https://doi.org/10.1007/s42979-021-00592-x>.
31. Chowdhury, R., Islam, M., & Rahman, F., 2021. "Sentiment Analysis in Transportation Apps: A Case Study of Taxi Services," *International Journal of Data Science*, 8(2), pp. 104-118.
32. Jun (2023) "What's new in forms," *Techcommunity.microsoft.com*, 30 June. Available at: <https://techcommunity.microsoft.com/blog/microsoftformsblog/what%E2%80%99s-new-in-forms--june-2023/3860976> (Accessed: December 1, 2024).
33. Yasar, K., Lawton, G. and Burns, E. (2024) *What is logistic regression?*, *Search Business Analytics*. TechTarget. Available at: <https://www.techtarget.com/searchbusinessanalytics/definition/logistic-regression> (Accessed: December 2, 2024).
34. (No date) *Datacamp.com*. Available at: <https://www.datacamp.com/tutorial/what-is-a-confusion-matrix-in-machine-learning> (Accessed: December 2, 2024).
35. Jain, J. (2021) *Confusion Matrix: The two types...* - Jayesh Jain, *Medium*. Available at: <https://jjain2201.medium.com/confusion-matrix-the-two-types-d017a8871d69> (Accessed: December 2, 2024).
36. Jain, V. (2024) *What riders want: Insights from customer feedback on taxi apps*, *Appical Applications*. Available at: <https://www.appical.com/blog/what-riders-want-insights-from-customer-feedback-on-taxi-apps.html> (Accessed: December 2, 2024).

【評語】 190033

1. It is suggested to compare the proposed method with more advanced AI methods.
2. Apart from text input, it is also appealing to take speech as input.
3. It remains to be seen whether the proposed method can equally applicable to other languages than English.
4. Customer sentiments analysis is very important to companies to get true and accurate feedback from their customers to upgrade their services.
5. For future work, LLM and voice data can be considered together to provide an even better and faster feedback to serve original and new customers.