**Project Title:**

**Use of social media for consumer opinion mining**

Student name

University

Course

Date

**Introduction**

Online communication has rapidly increased in popularity in recent years. There are a lot of social networking websites and mobile applications related to them, and more are constantly being developed. These websites produce enormous amounts of data every day, and you can utilize that data as a basis for a variety of analyses. With millions of members, Twitter is one of the most well-known networking services. There are users with various opinions, and they produce a range of reviews in the form of tweets. I focused on offering an opinion on the specific product in this article utilizing data from Twitter. There are millions of reviews for a single product, making it impossible for a customer or company to read them all and assess the product's quality. This implementation paper offers opinion mining on a specific product using customer reviews. The process comprises determining the positivity and negativity of tweets as well as providing an overall percentage of tweets that are favorable, negative, or neutral. This project's fundamental concept is that the consumer should automatically receive product suggestions based on prior tweets.

Social media involves a collection of online forums where users may share and discuss their interests, ideas, and activities with people in their social networks or the general public (Andrlic & De Alwis, 2016). By gathering words and feelings expressed on social media platforms including blogs, forums, and social networking sites like Facebook and Twitter, social media research aims to understand customer opinions. Although customer views aid marketers in comprehending their needs, desires, and preferences, there is a challenge that many marketers want to solve: It can be challenging to predict what people will want before they express it. Before expressing their opinions publicly, researchers might utilize social media to obtain information on consumer sentiments. The survey approach does not**,** however, explicitly ask each responder a question.

**Consumer opinion mining** involves drawing conclusions from publicly available material and applying them to market research (Tsirakis et al., 2017). There are several methods to do this, from using simple software to scrapbook articles on a blog to being created with real-time human interactions. Posting the results of your study on social media is the easiest method. It is simple to preserve and analyze conversations. You may get a ton of data without performing any actual work using social media platforms like Twitter, Facebook, and others, at least in the early phases of your project.

**Related works**

Instead than focusing on a specific consumer demographic, traditional market research techniques, including interviews, surveys, and observations often evaluated consumer attitudes from sample groups. However, since the advent of the internet, customers and businesses now frequently communicate online, which has altered marketplaces and organizations (Lusch, 2010), (Chen, 2010) . In this regard, consumer WOM, which refers to interpersonal communication, has been acknowledged as a vital source that helps businesses to comprehend consumers' interests, attitudes, and views about goods and services. In contrast, past research on WOM analysis only took customer evaluations and popularity ratings into account.

Social media and conventional market research techniques.

Traditional approaches are expensive, but social media is a cost-effective tool to learn what consumers think about goods and services (Rathore et al., 2016). Traditional approaches rely on surveys and focus groups, which may be costly and time-consuming for the researcher. Social media is available to market researchers and does not call for in-person participant engagement, a protracted, difficult survey, or a sizable sample. You can draw generalizations about the population from data collected from a much smaller set of customers.

Traditional techniques are sometimes accused for being skewed toward what is popular, while social media is supposed to be open and democratic by enabling participation for anybody with an internet connection, as opposed to only those with the resources for research and the time to do it. The participants are always honest about their ideas and do not feel pressured to provide a comprehensive answer, therefore the data gathered on social media may be more accurate (Dusek et al., 2015). On social media, tweets, comments, and posts frequently have a spontaneous aspect, but in traditional research focus groups and surveys, participants could be less honest about their opinions.

**Method and Implementation Technique view**

**Interesting Blog posts**. Can be useful since you can learn about the users and they include a lot of info. You may find out information on how people use particular technologies and goods, what they like and dislike, what interests them, what they believe to be important, etc. However, since blogging is less organized than social media, it takes more work to gather data from interesting blog posts, so you need to make sure you can act on your discoveries.

**Forums.** Are teams of individuals who have a same passion or interest, such as those for computers or vehicles. Forums function as a sort of "virtual gathering place" where individuals may converse about things they find interesting and get to know one another. It could be difficult to strike up talks and bring up subjects about which you'd like further information, though, if you don't have a particular interest in the forum.

**Blog comments**. This approach is comparable to blogging. If you want to remark on a blog article that has grabbed your interest and you want to react to it, you may use the site's commenting mechanism to do so instead of producing your own material. The blogger is not doing this using the blogging software. Using Google Reader and following blogs is one method to achieve this.

**Dataset Collection**

Twitter API provided by twitter can be used to access, retrieve and manipulate number of tweets posted on Twitter. Before getting to access, authorization is required, which means that one has to sign up for a developer account here [(twitter account)](developer.twitter.com). Thereafter one is provided with necessary credentials that can be used to access the tweets. Tweepy is a python library that simplifies the task of accessing tweets with the necessary credentials.

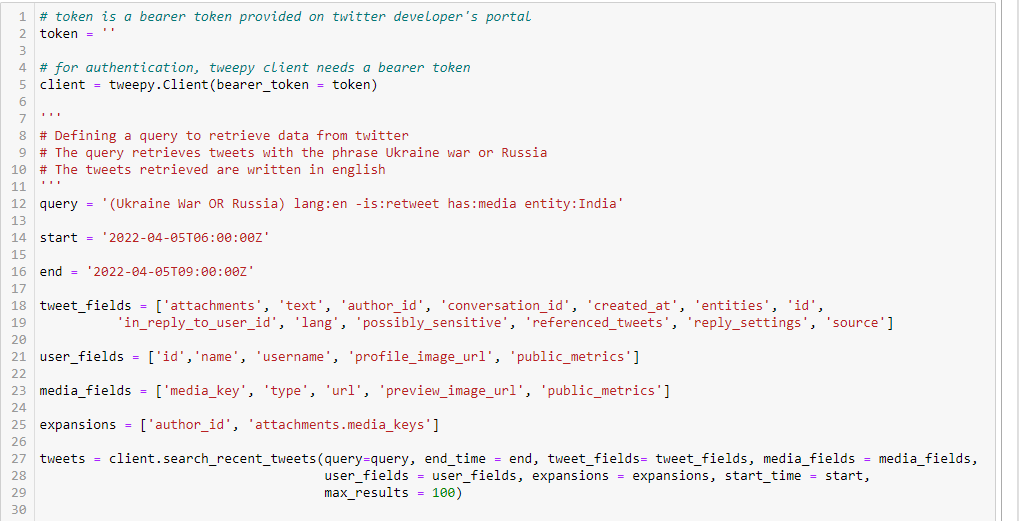
**Techniques for gathering data from social media**

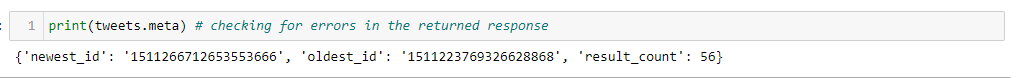
**Scrapbooking-** Given that it doesn't involve any actual work, this is the easiest type of social media. To save the content you are interested in, all you have to do is take it. Subsequently, publish it on your blog or even just your Facebook wall. Direct content scraping from Facebook pages or Twitter feeds is another option.

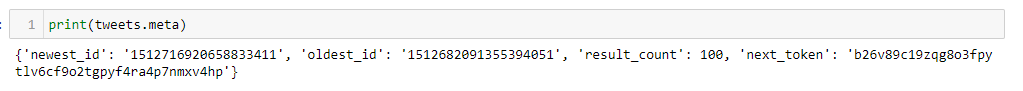
**Semiautomatic extraction**. Utilizing tools like SocialMiner, it is possible to automate the process of extracting material from social networking networks. With the help of this program, we may scrape material from a variety of social networking networks and save it as RSS feeds or keep it in a database for later use. This approach is more complicated and time-consuming, but it typically produces better results than just scraping content directly from the web.

**Crawling** - Utilizing social media tools like StatCrunch, IpsoFacto, and Infographic is necessary for this. Direct data collection from social media websites like Twitter, Facebook, LinkedIn, YouTube, blogs, and websites with RSS feeds that include text, photographs, and other multimedia information may be automated.

**Human data collection** - This type of social media data collecting is the most difficult since it requires a native speaker of the target language who can comprehend the content posted on the website, which is often written in English.

This project compiles tweets related to the conflict in Ukraine. Tweepy demands this as a query to be supplied, indicating every other item you need to have with a query result in order to make collecting easier. For instance, tweets from India, Hindu, or anybody else connected to India were collected in this instance. The query and requests are made as seen in the following screenshots.

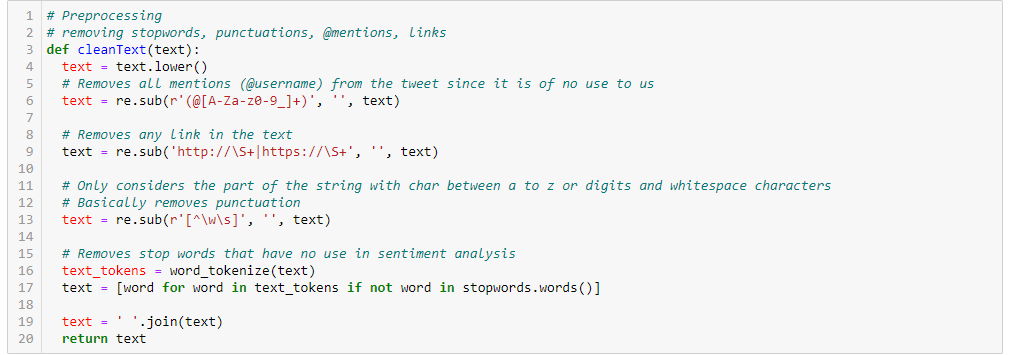
The answer contains information on the retrieved tweets, including their total number.

The number of tweets that were obtained may be seen using the Meta object, as seen below.****

The graphic depicts the response's 100 tweets.

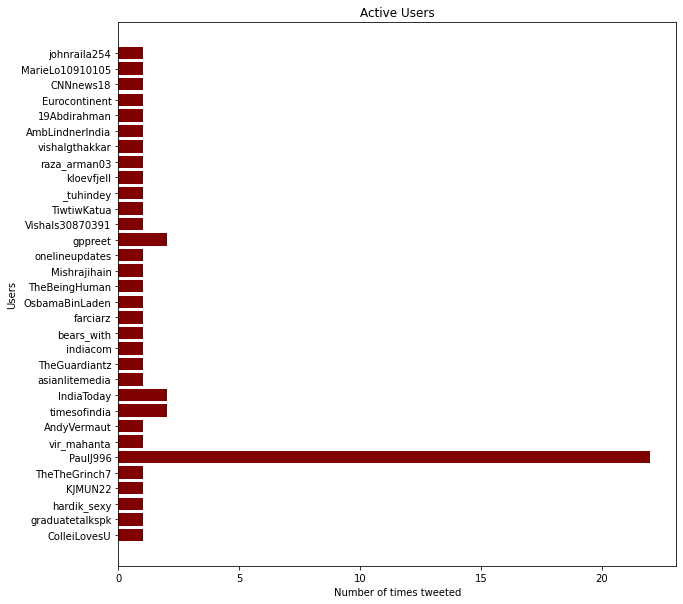
**Analysis**

**Preprocessing feature**

Preprocessing entails transforming data into a format that is understandable. Social media data is unstructured and includes text, photographs, videos, emoticons, and a combination of various types. Some characteristics are superfluous for text data retrieved from Twitter. An analysis would not be necessary if a tweet had mentions, stop words, or punctuation, for instance. Also unnecessary are links that are included in a tweet. The following function can be used for preprocessing.****

**Feature classification**

**Most active users**

The feature selection is described with users and number of times the tweets got tweeted. First, users are independent variable while the number of tweets feature is a dependent variable. Below are users who have the most tweets relevant to the search parameters. Using a Python script and the obtained data, the graph is created. Information on tweeting users is provided in Twitter's answer, which may be utilized as needed.****

Python may classify the most popular hashtags in the retrieved tweets as follows:****

The word cloud generated by the code above is shown below. The most popular hashtags in the dataset are visible to us.

In this figure below, I give an example of a word cloud or tag cloud that I created by extracting trending topics and high-frequency keyword phrases from the desired time frame. The size of a word in the word cloud indicates the volume of the topic, and the color highlights the topic.

This kind of word cloud can provide more information.****

**Case study: Celebrity Profile**

Conducting a review of celebrity’s social media presence, and identifying keywords, hashtags or phrases that have been associated with their accounts.

**Preliminary review and findings**

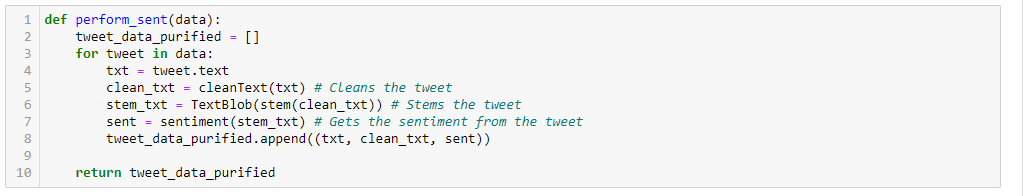
Following the incident between Will Smith and Chris Rock at the Oscars in 2022, in which Will Smith seemed to hit Rock, since then, this has stirred the internet, prompting users to create a large number of hashtags to spread the news. The three most used hashtags are #Oscars2022, #Willsmith, #Willsmithassault, and #Chrisrock.

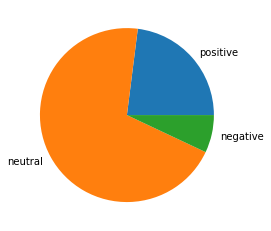
**Sentiment Modelling and Analysis**

Sentiment modelling is a method of dividing text into positive, negative, and neutral categories. Opinion mining is another name for sentiment modeling. It is a branch of machine learning called natural language processing that searches for and extracts insights from text. Textual data is frequently subjected to sentiment analysis to assist businesses in tracking the tone of consumer reviews of brands and products and potentially comprehending client wants.

Data preparation is the first step you must do before starting sentiment analysis. Eliminating stop words and punctuation from texts is part of this stage. For categorization, these characteristics are not required. Links that are present in the text do not aid sentiment analysis either. Also, all lowercase letters are used in the text.

The second critical stage for a successful sentiment model. The process involves breaking down a given word into its stem, which is then attached to its suffixes, prefixes, or root.

The classification of text as positive or negative is the final phase. As seen below, this was accomplished using Python code and libraries.****

I did give a brief statistical overview of the displayed opinion-mining outputs for the data obtained. The outcome of sentiment analysis, which includes a ratio of polarity (positive, neutral, and negative), is the output. It has been categorized using the table below:****

**Sentiment mining Issues**

There are different types of issues with sentiment analysis, but sometimes the biggest issue arises when an opinion that is beneficial to a user group turns out to be detrimental to a manufacturer or vice versa.

Bing Liu elaborates on these issues in his book "Sentiment Analysis and Opinion Mining."

a. It is difficult to determine by a keyword's meaning whether a sentiment is positive or negative because they may have the opposite meaning in a certain area.

b. Questioning Sentence

Neither a positive nor a negative attitude may be present in an interrogative phrase, although the opinion expressed by the key word may be either one.

c. Lighthearted Phrases

A few sentences that are written in the jock style might make the entire statement meaningless. Such a statement requires the reader to pay close attention to the keywords and sentences. These absurd phrases not only undermine the meaning of the sentence in question but also diminish the overall quality of the essay.

**References**

Andrlic, B., & De Alwis, A. C. (2016). Social media in destination marketing.

Dusek, G., Yurova, Y., & Ruppel, C. P. (2015). Using social media and targeted snowball sampling to survey a hard-to-reach population: A case study. *International Journal of doctoral studies*, *10*, 279.

McDaniel Jr, C., & Gates, R. (2018). *Marketing research*. John Wiley & Sons.

Rathore, A. K., Ilavarasan, P. V., & Dwivedi, Y. K. (2016). Social media content and product co-creation: an emerging paradigm. *Journal of Enterprise Information Management*.

Tsirakis, N., Poulopoulos, V., Tsantilas, P., & Varlamis, I. (2017). Large scale opinion mining for social, news and blog data. *Journal of Systems and Software*, *127*, 237-248.

Lusch, R., Liu, Y., & Chen, Y. (2010). The phase transition of markets and organizations: the new intelligence and entrepreneurial frontier. *IEEE Intelligent Systems, IEEE Intelligent Systems*, *25*(1), 71-75.

Chen, H. (2010). Business and market intelligence 2.0. *IEEE intelligent systems*, *25*(1), 68-71.