



# NATURAL LANGUAGE PROCESSING ANALYSIS WEBSITE USING PYTHON/DJANGO

**K.Praveen<sup>1</sup>, J.J.Nanthakumar<sup>2</sup>, N.Nirmal Kumar<sup>3</sup>**

Student, CSE, Panimalar Engineering College, Chennai, India<sup>1,2,3</sup>

**Abstract:** Reviews act as a valuable source of information for decision making. Online e-commerce sites has provided their users to make their opinion about products and services. Huge amount of such opinions are publicly available in the form of reviews. Manufacturers, retailer as well as customers have great interest in customer reviews. Due to large number of reviews available on internet for analysis, it is not cost worthy to read these manually. To optimize this time consuming task there is a need of an automated system which provides summarized result of user sentiments. Opinion Mining (OM) is the field of study that analyzes people's sentiments or opinion from reviews or opinionated text. Opinion Mining can be viewed as a natural language processing task, the task is to develop a system that understands the people's language. Opinion Mining is a difficult task due to ambiguous nature of human languages( like English).

**Keywords:** Django, API, tweepy, NLP, heroku, textblob

## I. INTRODUCTION

Natural Language Processing or NLP is a field of Artificial Intelligence that gives the machines the ability to read, understand and derive meaning from human languages. It is a discipline that focuses on the interaction between data science and human language, and is scaling to lots of industries. Today NLP is booming thanks to the huge improvements in the access to data and the increase in computational power, which are allowing practitioners to achieve meaningful results in areas like healthcare, media, finance and human resources, among others. In simple terms, NLP represents the automatic handling of natural human language like speech or text, and although the concept itself is fascinating, the real value behind this technology comes from the use cases.

## II. EXISTING SYSTEM

Twitter, with 500 million users and million messages per day, has quickly become a valuable asset for organizations to invigilate their reputation and brands by extracting and analyzing the sentiment of the tweets by the public about their products, services market, and even about competitors highlighted that, from the social media generated opinions with the mammoth growth of the world wide web, super volumes of opinion texts in the form of tweets, reviews, blogs or any discussion groups and forums are available for data analysis, thus making the world wide web the very fast, most comprising and easily accessible medium for Opinion Mining.

## III. PROPOSED SYSTEM

Imagine you just launched a new product feature and notice a sharp increase in mentions on Twitter. Are customers tweeting more because they are delighted with the new feature? Or, are they actually complaining about the feature?

### Problem definition

- Going through each of these comments manually would take far too much time. You did miss out on valuable feedback that could help you instantly improve a customer's experience with the latest feature (bug issues, user experience).
- By performing analysis with NLP, you can quickly understand the tone and context of social mentions on Twitter.

### Advantages:

- Opinion Mining
- Advanced Visualization for tweets



- Authentication via Different Social Media Platforms
- Exporting Tweets to Excel Sheet

### Modules

- Basic Authentication
- Object Data collection
- Data Preprocessing
- Data Visualization

### Module Description:

#### Basic Authentication:

The Django authentication system handles both authentication and authorization. Briefly, authentication verifies a user is who they claim to be, and authorization determines what an authenticated user is allowed to do. Here the term authentication is used to refer to both tasks.

The authentication system consists of:

- ✓ Users
- ✓ Permissions: Binary (yes/no) flags designating whether a user may perform a certain task.
- ✓ Groups: A generic way of applying labels and permissions to more than one user.
- ✓ A configurable password hashing system
- ✓ Forms and view tools for logging in users, or restricting content
- ✓ A pluggable backend system

So the user will login/register to our website in a secure manner, we have provided privacy policy to the users that how we use the data.

#### Object Data collection:

Real-time data collected from Twitter. The collection of data is one of the major and most important tasks of any machine learning project. Because the input we feed to the algorithms is data. So, the algorithm's efficiency and accuracy depend upon the correctness and quality of data collected. So the data should be clear as possible for better results. Depending on the hashtag entered by the user we will get the most popular and recent tweets.

#### Data Preprocessing:

Collection of data is one task and making that data as meaningful is another vital task. Data collected from various sources like Twitter, Facebook, etc. means those data will be in an unorganized format and there may be a lot of null values, invalid data values, and unwanted data. Cleaning all these data and replacing them with appropriate or approximate data and removing null and missing data and replacing them with some fixed alternate values are the basic steps in pre-processing of data. Even data collected may contain completely garbage values. It may not be in the exact format or the way that is meant to be. All those cases should be verified and replaced with some other alternate values to make the data meaningful and useful for further processing. Data must be kept in an organized format. So here we have used some regex expressions to clean some of the text of the tweets.

#### Data Visualization:

Data visualization is the process of translating information into a visual context, such as a map or graph, to make data much easier for the human brain to understand those data and get some insights from it. The main goal of data visualization is to make it easier to identify patterns, trends, and outliers in large data sets. So when the user hits analyze button on our website it will process those tweets and send back those results to the front end and the user can visualize those data based on sentiments, retweet status counts, favorite counts, etc. even the user can view individual positive tweets, negative tweets and neutral tweets then he/she can export it to excel if needed.

#### Software Requirements

- Windows 7 and above
- Python 3.x
- Django 3.x
- Twitter API Credentials
- Other dependencies find [here](#)



### Hardware Requirements

- Hard Disk : 80GB and Above
- RAM : 4GB and Above
- Processor : Intel Core i3 and Above

### Technology Used

- JavaScript, HTML5, CSS3, AJAX, JQuery.
- Django Framework
- MVC Pattern
- Tweepy, Chart.js

### CONCLUSION:

Thus the main aim is to help business people with their product analysis based on the hashtag that they will post on Twitter. Our website has successfully tested and deployed successfully.

### REFERENCES

- [1] G.Vinodhini and RM.Chandrasekaran, "Sentiment Analysis and Opinion Mining: A Survey", Volume 2, Issue 6, June 2012 ISSN: 2277 128X International Journal of Advanced Research in Computer Science and Software Engineering.
- [2] Zhongwu Zhai, Bing Liu, Hua Xu and Hua Xu, "Clustering Product Features for Opinion Mining", WSDM'11, February 9–12, 2011, Hong Kong, China. Copyright 2011 ACM 978-1-4503- 0493- 1/11/02...
- [3] Singh and Vivek Kumar, "A clustering and opinion mining approach to socio-political analysis of the blogosphere", Computational Intelligence and Computing Research (ICCRIC), 2010 IEEE International Conference.
- [4] Alexander Pak and Patrick Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining".
- [5] Bing Liu. "Sentiment Analysis and Opinion Mining, Morgan & Claypool Publishers, May 2012."

### Architecture Diagram:

