**Sentiment Analysis of Vodafone Reviews**

Made by AbdelRahman Berery 202000001

Karim Hussein Ibrahim 202001674

**Introduction**

Sentiment analysis, also known as opinion mining, is a branch of natural language processing (NLP) focused on extracting and analyzing subjective information from text data. The purpose is to determine whether the mood expressed in the text is positive, negative, or neutral. Sentiment analysis has received a lot of attention in recent years due to its wide range of applications, including brand monitoring, market research, customer feedback analysis, and social media analysis.

Vodafone, one of the leading telecommunications companies, has an extensive online presence, where customers often share their experiences and opinions about the company's products and services on various platforms. Trustpilot, a popular consumer review site, has tons of customer reviews about Vodafone. Analyzing these ratings provides valuable insight into customer sentiment and helps Vodafone understand customer perceptions, identify opportunities for improvement and improve customer satisfaction.

To overcome the challenges of analyzing large amounts of text data, this project leveraged the NLTK (Natural Language Toolkit) library, a powerful NLP toolkit written in Python. The focus of the project was to leverage the SentimentIntensityAnalyzer class from the NLTK library. This class uses pre-trained models to assign sentiment scores to individual words to generate an overall sentiment score for a given text.

The algorithms and techniques used in this project are based on lexical-based sentiment analysis. A lexicon-based approach is based on a pre-defined sentiment lexicon or dictionary that associates sentiment scores with words. NLTK's SentimentIntensityAnalyzer uses a lexical-based approach to assign a sentiment score to each word in the text and combine these scores to get an overall sentiment score for the text.

This project used web scraping techniques to extract Vodafone reviews from the Trustpilot website. Python's BeautifulSoup and Requests libraries were used to extract relevant information from the HTML structure of the Trustpilot web page. The extracted ratings were saved in CSV files for further analysis.

By combining the power of NLTK's SentimentIntensityAnalyzer, web scraping and data manipulation techniques, this project will provide an up-to-date understanding of the sentiment expressed in Vodafone reviews from Trustpilot, providing a comprehensive view of customers' perceptions and sentiments towards the company. It was intended to allow analysis.

**Methodology**

In this project we’ve utilized a couple of modules to assist us in web scraping the reviews from the website and also to allow us to save the reviews in a csv file and to be able to manipulate this file for our preprocessing needs.

A screenshot of a computer

Description automatically generated with medium confidence

**1.Web Scraping:**

The request module was used to connect us to the target website and allows us to extract the page source html code subsequently beautiful soup was used to parse the html content after words the command (**soup.findall** ) was used to find all html content with the specific tag of (p and class=typography\_body-l\_\_KUYFJ typography\_appearance-default\_\_AAY17 typography\_color-black\_\_5LYEn

).Finally, the extracted reviews were saved in a csv file.

A screenshot of a computer

Description automatically generated with medium confidence

**2.Preprocessing:**

To be able to use the reviews in the csv file in a sentiment Analyzer it is first required to clean the data.

The extracted reviews from the CSV file were preprocessed to ensure better accuracy in sentiment analysis. The preprocessing steps included: Removing any irrelevant information such as HTML tags, special characters, and punctuation marks. Converting the text to lowercase to maintain consistency. Removing stop words (commonly used words such as "and", "the", "is") that do not contribute much to sentiment analysis. Tokenizing the text by splitting it into individual words.

A picture containing text, font, line, number

Description automatically generated

**3.Sentiment Analysis:**

The NLTK (Natural Language Toolkit) library was used for sentiment analysis. Specifically, the SentimentIntensityAnalyzer class from the NLTK library was employed. The SentimentIntensityAnalyzer uses a pre-trained model that assigns sentiment scores to individual words in a text and combines them to generate an overall sentiment score for the text. For each review in the preprocessed data, the SentimentIntensityAnalyzer was applied to obtain sentiment scores. The sentiment scores consisted of positive, negative, neutral, and compound scores.

**Implementation and Testing:**

The implementation involved writing code in Python using libraries such as NLTK, BeautifulSoup, and pandas for data manipulation and analysis. The sentiment analysis was performed on the preprocessed data, and the sentiment scores were assigned to each review. To evaluate the accuracy of the sentiment analysis. The compound scores, which range from -1 (extremely negative) to 1 (extremely positive), were used to categorize the sentiment as positive, negative, or neutral. The results were analyzed by calculating the percentage of positive, negative, and neutral reviews from the dataset. Additional analysis such as word frequency analysis or word cloud generation could also be performed to gain further insights into the sentiment expressed in the reviews.

**Majority Sentiment Analysis:**

After obtaining the sentiment scores for each review, the next step was to determine the overall sentiment of the Vodafone reviews. To achieve this, the sentiment scores were categorized into three classes: positive, negative, and neutral. The compound score was used as the basis for classification. A threshold value was set to distinguish positive and negative sentiments. Reviews with a compound score greater than the threshold were classified as positive, those with a compound score lower than the negative threshold were classified as negative, and the remaining reviews were labeled as neutral. By calculating the percentage of positive, negative, and neutral reviews, the majority sentiment of the Vodafone reviews was determined. In this project, the majority sentiment was found to be positive, indicating that the majority of the reviews analyzed had a positive sentiment towards Vodafone

****

**Conclusion:**

Based on the sentiment analysis results, conclusions can be drawn about the overall sentiment of the Vodafone reviews extracted from Trustpilot. The methodology used in this project provides an automated approach to sentiment analysis using the NLTK library, allowing for efficient analysis of large volumes of text data. The accuracy of the sentiment analysis can be further improved by incorporating domain-specific lexicons or by training a custom sentiment analysis model on domain-specific data.