# ABSTRACT

Twitter is a platform widely used by people to express their opinions and display sentiments on different occasions. Sentiment analysis is an approach to analyze data and retrieve sentiment that it embodies. Twitter sentiment analysis is an application of sentiment analysis on data from Twitter (tweets), in order to extract sentiments conveyed by the user. In the past decades, the research in this field has consistently grown. The reason behind this is the challenging format of the tweets which makes the processing difficult. The tweet format is very small which generates a whole new dimension of problems like use of slang, abbreviations etc. In this paper, we aim to review some papers regarding research in sentiment analysis on Twitter, describing the methodologies adopted and models applied, along with describing a generalized Python based approach.

With the advancement of web technology and its growth, there is a huge volume of data present in the web for internet users and a lot of data is generated too. Internet has become a platform for online learning, exchanging ideas and sharing opinions. Social networking sites like Twitter, Facebook, Google+ are rapidly gaining popularity as they allow people to share and express their views about topics, have discussion with different communities, or post messages across the world. There has been lot of work in the field of sentiment analysis of twitter data. This survey focuses mainly on sentiment analysis of twitter data which is helpful to analyze the information in the tweets where opinions are highly unstructured, heterogeneous and are either positive or negative, or neutral in some cases. In this paper, we provide a survey and a comparative analyses of existing techniques for opinion mining like machine learning and lexicon-based approaches, together with evaluation metrics. Using various machine learning algorithms like Naive Bayes, Max Entropy, and Support Vector Machine, we provide research on twitter data streams. We have also discussed general challenges and applications of Sentiment Analysis on Twitter.

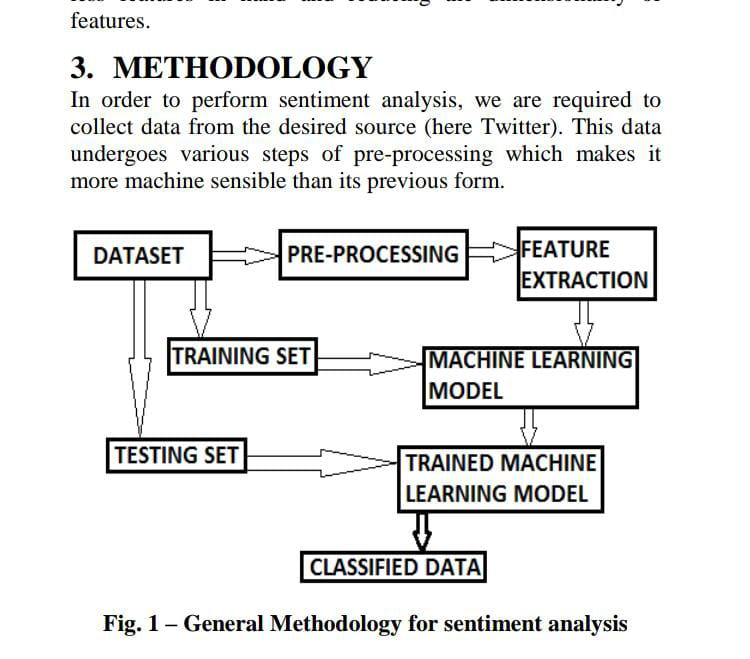
# Introduction

Twitter has emerged as a major micro-blogging website, having over 100 million users generating over 500 million tweets every day. With such large audience, Twitter has consistently attracted users to convey their opinions and perspective about any issue, brand, company or any other topic of interest. Due to this reason, Twitter is used as an informative source by many organizations, institutions and companies.

On Twitter, users are allowed to share their opinions in the form of tweets, using only 140 characters. This leads to people compacting their statements by using slang, abbreviations, emoticons, short forms etc. Along with this, people convey their opinions by using sarcasm and polysemy.

Hence it is justified to term the Twitter language as unstructured. In order to extract sentiment from tweets, sentiment analysis is used. The results from this can be used in many areas like analyzing and monitoring changes of sentiment with an event, sentiments regarding a particular brand or release of a particular product, analyzing public view of government policies etc. A lot of research has been done on Twitter data in order to classify the tweets and analyze the results. Sentiment analysis is a process of deriving sentiment of a particular statement or sentence. It’s a classification technique which derives opinion from the tweets and formulates a sentiment and on the basis of which, sentiment classification is performed. Sentiments are subjective to the topic of interest. We are required to formulate that what kind of features will decide for the sentiment it embodie

**Data Flow Diagram**

****

**PROBLEM STATEMENT**

### Sentiment analysis is ever-growing sub-field of natural language processing. There are many research works that have addressed the problems of SA from text, emoticon, images, and audio or videos separately [1], [6]. Very few researches have been done on emoticon for finding sentiments [3]. Moreover, related works section shows there is a scope for further extension in the field of SA with both text and emoticon. Therefore, the objectives of this research are:

### Sentiment analysis using bi-mode (text and emoticons) on social media data.

### Developing emoticon lexicon

### To improve the classification accuracy of sentiment analysis by using both ML and DL algorithm

**Pre-processing of tweets**

The preprocessing of the data is a very important step as it decides the efficiency of the other steps down in line. It involves syntactical correction of the tweets as desired. The steps involved should aim for making the data more machine readable in order to reduce ambiguity in feature extraction.

Below are a few steps used for pre-processing of tweets -

 Removal of re-tweets.

 Converting upper case to lower case: In case we are using case sensitive analysis, we might take two occurrence of same words as different due to their sentence case. It important for an effective analysis not to provide such misgivings to the model.

 Stop word removal: Stop words that don’t affect the meaning of the tweet are removed (for example and, or, still etc.). [3] uses WEKA machine learning package for this purpose, which checks each word from the text against a dictionary ([3], [5]).

 Twitter feature removal: User names and URLs are not important from the perspective of future processing, hence their presence is futile. All usernames and URLs are converted to generic tags [3] or removed [5].Stemming: Replacing words with their roots, reducing different types of words with similar meanings [3]. This helps in reducing the dimensionality of the feature set.

 Special character and digit removal: Digits and special characters don’t convey any sentiment. Sometimes they are mixed with words, hence their removal can help in associating two words that were otherwise considered different.

 Creating a dictionary to remove unwanted words and punctuation marks from the text [5].

 Expansion of slangs and abbreviations [5].

 Spelling correction [5].

 Generating a dictionary for words that are important [7] or for emoticons [2].

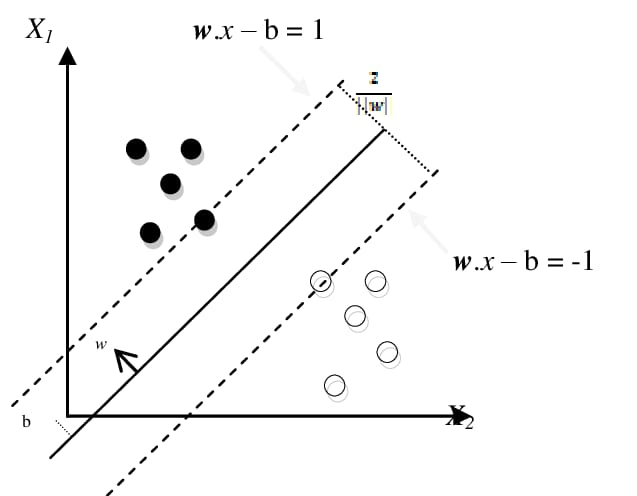
 Part of speech (POS) tagging: It assigns tag to each word in text and classifies a word to a specific category like noun, verb, adjective etc. POS taggers are efficient for explicit feature extraction.

**Sentiment Analysis using SVM:**

Sentiment Analysis is the NLP technique that performs on the text to determine whether the author’s intentions towards a particular topic, product, etc. are positive, negative, or neutral.

What is SVM?

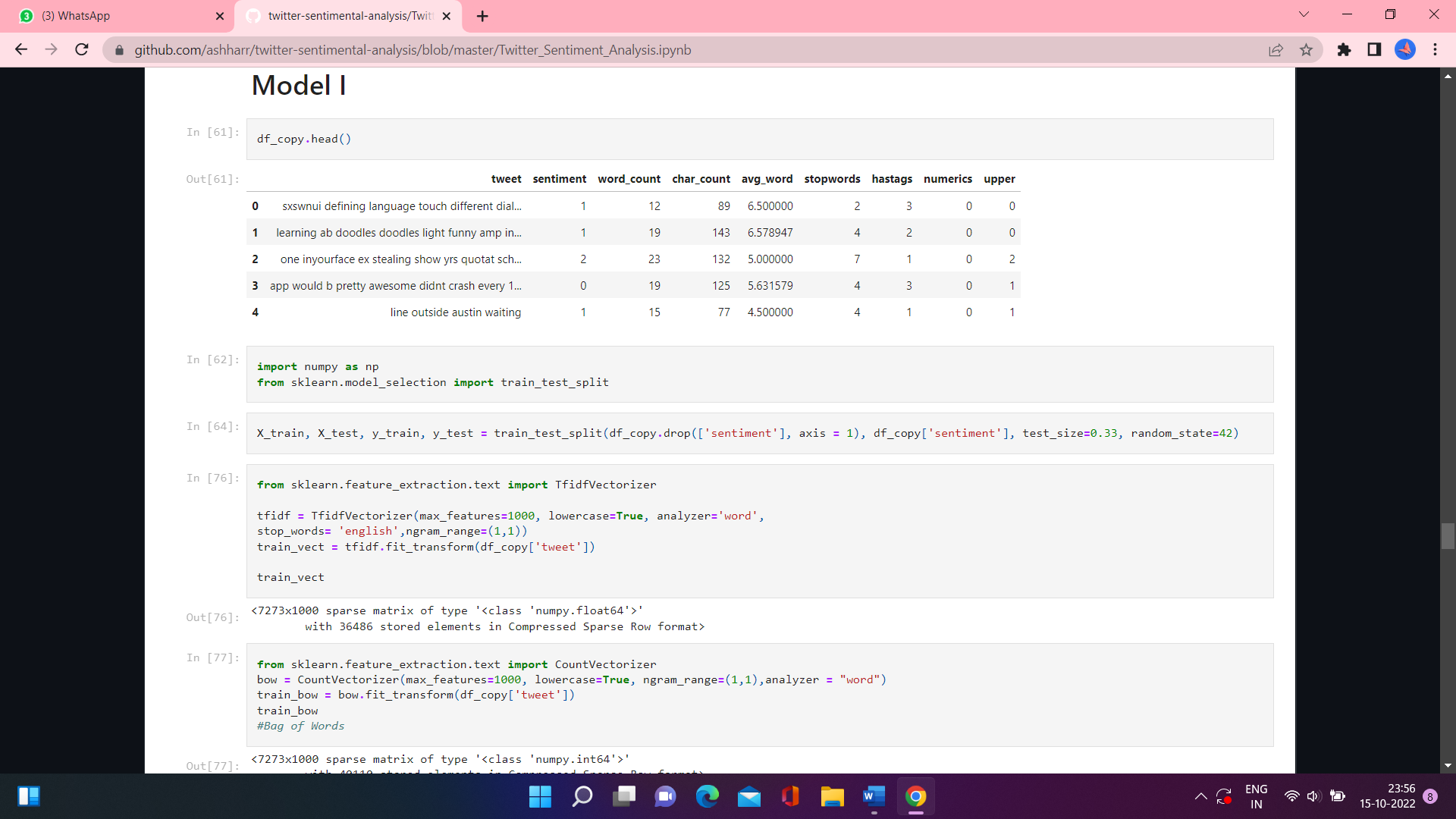
SVM is a supervised(feed-me) machine learning algorithm that can be used for both classification or regression challenges. Classification is predicting a label/group and Regression is predicting a continuous value. SVM performs classification by finding the hyper-plane that differentiate the classes we plotted in n-dimensional space.

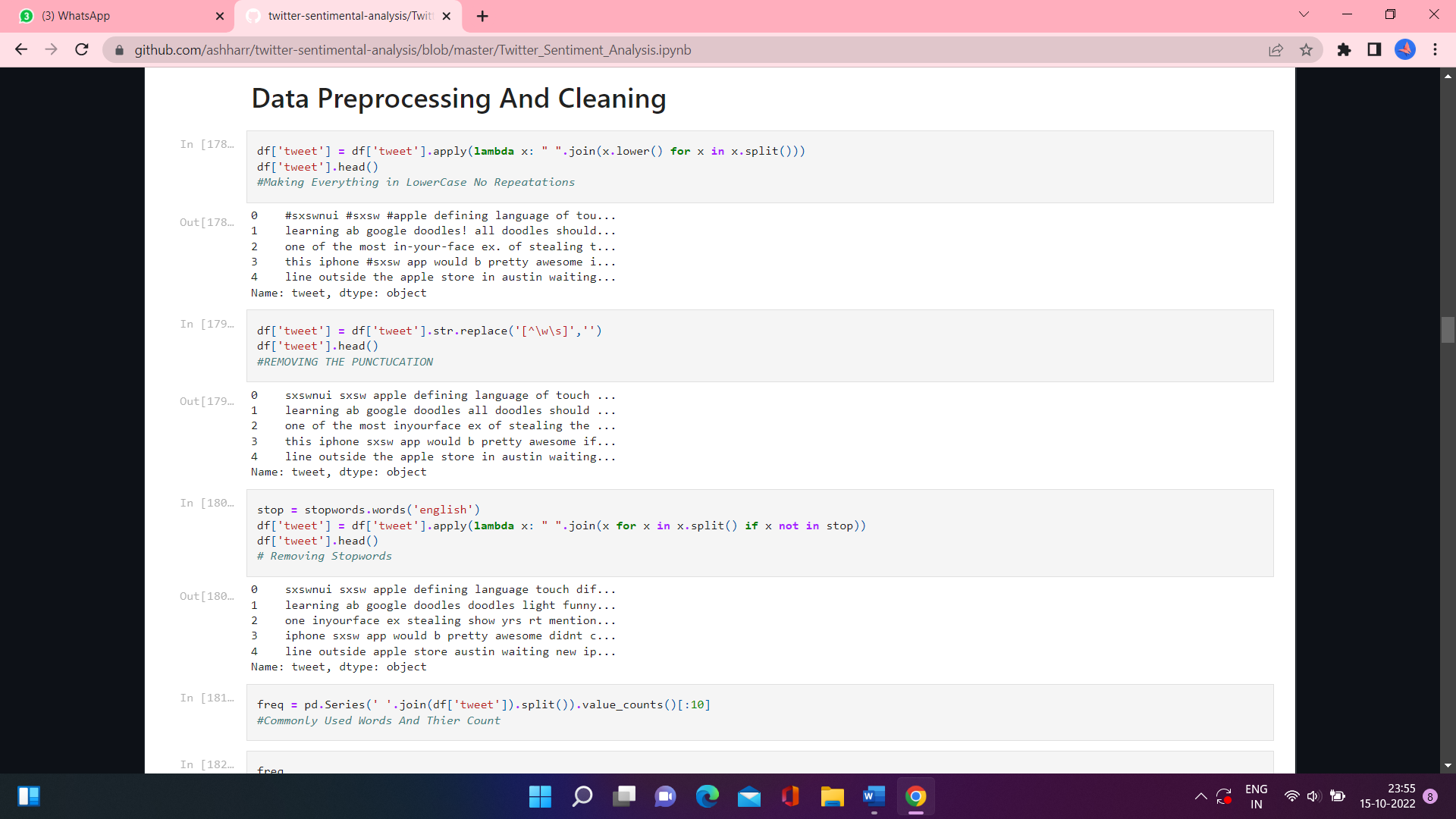
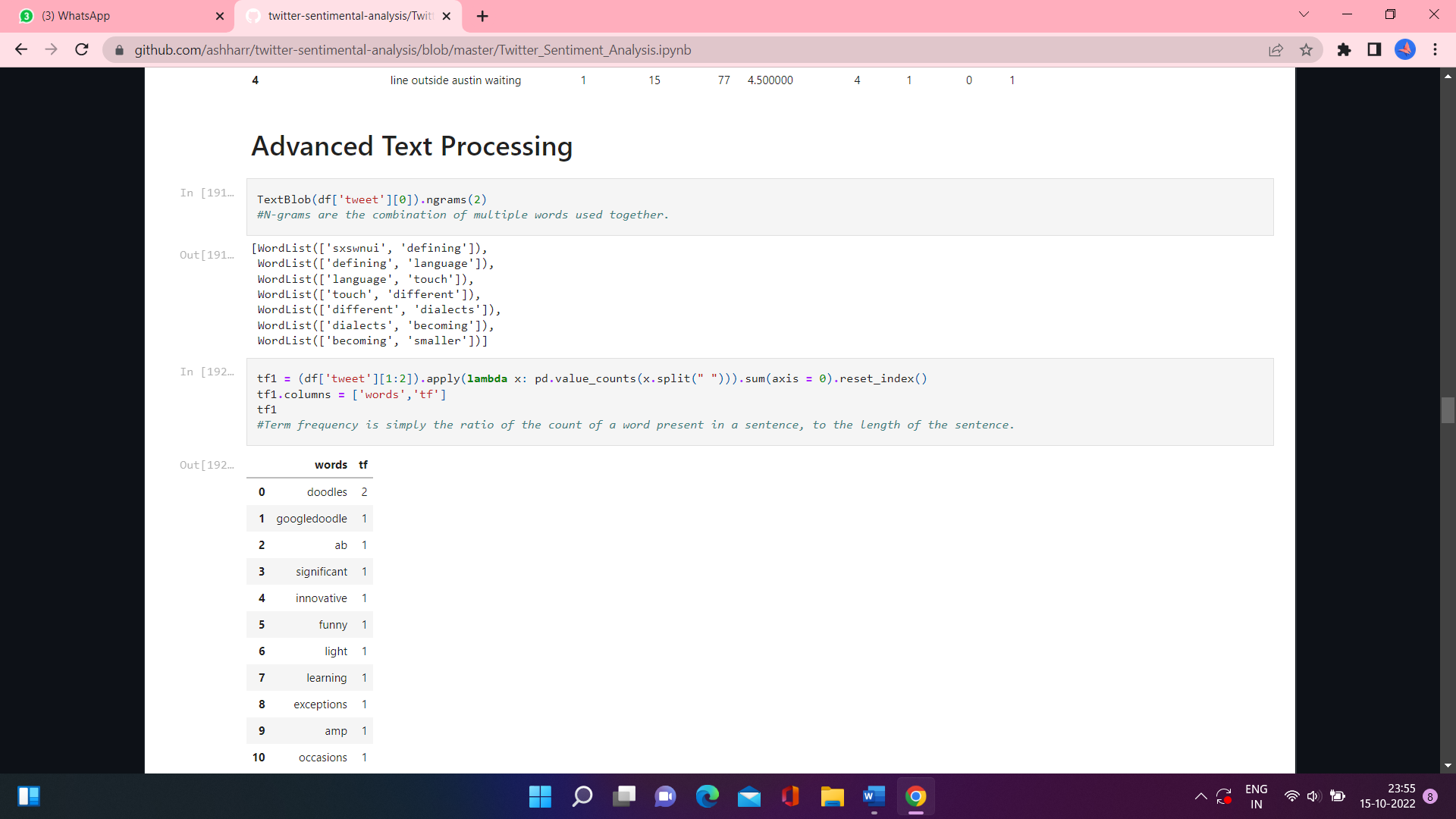
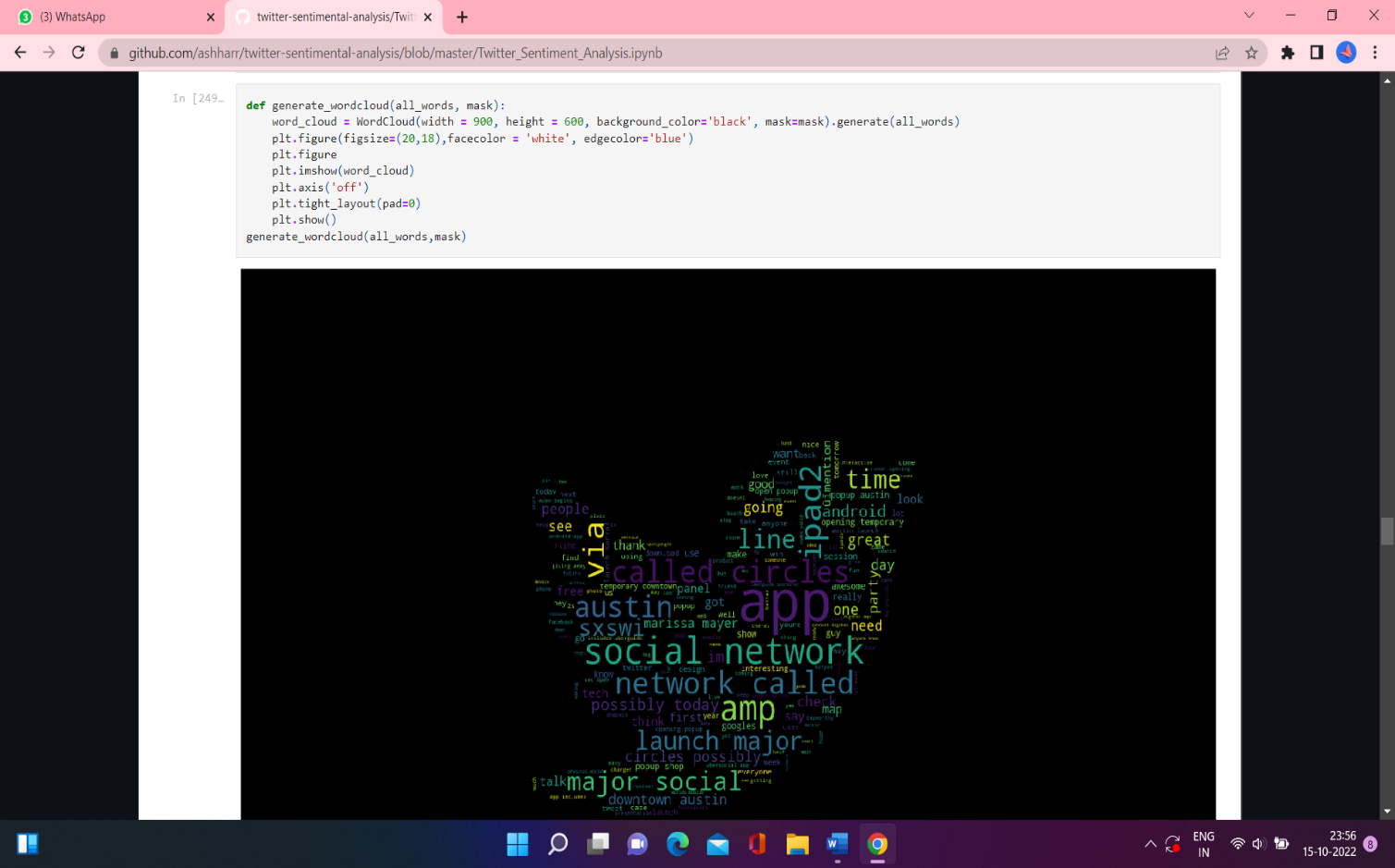
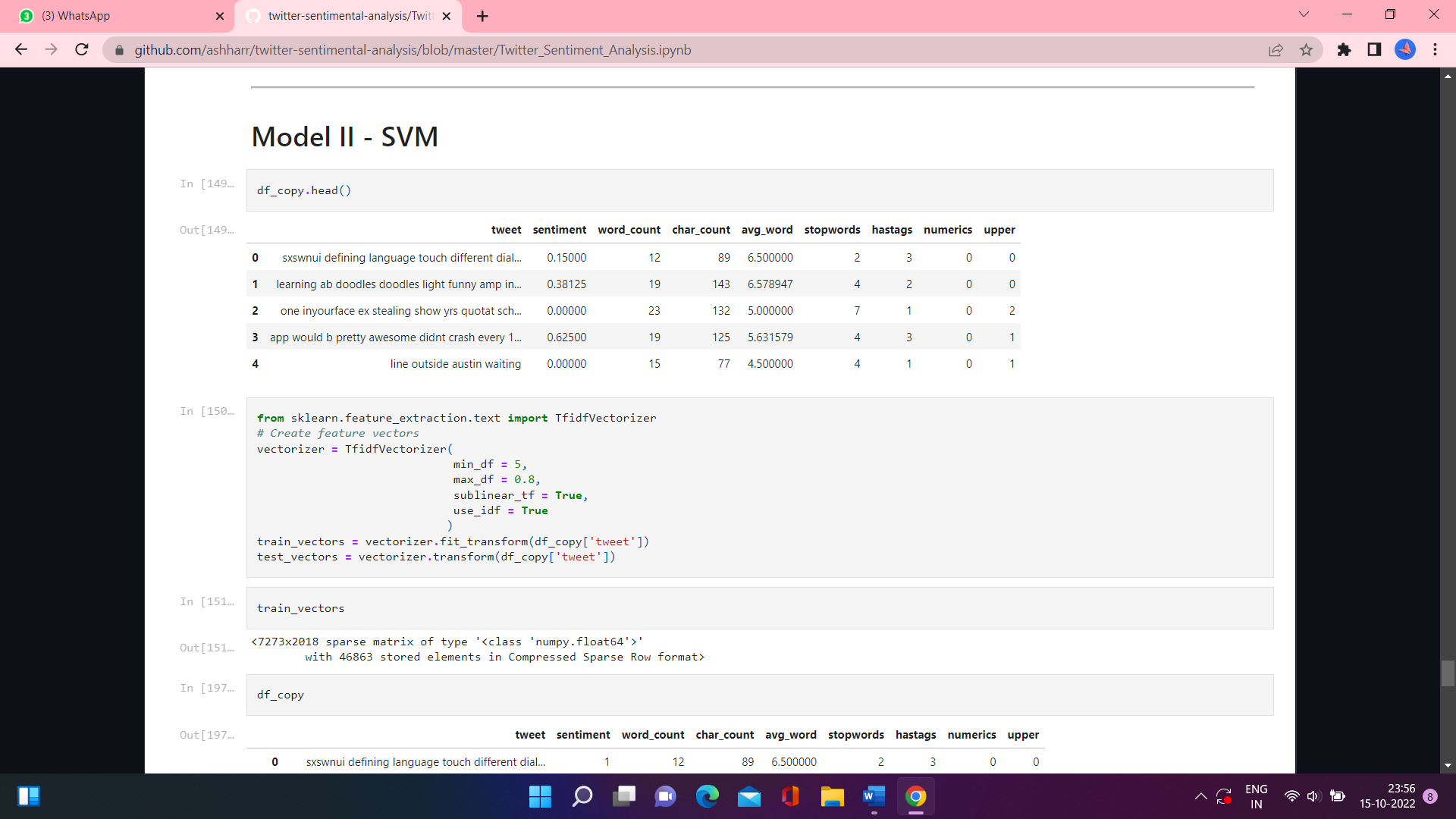
****

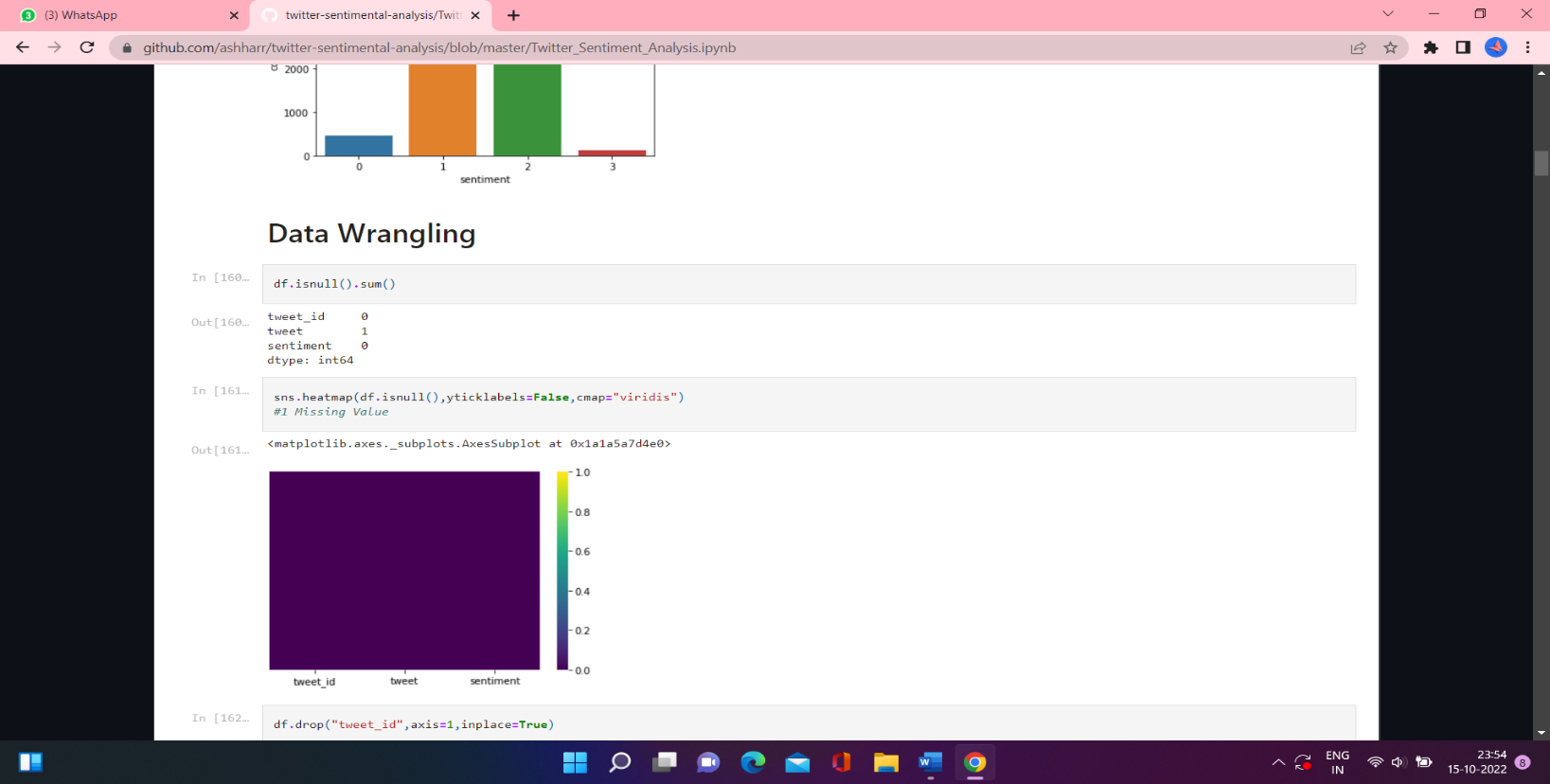
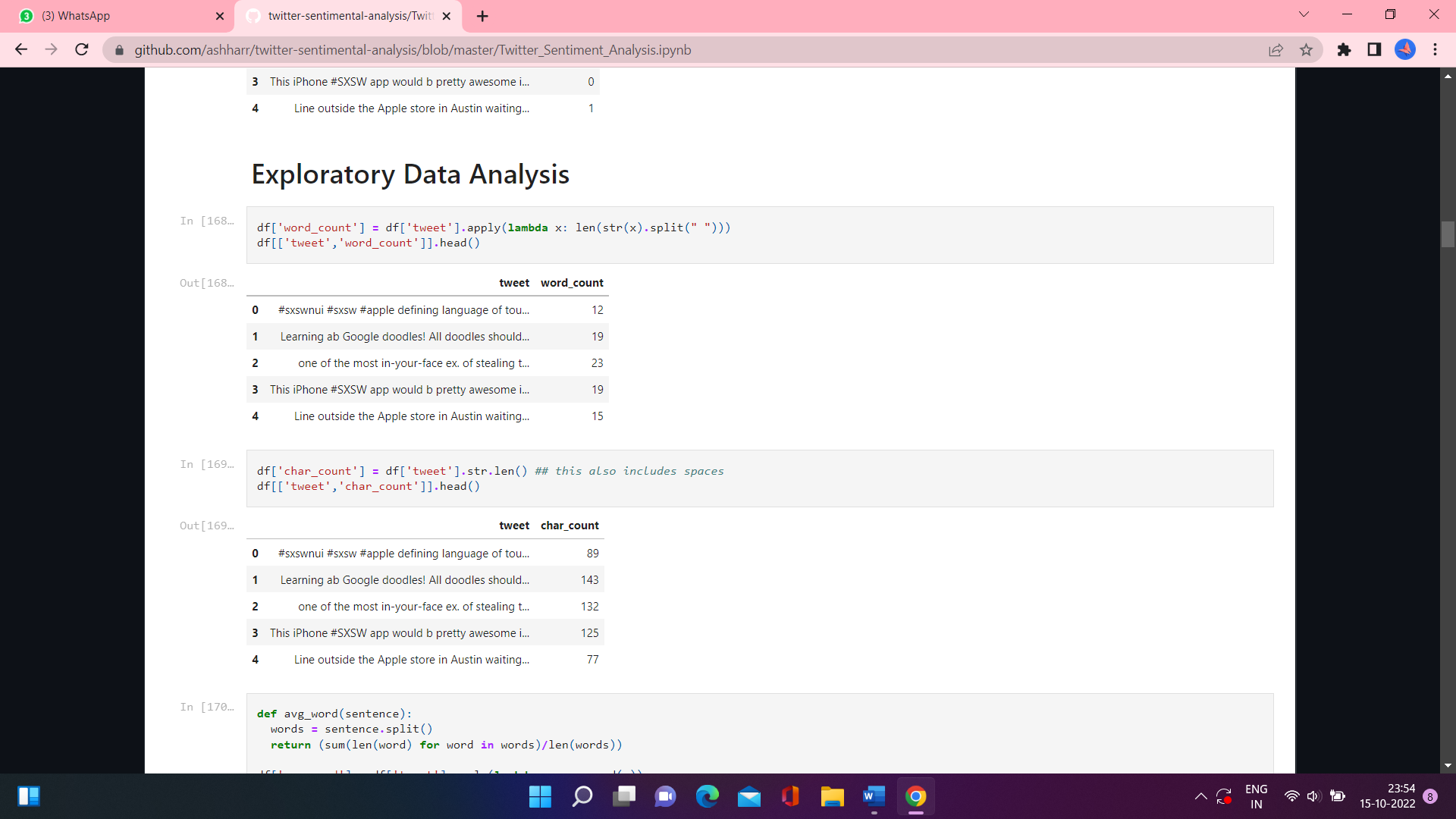
**IMPLEMENTATION**

The project is run in jupyter by using different types of libraries such as numpy, pandas,seaborn, matplotlib, scipy, sklearn, textblob, etc. After we took 2 datasets i.e train.csv and test.csv After we started the preprocessing of the dataset and clean the data which we required to project. After analysing the dataset we get the different types of result about the Twitter Sentiment Analysis.

**OUTPUT SCREENSHOTS :**







**CONCLUSION**

Sentiment Analysis is considered as one of the hot research topics in the domain of knowledge discovery. Large amount of online data is being added on daily basis ranging from social media posts and comments to movie and software reviews. By using sentiment analysis techniques, these data sources can be used to fetch the useful information such as: prediction of election results, getting user’s feedback about any software,

analyzing the market reputation of particular brand and obtaining public opinion before launching a new product etc.

Multiple approaches are available for sentiment analysis such as lexicon based, machine learning based and the hybrid of both. SVM is one of the widely used machine learning techniques for the detection of polarity from text. Now days, along with conventional machine learning classification techniques, many customized and integrated models have been proposed by researchers for sentiment analysis and polarity detection. This study has provided a compact and comprehensive review of latest research by focusing on SVM technique of sentiment analysis. This study has followed a systematic framework for review and provided the answers of identified research questions after critical review of selected papers. For future work it is suggested to perform a comparative analysis of the customized techniques with same dataset.

**REFERENCE**

* <https://www.google.com/url?sa=t&source=web&rct=j&url=https://scholar.google.co.in/scholar_url%3Furl%3Dhttps://www.academia.edu/download/55982245/Paper_26_Sentiment_Analysis_using_SVM.pdf%26hl%3Den%26sa%3DX%26ei%3DUfxKY46SB8amywSgjJiQDA%26scisig%3DAAGBfm3pJuH6BmrMFqueBfffWN5MpW9taA%26oi%3Dscholarr&ved=2ahUKEwiM3aSE7>L6AhWcwjgGHe3eB40QgAN6BAgIEAE&usg=AOvVaw2cil7jTD-NBYBOlqkZtOcg
* <https://www.researchgate.net/publication/286583940_Sentiment_analysis_using_Support_Vector_Machine>
* <https://www.slideshare.net/SavioAberneithie/twitter-sentimentanalysis-report>
* <https://www.researchgate.net/publication/301408174_Twitter_sentiment_analysis>