

RESUME CLASSIFICATION USING NAÏVE BAYES CLASSIFICATION ALGORITHM BY THE ANALYSIS OF TWITTER DATA

A PROJECT REPORT

Submitted by

GAYATHRI J (15C032)

SHINY R S (15C096)

SIVARANJANI J (15C099)

in partial fulfillment for the award of the degree

of

BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

THIAGARAJAR COLLEGE OF ENGINEERING, MADURAI – 15

(A Government Aided ISO 9001: 2008 Certified Autonomous Institution Affiliated to Anna University)



ANNA UNIVERSITY: CHENNAI 600 025

APRIL 2019

THIAGARAJAR COLLEGE OF ENGINEERING MADURAI-15

(A Government Aided ISO 9001: 2008 Certified Autonomous Institution Affiliated to Anna University)



BONAFIDE CERTIFICATE

Certified that this project report “**Resume Classification Using Naïve Bayes Classification Algorithm By The Analysis Of Twitter Data**” is the bonafide work of “**GAYATHRI J (15C032), SHINY R S (15C096), SIVARANJANI J (15C099)** ” who carried out the project work under my supervision during the Academic Year 2018-2019.

SIGNATURE

Dr.S.Mercy Shalinie, M.E., Ph.D.,
HEAD OF THE DEPARTMENT
COMPUTER SCIENCE AND ENGINEERING
THIAGARAJAR COLLEGE OF ENGG
MADURAI – 625 015

SIGNATURE

Mrs.B.Subbulakshmi B.E., M.E.,
Dr.M.Vijayalakshmi M.E., Ph.D.,
Mr.S.Prasanna B.E., M.E.,
SUPERVISORS AND ASSOCIATE PROFESSORS
COMPUTER SCIENCE AND ENGINEERING
THIAGARAJAR COLLEGE OF ENGG
MADURAI – 625 015

Submitted for the VIVA VOCE Examination held at Thiagarajar College of
Engineering on

INTERNAL EXAMINER

EXTERNAL EXAMINER

ACKNOWLEDGEMENT

I wish to express my deep sense of gratitude to **Dr.V.Abhai Kumar**, Principal of Thiagarajar College of Engineering for his support and encouragement throughout this project work.

I wish to express my sincere thanks to **Dr.S.MercyShalinie**, Head of the Department of Computer Science and Engineering for her support and ardent guidance.

I owe my special thanks and gratitude to **Dr.M.Vijayalakshmi, Mrs.B.Subbulakshmi, Mr.S.Prasanna** Associate Professors, Department of Computer Science and Engineering for her guidance and support throughout our project

I am also indebted to all the teaching and non-teaching staff members of our college for helping us directly or indirectly by all means throughout the course of our study and project work.

I extremely thank my parents, family members and friends for their moral support and encouragement for my project.

ABSTRACT

Nowadays candidates are often using Internet searches, social networking websites, and social media to find their suitable jobs where selecting an apt candidate with same set of skills is becoming a challenging task for the recruiters. We evaluate the feasibility of using twitter data to enhance the effectiveness of a recruitment system, especially for resume classification by using social network analysis method. So, the concept of social media analysis (twitter data) is recommended to extract the personal data of a candidate to examine the differences depending on the intended use of twitter and their degrees of friendship between twitter and real world friends. This experiment uses text classification to predict personality based on text written by Twitter users. The user's interactions with brand page as tweets, reactions or emotions are getting more relevance in present days.

TABLE OF CONTENTS

CHAPTER NO	TITLE	PAGE
	Abstract	ii
	List Of Figures	v
	List Of Abbreviations	vi
1	Introduction	1
2	Literature Survey	2
3	Problem Description	12
	3.1 Scope of the project	12
	3.2 Existing System	12
4	Proposed System	13
5	System Analysis	14
	5.1 Introduction	14
	5.2 Requirement Analysis	14
	5.2.1 Functional Requirements	14
	5.2.2 Non Functional Requirements	15
	5.2.3 Hardware Requirements	16
	5.2.4 Software Requirements	16
6	System Design	17
	6.1 Introduction	17
	6.2 Dataflow Diagram	17
	6.3 Use Case Diagram	18
7	METHODOLOGY	19
	7.1 Data Extraction	20
	7.2 Data Transformation	20
	7.2.1 Text Pre-Processing	21
	7.2.2 Tweet Cleaning	22

	7.2.3 Tweet Parsing	22
	7.3 Term Frequencies	23
	7.4 Naive Bayes	23
8	Implementation	24
	8.1 PHP	24
	8.2 WAMP	30
	8.3 Python	31
9	Results and Discussions	32
	9.1 Extraction Of Twitter Data	32
	9.2 Splitting Of Words	33
	9.3 Dropping Of Special Characters	34
	9.4 Term Frequency	35
	9.5 Remove The Words With The Frequency Lesser Than 10	36
	9.6 Finding The Positive Instance And Negative Instance	37
	9.7 Sentimental Analysis	38
	9.8 Finding The Accuracy	39
10	Conclusion And Future Work	40
	10.1 Future Work	40
	10.2 Conclusion	40
11	References	42
12	Plagiarism Report	45

LIST OF FIGURES

FIGURE NO	TITLE	PAGE NO
6.1	Dataflow Diagram	17
6.2	Use Case Diagram	18
7.1	Flow Graph	19
7.2	Data Extraction Diagram	20
7.3	Text Pre-Processing Diagram	21
8.1	User Login	25
8.2	Home Page	26
8.3	Job Listing	26
8.4	Job Description	27
8.5	Upload Resume	28
8.6	Job Search	29
8.7	Employer Login	30
9.1	Extraction Of Twitter Data	32
9.2	Splitted Words	33
9.3	Dropping Special Characters	34
9.4	Term Frequencies	35
9.5	Frequencies Greater Than 10	36
9.6	Positive And Negative Instances	37
9.7	Sentiment Analysis	38

LIST OF ABBREVIATIONS

PHP Hypertext Pre-processor

WAMP Windows, Apache, MySQL, and PHP

API Application Program Interface

CSV Comma-separated values

CHAPTER 1

INTRODUCTION

Nowadays people search job opportunity mainly online using LinkedIn, Firstnaukri.com, Guru, Career Builder, Class door, Cool-Works, Monster etc. The key problem is that most of the job hunting websites just display recruiting information to job applicants. The aim of the Project is developing an online search portal for the placement details for job seekers. To make works of Human Resource (HR) and hiring process easier, we analysis the social network data (Twitter data). Twitter is a social networking service attracts attention of researches in the past years. It provides many useful, both implicit and explicit social data. An example of the explicit data is user biography consisting of age, gender, and so on. An example of the implicit data is user behaviours. In Twitter there are two main social actions that user performs: post and tweets. It also allows additional actions like check-in and tag friends. Post and tweets are non-structures data which requires much effort to pre-process. On the other hand, the promising action that relate to user preference identification is 'like' pages and 'share' posts by pages. Main contribution of this work is to present technique that generates classification of CV based on user performances on social network and also their behaviours. In this work, we are using the popular classification algorithms: Naïve Bayes (NB) for evaluating performances, determining the probability of the polarity for Sentiment Analysis. By performing sentiment analysis, the particular user characteristics has been classified into either positive or negative traits by using which the recruiters choose what they need.

CHAPTER 2

LITERATURE SURVEY

[1] K. Shin and J. Lee, "A job applicants resume verification method using a social network analysis for a recruiting," 2017 11th International Conference on Software, Knowledge, Information Management and Applications (SKIMA), Malabe, 2017, pp. 15.

The research used a Facebook application, netvizz, to collect network data of subjects and Gephi for an analysis. Modularity was used to detect communities from the Facebook data that considered as experiences on resumes. Each detected community labeled with the text that showed the highest affiliation likelihood among the common words in an each of community members' Facebook profile. These labels are considered as experiences, and compared subjects' resume for verification. This research suggests a method to use SNS data to verify job applicants ' summary. The naturally recorded life log data of peoples could be a useful source to confirm the experiences of each person. Our main objective is to match detected community information from the Facebook data of a subject with described experience from the summary of a subject (Community=Experience).To examine this method for general user cases, we assessed the hypotheses: a recall rate will differ based on subjects using Facebook pattern.Experiments were conducted to learn how to adopt data from the social network to evaluate job applicants: two goals were suggested. We collected Facebook network data and tried to verify some job applicant experiences they described in their abstracts.

[2]R. Jiamthapthaksin and T. H. Aung, "User preferences on Facebook page categories based on user behaviour" 2017 9th International Conference on Knowledge and Smart Technology (KST), Chonburi, 2017, pp. 248-253.

User preferential profiling is important in both mining and recommendation systems for social networking. Facebook provides more than two hundred page category information related to user preferences, but the predefined categories may not fit the application well. Mapping these categories explicitly to a desirable set of user preferences is a tedious task. This paper proposes an effective user profiling technique using user behavior features built on the Facebook page in different categories. Models created from three well-known classification algorithms: Naïve Bayes (NB), Artificial Neural Network (ANN), and Support Vector Machine (SVM) turn user behavior raw data into an application-defined set of user preferences. The experiments conducted on the Facebook dataset show that the constructed features implicitly reflect user preferences and can be used to tailor preferences as needed. Among the three algorithms, SVM leverages the most accurately over 72 percent of classification performance. The main contribution of this work is to present technique that generates user preferences classification model from user behaviors on Facebook pages. Model users must specify a set of desirable user preferences based on the application. This work then builds features based on Facebook categories and uses them to build a classifier.

[3]A. Zaroor, M. Maree and M. Sabha, "JRC: A Job Post and Resume Classification System for Online Recruitment," 2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI), Boston, MA, 2017, pp. 780-787.

Because of the increasing growth in online recruitment, traditional hiring methods are becoming inefficient. This is due to the fact that job portals receive huge numbers of unstructured resumes-in various styles and formats-from applicants with different fields of expertise and specialization. Therefore, the extraction of structured information from the applicant's resumes is needed not only to support the candidates ' automatic screening, but also to effectively route them to their respective occupational categories. This helps minimize the effort required by employers to manage and organize resumes as well as screen out irrelevant candidates. To demonstrate the effectiveness of the proposed system, we conducted several experiments using a real-world recruitment dataset. In addition, we evaluated the efficiency and effectiveness of the proposed system against state-of - the-art online recruitment systems. First, the proposed system uses section-based segmentation module to segment the resumes and extract a set of skills that are used in the classification process. Next, the system uses an integrated knowledge base of skills to perform the classification task.

[4] B.Y.Pratama and R. Sarno, "Personality classification based on Twitter text" 2015 International Conference on Data and Software Engineering (ICoDSE), Yogyakarta, 2015, pp. 170-174.

A person has to take various personality tests to find out their personality. Personality tests can be self-descriptive psychologist reporting, interviewing, or observation. These traditional methods are expensive and less practical. Recently there has been a personality test in the form of an online questionnaire via the website. While this can be quite practical, the user still has to take action to answer various questions. A recent study shows that personality traits can be automatically obtained from the text they wrote. The choice of the most frequently used words can describe that particular person's personality. Social media is a place where users represent themselves in the world. Social media account is private and personal in order to reflect their personal lives. Social media activities such as posting, commenting and updating status can reveal personal information. Text left by users can be analyzed to obtain information, in this case the user's personality. People are required to take a personality test in order to find out their personality. Social media is a place where users are expressing themselves to the world. Posts made by social media users can be analyzed to obtain their personal information. This experiment uses text classification to predict personality based on text written by users of Twitter.

[5] A. Teixeira and R. M. S. Laureano, "Data extraction and preparation to perform a sentiment analysis" 2017 12th Iberian Conference on Information Systems and Technologies (CISTI), Lisbon, 2017, pp. 1-6.

Social media is a very much discussed topic in the present. Opinion Mining and Sentiment Analysis was driven by the increasing availability of the internet and the growth of social media platforms. One of Social Media Platform's most popular is Facebook. Users / consumers can express their feelings with comments or emotion buttons on this Social Media Platform. The interactions of the user with a brand page as posts, likes, shares or comments are becoming more relevant today. The analyzed data can provide decision makers with new approaches for running their business, understanding their brand value against competitors, or even better understanding what customer or potential customer evaluations are about brand products or services. Using open source tools to perform a Sentiment Analysis, extract and prepare data collected on Facebook. Sentiment Analysis can provide useful information on current and new consumers ' preferences. Knowing opinions, likes, dislikes the company can improve strategies by creating or enhancing methods for better approaching their consumers to create a positive feeling and eliminate the negative ones. You can provide the information sets collected through Sentiment Analysis and add them to Business Intelligence. Using this information correctly can be valuable to organizations and can help measure impact decisions in order to gain benefits and leverage the business.

[6] Nadeem Ahmad and Jawaid Siddique, "Personality Assessment using Twitter Tweets" 2017 21st International Conference on KES.

Social media establishes uninterrupted connectivity between its users and the outside world by revealing personal details and viewpoints in every aspect of life. The focus of this study is to analyze how twitter (dataset) can be used with personality assessment to improve the user experience. The article, in its primary context, shows users' psychological profiling based on the dataset. This profiling can be a very useful tool for career progression, job satisfaction and setting preferences in various interfaces. We propose a way to predict the personality of the user by using DISC (Dominance, Influence, Compliance, Steadiness) assessment to map information available to the public on their personal Twitter. The results of this study may be useful for information retrieval (search engine), content selection mechanism, and product & services positioning. The growth effects of social network sites (SNS) are very heavy on the techniques developed in social networks for text mining. Text mining provides an efficient way to execute and use data sets. In this paper, we reviewed the text mining technique (clustering) used to analyze social networks and mapped its results to a framework called DISC. The practical implementation of the results can be in the field of user information retrieval, content selection, product positioning and psychological assessment. The peculiarity of this research study is that it provides a framework for the entire use of text mining technique in a novel area. Keeping the same track, geolocation parameters (latitude & longitude) can be added to regionalize the assessment. Trend and taste parameters can be added to gain a comprehensive personality profile that can provide a fundamental basis for extending the findings of this research to other disciplines. We invite other researchers to come forward and take initiatives to enrich the body of knowledge in the very

right perspective of this new area and take the matter further by exploring new-captured dimensions.

[7]. Chris Sumner, Alison Byers, Rachel Boochever, Gregory J.Park, “Predicting Dark Triad Personality Traits from Twitter Usage and a Linguistic Analysis of Tweets”, 2012 11th International Conference on Machine Learning and Applications.

This mainly focused on the Five Big characteristics of personality, but the area that is relatively unexplored is the anti-social characteristics of narcissism, machiavellians and psychopathy, commonly known as the Dark Triad. This study explored the extent to which it is possible to determine anti-social personality traits based on Twitter use. This was accomplished by comparing 2,927 Twitter users' Dark Triad and Big Five personality traits with their profile attributes and language usage. Analysis shows that there are some statistically significant relationships between these variables. Through the use of crowd sourced machine learning algorithms, we show that machine learning provides useful prediction rates, but is imperfect in predicting an individual's Dark Triad traits from Twitter activity. While predictive models may be inappropriate to predict the personality of an individual, they may still be of practical importance when models are applied to large groups of people, such as gaining the ability to see if anti-social traits increase or decrease over a population. Our results raise important questions concerning the unregulated use for screening purposes of social media analysis. It is important to have a better understanding of the practical and ethical implications of drawing conclusions on personal information embedded in social media sites. This study highlights the relationships between Twitter activity, Dark Triad, and Big Five personality

traits, but the practical performance of machine prediction is currently poor when applied directly to an individual.

[8]. Alireza Souri, Shafigheh Hosseinpour and Amir Masoud Rahmani, “Personality classification based on profiles of social networks’ users and the five-factor model of personality”, 2018 Human-centric Computing and Information Sciences.

This paper presents a hypothesis that when cooperating through social networks, users of similar personality are expected to show mutual behavioral patterns. With the goal of personality recognition in terms of analyzing user activity on Facebook, we collected information about users' personality traits and their profiles on Facebook, making an application flourish using Facebook API. This study's participants are 100 Facebook user volunteers. Based on all the data collected, classifiers were learned to recognize user personality by their profile using different data mining techniques and without completing any questionnaire. By comparing the results of the classifiers, our proposed model with 82.2% accuracy was the boosting-decision tree more accurate than previous studies which were able to predict personality according to the variables in their profiles in five factors to use it as a model for personality recognition.. Therefore, in this work, we got assistance from data mining which discovered fruitful information from a series of irrelevant data. By knowing the personality, this model can be used for other purposes, such as the recommender system of friends and social groups, and even can be used for promotional purposes. With other techniques such as text mining via utilizing words in posts and comments in user's timelines we can also predict personality. Moreover, the other proposal is to find out what kind of photos will be used by each of the five kinds of personalities as their profile photos. Ultimately for increasing accuracy of classifiers, we can use fuzzy classification in future works.

[9]. Xiao-FengZhong, Shi-ZeGuo, Liang Gao, Hong Shan and Di Xue, “ A General Personality Prediction Framework Based on Facebook Profiles”, ICMLC 2018 Proceedings of the 2018 10th International Conference on Machine Learning and Computing.

Social network services are growing significantly and in the last decade they have generated massive data. Personality prediction has been an important research topic not only in psychology but also in computer science to describe user profiles and people. However, the prediction results were usually not able to obtain the user's actual personality due to the diversity of online users. To solve the problem, this paper proposes a framework for personality prediction consisting of outlier elimination, selection of training datasets and personality prediction. Mahalanobis distance and Z-value are utilized to compute the distance of samples and identified outliers. Maximum Margin Criterion is used to select the training dataset and SVM algorithm is applied to predict personality. The experimental results on Facebook dataset show that our framework is effective and capable of improving the prediction accuracy.

[10]. Gabriel Yakub N.N. Adi, Michael Harley Tandio, Veronica Ong, and DerwinSuhartono, “Optimization for Automatic Personality Recognition on Twitter in Bahasa Indonesia”, 2018 3rd International Conference on Computer Science and Computational Intelligence.

This paper presents optimization techniques for automatic personality recognition (APR) based on Twitter in Bahasa Indonesia, the mother tongue of Indonesians. Foremost, we discuss Twitter and its utilization as a resource for many types of research. Several previous studies have been attempted to predict users’ personality automatically. However, only a few of them have done their research for Bahasa Indonesia data. Therefore, this paper discusses the optimization of APR in Bahasa Indonesia. We evaluate a series of techniques implementing hyperparameter tuning, feature selection, and sampling to improve the machine learning algorithms used. The personality prediction system is built on machine learning algorithms. There are three machine learning algorithms used in this study, namely Stochastic Gradient Descent (SGD), and two ensemble learning algorithms, Gradient Boosting (XGBoost), and stacking (super learner). By implementing this series of optimization techniques, the current study’s evaluation results show huge improvement by achieving 1.0 ROC AUC score with SGD and Super Learner. In this research, we emphasize on the optimization of APR by implementing three different techniques. Whereas previous study highlights methods for performing APR on contents in Bahasa Indonesia. Series of training and testing was done using different scenarios to monitor the improvements of the machine learning performance.

CHAPTER 3

PROBLEMDESCRIPTION

Selecting an apt candidate with same set of skills is tedious job for recruitment system. Processing the similar resumes is challenging task for recruiters. Nowadays, filtering of a candidate is becoming a tedious process. To make it simple and easier Sentimental analysis is used Sentiment analysis for job recruitment is not implemented for the classifications of resumes. To do sentiment analysis for determining the personality of a candidate to classify opinions expressed in a tweet (document) in order to determine whether the candidate's personality towards the recruiter is positive or negative. Based on the probability result of the positive and negative tweets the recruiters can able to predict the personality of the candidate by which the recruiters will be able to know more about the candidate that how he interacts with the peoples and how he is socialized.

3.1 SCOPE OF THE PROJECT

The scope of the project is used to get the user behaviour in social media for choosing a candidate for a job.

3.2 EXISTING SYSTEM

In existing system, there is screening and checking the skill mapping and authenticity of millions of resume. With only idea of a candidate's qualifications recruiters fail to notice their personality or mind-set.

CHAPTER 4

PROPOSED SYSTEM

The general concept of the proposed framework is to help the job recruiters by analysing the personality of a particular candidate by extracting the twitter data. Nowadays Many of the job recruiters not only looking the knowledge of the candidates more than that analysing the personality is important to predict whether the candidate with a positive thoughts or with a negative thoughts. Performs sentiment analysis (whether the polarity is positive or negative). By performing the sentimental analysis we can predict whether tweets of the user is a positive or negative. In sentimental analysis various Machine Learning algorithm is used. In our problem we used a naïve bayes algorithm which will produce a probability as output. Based on the output we can predict whether the tweets are Positive one or a Negative one. Now it is easy for the job recruiters whether to select a candidate or not.

CHAPTER 5

SYSTEM ANALYSIS

5.1 INTRODUCTION

This chapter contains the detailed description of the requirement used in the entire project. They are segregated into functional and non-functional requirements.

5.2 REQUIREMENT ANALYSIS

5.2.1 Functional Requirements

UI Design

Candidates are allowed to submit the resume after proper log in and registration. They can search the suitable jobs of their interest and apply for their favorite job. HR can view the submitted resumes for choosing the candidate. Notifications are pushed to the particular candidate, if their resume is matched with the recruiter's expectations

INPUT

Tweets- The textual representation of candidate's tweets and their resumes submitted by candidate.

PROCESS

According to the given input, the code was specified with a condition to predict the personality of user by using sentiment analysis. The polarity value was calculated for each candidate and then classified the resumes based on the calculated polarity value.

OUTPUT

- Polarity analyzed
- Predicted the candidate's personality

5.2.2 Non Functional Requirements

USER FRIENDLY

Since the project is developed in WAMP server. WAMP stands for Windows (W), Apache (A), MySQL (M), PHP (P) and Perl (P). It is a simple, lightweight Apache distribution; it is easy for developers to create a local web server for testing purposes.

PLATFORM INDEPENDENT

The Project is implemented using PHP, it is Platform independent.

EASE OF USE

Since it is available in the web, people can use it freely and increases efficiency along with the number of users.

Some Other Non-Functional Requirements:

- Security
- Reliability
- Memory Efficiency
- Authentication
- Confidentiality
- Data Integrity
- Non-repudiation

5.2.3 Hardware Requirements

Processor: Intel(R) Core(TM) i3-4005U CPU

@1.70GHz

Installed Memory (RAM):4.00GB

System Used: 64-bit Operating System

5.2.4 Software Requirements

Technology Used : BOOTSTRAP, PHP, PYTHON

Database : MY-SQL

Database Connectivity : PHP

CHAPTER 6

SYSTEM DESIGN

6.1 INTRODUCTION

System design is the process of defining a system's architecture, components, modules, interfaces, and data to meet specific requirements. This stage's output will describe the new system as a module or subsystem collection.

6.2 DATAFLOW DIAGRAM

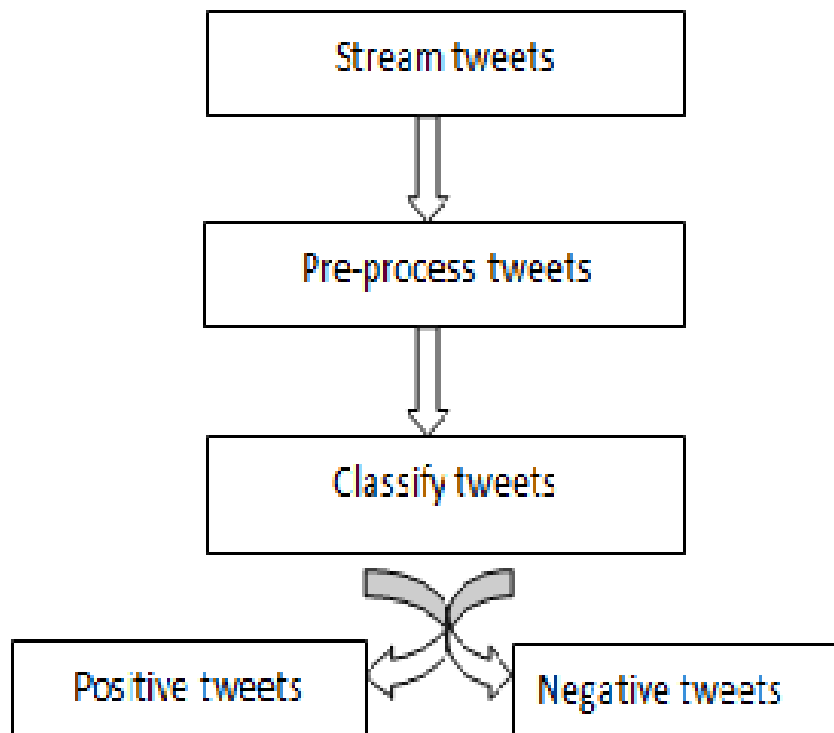


Fig 6.1 DATAFLOW DIAGRAM

6.3 USE CASE DIAGRAM

A Unified Modeling Language (UML) use case diagram is a type of behavioral diagram that is defined by and created from a use case analysis.

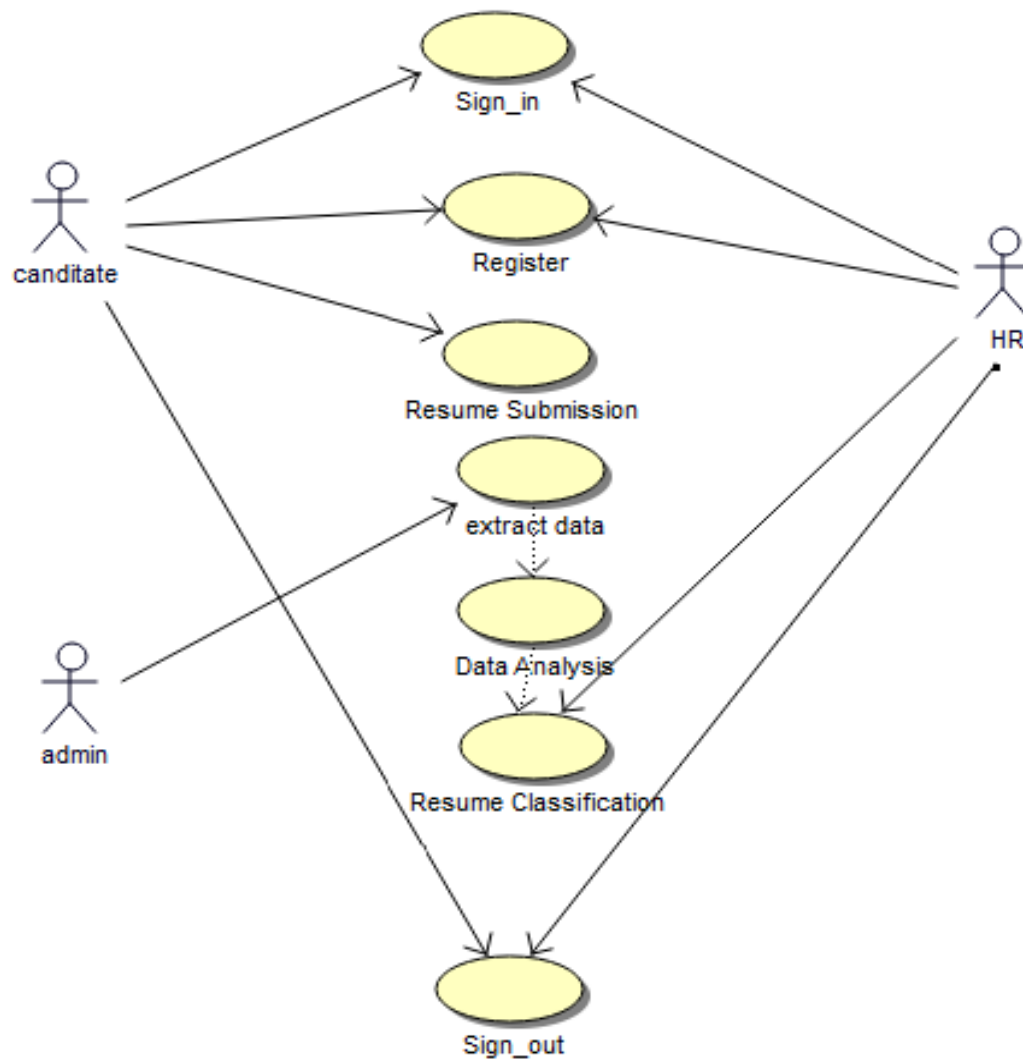


Fig 6.2 Use case Diagram

CHAPTER 7

METHODOLOGY

There are two major social actions performed by the user on Twitter: post and tweet. Additional actions such as check-in and tag friends are also allowed. Post and tweets are data from non-structures that require a lot of pre-processing effort. After selecting the tweet data set, these tweets were cleaned from emoticons, unnecessary punctuation marks were created and databases were created to store these data in a specific database. Once the candidate upload the resume, The admin who will extract the twitter data and process those tweets and categorized those tweets as “positive tweets” and “negative tweets” .By finding the frequency factor we can predict the personality of the candidate. Based on the result the recruitment team can either select the candidate or reject them.

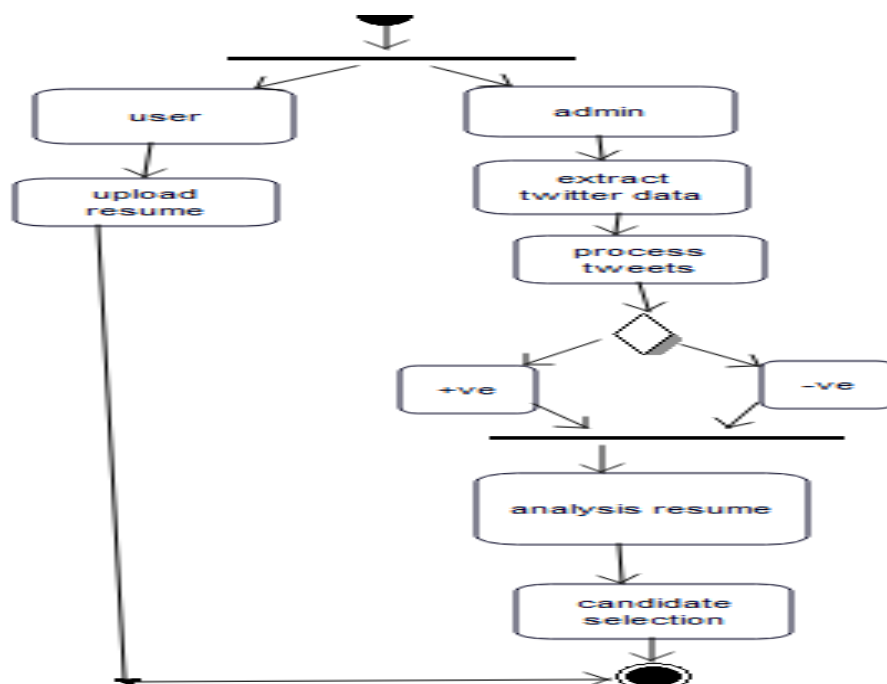


Fig 7.1 Flow Graph

7.1Data Extraction

It was the first step of the entire process. Python was used here as a scripting language to extract twitter data using a twitter API called tweepy. The data was extracted using tweets of all the users along with the major events occurred in that particular time frame to ensure that the tweets extracted are relevant and according to the requirements. Extracted dataset consisted of text of the tweet along with the date and time of the tweet.

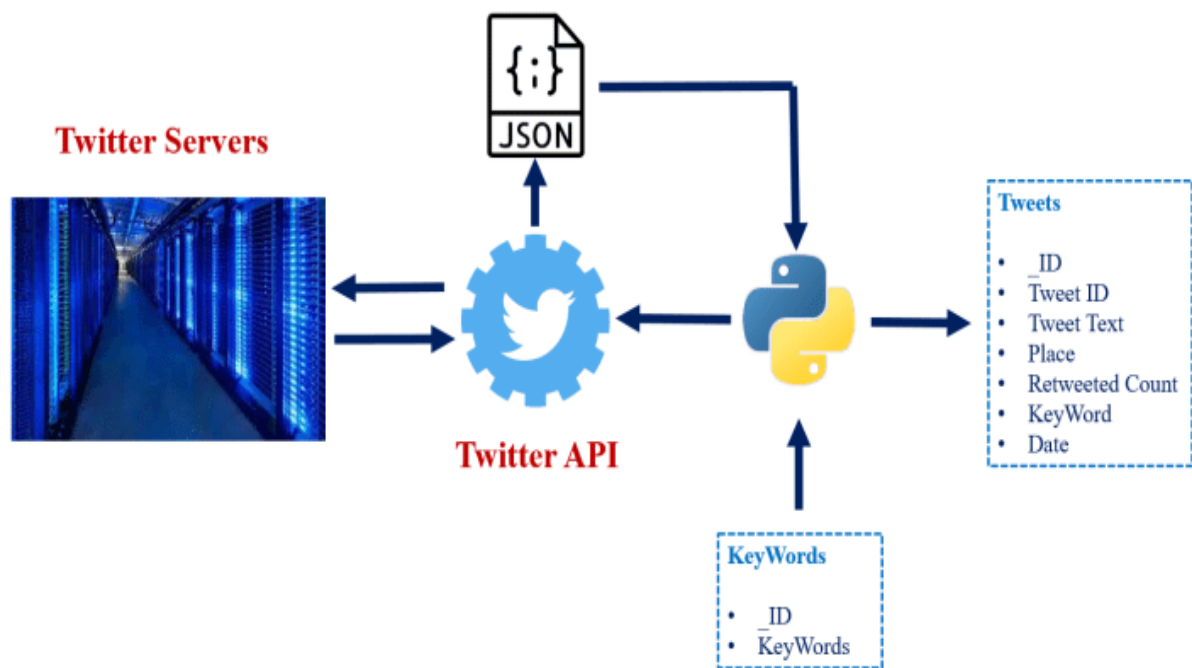


Fig 7.2 Data extraction diagram

7.2 Data Transformation

Data can be extracted through Twitter's API and accessed by tweepy library. Authentication of Twitter API is done by using the access token, access secret, consumer key and consumer secret. In order to use them, we had to clean and transform it into more usable structured dataset to ensure that our next phases of the process are smooth, easy efficient and effective.

7.2.1 Text Pre-Processing

Split the sentences into words or phrase is known as pre-processing. Breaking of the sentences in tweets into unique words from data frame. And then removing special character and emoticons in them. (☺, ☹, //, #, @, !, :-), :- ().

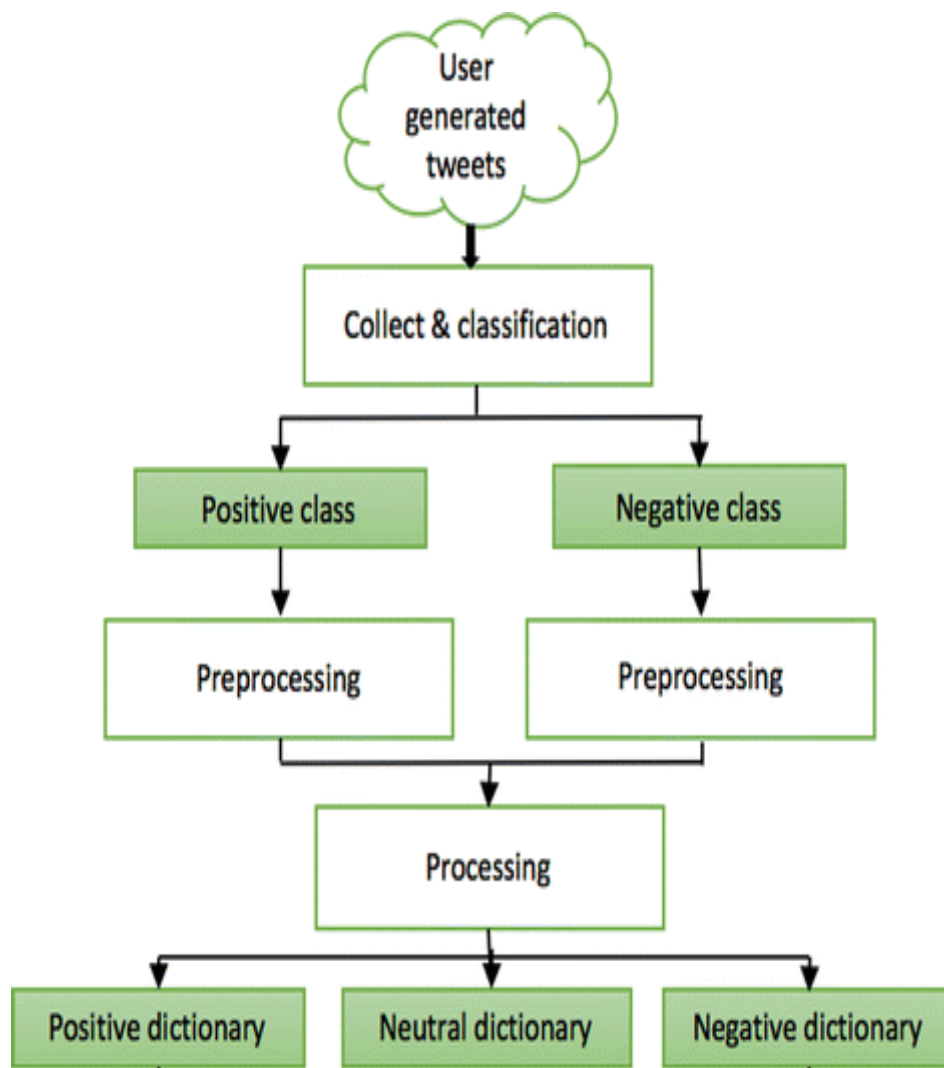


Fig 7.3 Text Pre-Processing Diagram

7.2.2 Tweet cleaning

To accomplish this process, this task consists of three subtasks.

1. Converting all Tweets into Lower case

In order to bring the tweets in a consistent form, we had to convert all tweets to lower case. We can carry out further transformation and classification without worrying about data non-consistency. This task is performed using the Lower() (function of python in Python IDE. This function converted all alphabets in lower cases and solved the problem of case sensitivity by making all data consistent in lower cases.

2. Removing emoticons and punctuations

We had to remove emotions and punctuations because they were not needed in our analysis. Here we will remove any character or string of the characters that are non readable in ASCII values. These cleaned data's are transformed in to the data frames and saved them as a csv file.

7.2.3 Tweet Parsing

After cleaning the data, the selected tweets are then parsed.

1. Removal of duplicates

We have the list of words that appear in each individual file(happy, sad).we want to combine them all in one data frame and save it in a csv file. Then drop duplicates get rid of any word that appears multiple times and then these data frames are saved in a new csv file.

2. Splitting of tweets

Now we have two tweet files (happy.csv, sad.csv) and each column would include the list of words that appeared as an array in the tweet. So each row is a tweet and in the text column there is an array of words that appeared in the tweet and stored as a data frame. If rows have a "nan" value, it will drop it and store these data frames in a new csv file

7.3 Term Frequencies

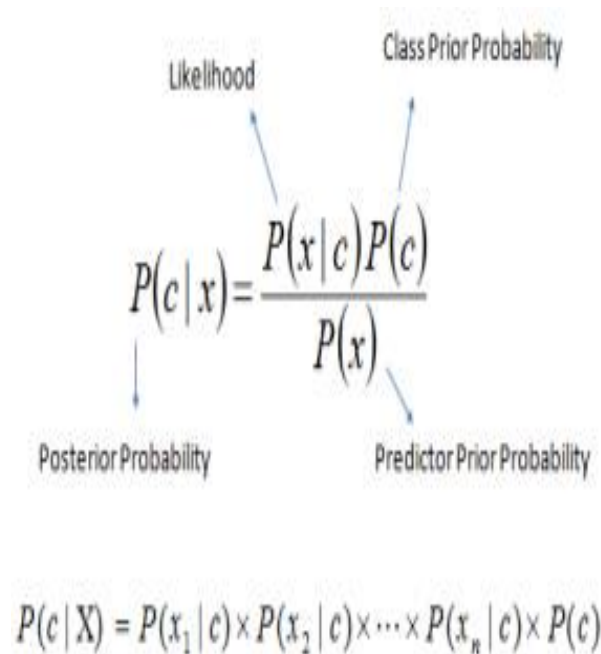
Term Frequency is to determine how important a word is to a document in a collection or corpus. It is often used in information retrieval and text mining as a weighting factor. Term frequency is simply the ratio of the count of a word in a sentence to the length of the sentence.

Term Frequency can be calculated by

TF = (Number of times term T appears in the particular tweet) / (number of terms in that tweet)

7.4 Naïve Bayes

It is used to determine exact probabilities for hypothesis and it is also robust to noise in input data



The diagram shows the Naïve Bayes formula with labels for its components:

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

Labels and arrows:

- Likelihood** points to $P(x|c)$
- Class Prior Probability** points to $P(c)$
- Posterior Probability** points to $P(c|x)$
- Predictor Prior Probability** points to $P(x)$

$$P(c|X) = P(x_1|c) \times P(x_2|c) \times \dots \times P(x_n|c) \times P(c)$$

CHAPTER 8

IMPLEMENTATION

8.1 PHP:

PHP (recursive acronym for PHP: Hypertext Pre-processor) is a widely used open source general-purpose scripting language that is particularly suitable for web development and can be embedded in HTML.

PHP focuses primarily on server-side scripting, so you can do anything that any other CGI program can do, such as collecting data, generating dynamic page content, or sending and receiving cookies. But PHP can do a lot more. There are three main areas in which PHP scripts are used.

1. Scripting on the server side. This is PHP's most traditional and main target field. To make this work, you need three things. The parser of PHP (CGI or server module), a web server and a web browser. You have to run the web server with a connected php installation. With a web browser, you can access the PHP program output by viewing the PHP page through the server. All of these can run on your home machine if you're just experimenting with PHP programming.
2. To run it without any server or browser, you can make a PHP script. You only need the PHP parser to use it this way. This type of use is ideal for scripts that are regularly executed using cron (Unnix or Linux) or Task Scheduler (on Windows). These scripts can also be used for simple tasks of text processing.
3. PHP is probably not the best language to create a desktop application with a graphical user interface, but if you know PHP very well and would

like to use some advanced PHP features in your client-side application, you can also use PHP-GTK to write such programs.

1. User Login

Created a login page for the candidate's . If the candidate is new user then the candidate should do registration otherwise the candidate can login to the page.

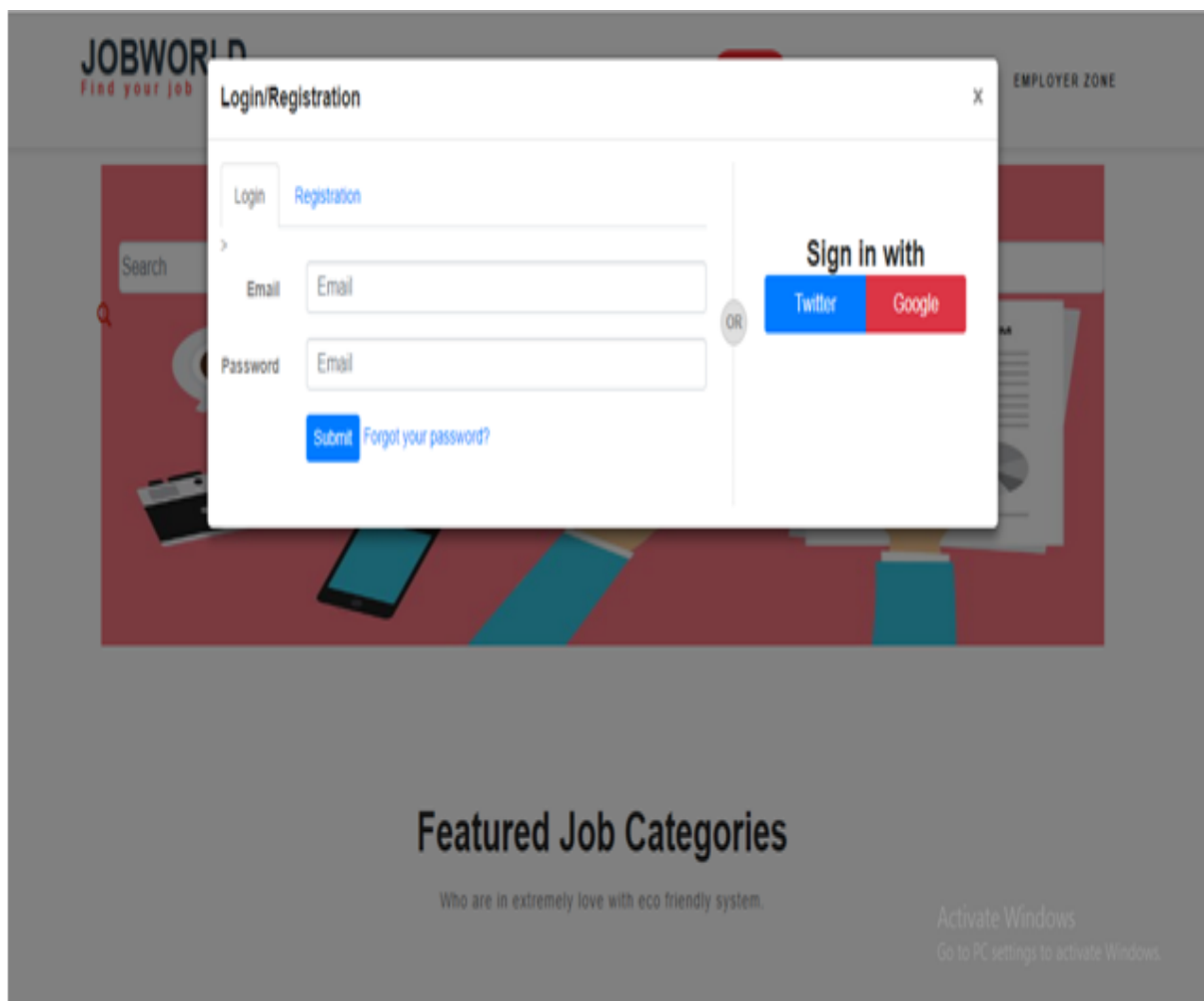


Fig 8.1 USER LOGIN

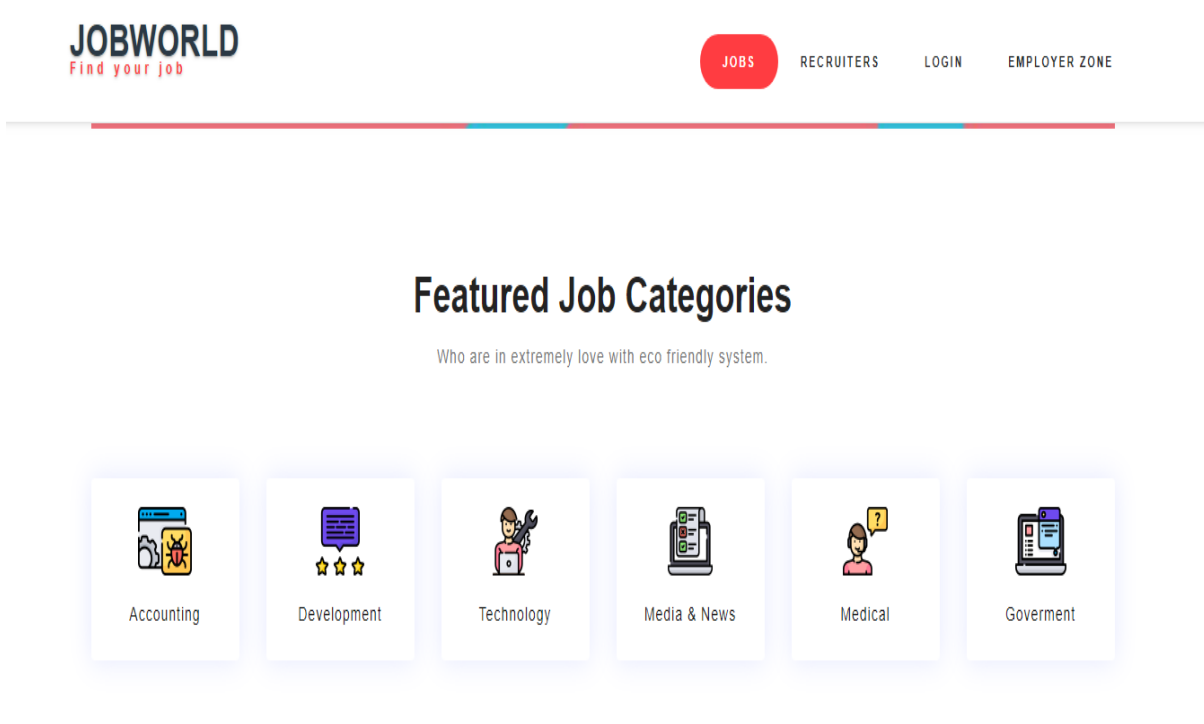


Fig 8.2 HOME PAGE

2. Listing of Job

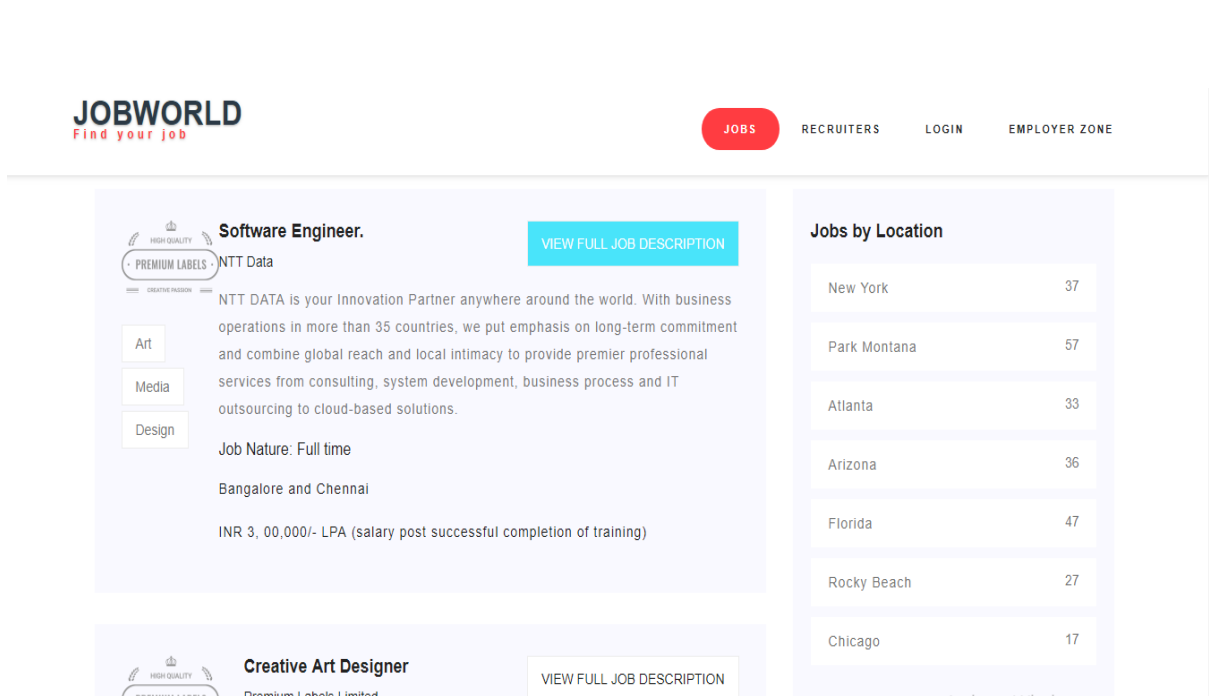




Fig 8.3 JOB LISTING

3. Job Description

The Candidate can able to see the Jobs available and the detailed Information of the jobs were also available. For example: Salary offered, stipend, job location etc...



[JOBS](#)[RECRUITERS](#)[NOTIFICATIONS](#)[LOGIN](#)[EMPLOYER ZONE](#)



NTT DATA is your Innovation Partner anywhere around the world. With business operations in more than 35 countries, we put emphasis on long-term commitment and combine global reach and local intimacy to provide premier professional services from consulting, system development, business process and IT outsourcing to cloud-based solutions.

Visit: www.nttdata.com/americas to learn how our consultants, projects, managed services, and outsourcing engagements deliver value for a range of businesses and government agencies.

Profile Offered: Software Engineer.

Stipend: Rs. 12,000/- as stipend during training period.(3 Months)

Salary offered: INR 3, 00,000/- LPA (salary post successful completion of training)

Experience: Fresher

Course Specialization: B.E/B.Tech(CS/IT), MCA (2018 Batch)

Job Location: Bangalore and Chennai

Interview process:

Group Discussion.

Technical Interview

Management Round

Qualification and Pre-requisite:

4 years full time Engineers 2018 (CSE & IT) & MCA.

60% aggregate throughout their education (X, XII and Engg).

History of Arrears not more than 2.

Activate Windows
Go to PC settings to activate Windows.

Fig 8.4 JOB DESCRIPTION

4. Uploading Resumes

The candidate can upload their resumes to get more Job opportunities.

The screenshot shows a web application interface for 'JOBWORLD' with the tagline 'Find your job'. The background is split into a grey top half and a dark green bottom half. In the top right, there are links for 'NOTIFICATIONS', 'LOGIN', and 'EMPLOYER ZONE'. On the left, under 'Quick links', are 'Home', 'About', 'FAQ', 'Get Started', and 'Videos'. A white modal window titled 'Upload Your Resume' is centered, featuring a close button 'X' in the top right. The form contains three input fields: 'Your name', 'Your email', and 'Phone Number'. Below these is a file upload section with an 'Upload' button, a 'Choose file' button, and a 'Browse' button. A 'Send' button is at the bottom of the modal. At the very bottom of the page, a small line of text reads: 'National Transaction Corporation is a Registered MSP/ISO of Elavon, Inc. Georgia (a wholly owned subsidiary of U.S. Bancorp, Minneapolis, MN)'.

JOBWORLD
Find your job

NOTIFICATIONS LOGIN EMPLOYER ZONE

Quick links

- Home
- About
- FAQ
- Get Started
- Videos

Upload Your Resume X

Your name

Your email

Phone Number

Upload Choose file Browse

Send

National Transaction Corporation is a Registered MSP/ISO of Elavon, Inc. Georgia (a wholly owned subsidiary of U.S. Bancorp, Minneapolis, MN)

Fig 8.5 UPLOAD RESUME

5Job Search

Once the candidate upload their resumes they can search the jobs based on location or by work permit or by the Domain that they were specialized.

JOBWORLD
Find your job

JOBS

RECRUITERS

LOGIN

EMPLOYER ZONE

All Candidates

Search

I'm looking for a ...

Job

in

Location


Industry / Role

Search

Specialism

Accountancy (2)


Rankings (2)



Comapany Name

Sales & Marketing

Istanbul / Turkey



4 Open Position

Artivista Windows

3 Open Position

Fig 8.6 JOB SEARCH

29

Login or Create a Free Account as Employee!

New Account:

☒ Please inform me of upcoming Promotions and news

Login:

[I forgot my Password!](#)

Activate Windows

Fig 8.7 EMPLOYER LOGIN

8.2 WAMP

WAMP stands for "Windows, Apache, MySQL, and PHP." WAMP is a LAMP variation for Windows systems and is often installed as a software bundle (Apache, MySQL, and PHP). It is often used for web development and internal testing, but can also be used to serve live websites. The most important part of the WAMP package is Apache (or "Apache HTTP Server"), which is used to run the Windows web server. A web developer can test web pages in a web browser without publishing them live on the Internet by running a local Apache web server on a Windows machine.

WAMP also includes MySQL and PHP, two of the most common technologies used to create dynamic websites. MySQL is a high-speed database, while PHP is a scripting language that can be used to access database data. By installing these two components locally, a developer can build and test a dynamic website before publishing it on a public web server. While Apache, MySQL, and PHP are open source components that can be individually

installed, they are usually installed together. One popular package is called "Wamp Server," which provides a user-friendly way to install and configure Windows "AMP" components.

8.3 PYTHON

Python is a language of high-level, interpreted, interactive and object-oriented scripting. Python is designed to be highly legible. It frequently uses English keywords where punctuation is used as other languages, and it has fewer syntactic constructions than other languages. Python was developed by Guido Van Rossum at the National Math Research Institute in the late eighties and early nineties. Python is derived from many other languages, including ABC, Modula-3, C, C++, Algol-68, Small Talk, Unix shell and other scripting languages. Python is protected by copyright. Python source code, like Perl, is now available under the GNU General Public License (GPL). Python is now maintained by a core development team at the institution, although Guido Van Rossum still plays a vital role in directing his profession

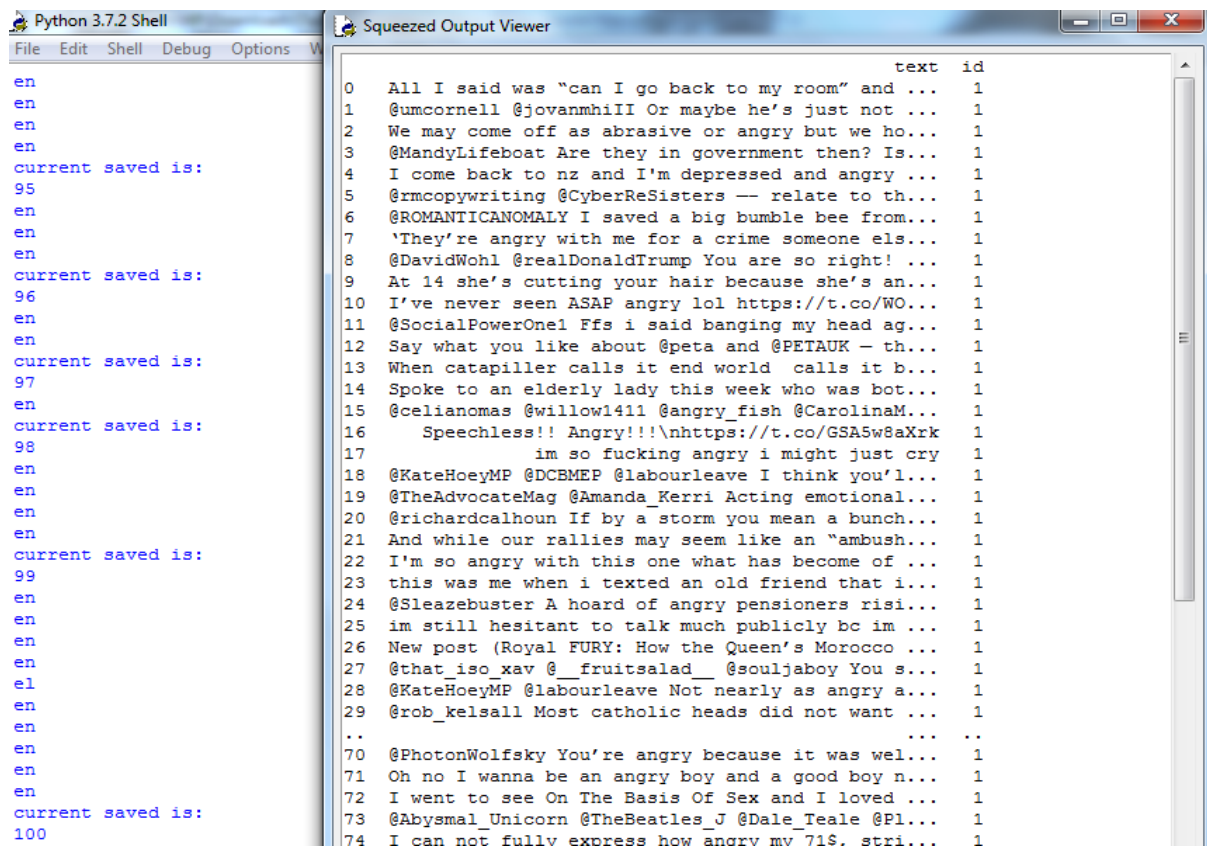
1. Python is interactive – you can actually sit at a Python prompt and interact directly with the interpreter to write your programs.
2. Python is Object-Oriented – Python supports Object-Oriented style or programming techniques that encapsulate code within objects.
3. Python is easy to use- It has many modules which can be installed using pip command.
4. To access the Twitter data in python using a Twitter Api's. By installing "Tweepy" using "pip installTweepy". Tweepy that it is used to interact with the twitter more easily and those data's can accessed using Tweepy library
5. It is important to authenticate the twitter Api using access token, access secret, consumer key, consumer secret

CHAPTER 9

RESULTS AND DISCUSSIONS

9.1 Extraction of Twitter Data

Here, we extract the tweets of one person and to get the data we will be using Twitter's API. Those data's are transformed into a dataframe and then as csv file



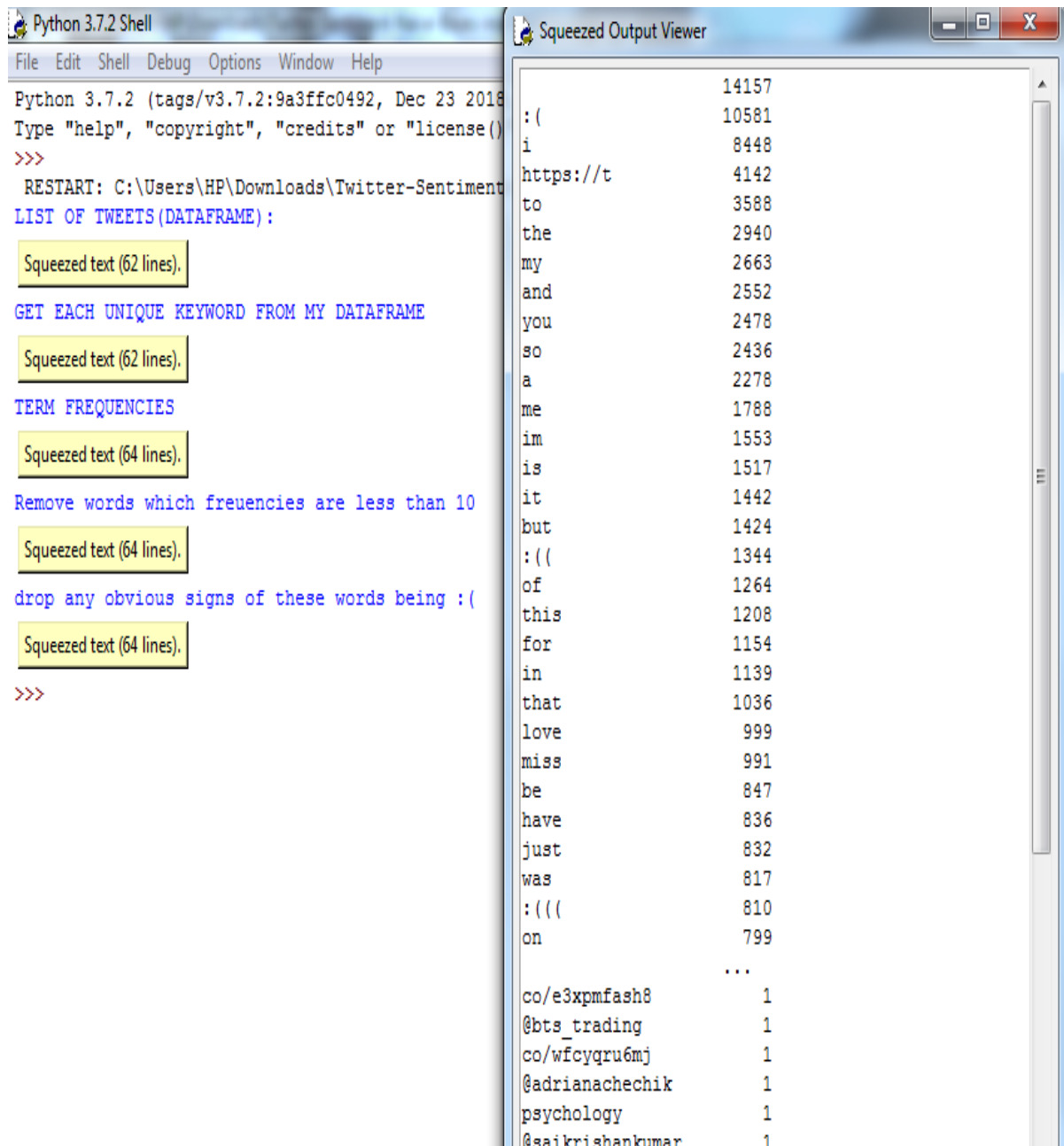
The screenshot displays two windows. The left window, titled 'Python 3.7.2 Shell', shows a series of 'en' commands and 'current saved is:' prompts, indicating a loop or iterative process. The right window, titled 'Squeezed Output Viewer', displays a list of tweets with their corresponding IDs. The tweets are numbered from 0 to 74, and each entry includes the tweet text and its ID. The tweets are as follows:

id	text
0	All I said was "can I go back to my room" and ...
1	@umcornell @jovanmhiII Or maybe he's just not ...
2	We may come off as abrasive or angry but we ho...
3	@MandyLifeboat Are they in government then? Is...
4	I come back to nz and I'm depressed and angry ...
5	@rmcopywriting @CyberReSisters — relate to th...
6	@ROMANTICANOMALY I saved a big bumble bee from...
7	'They're angry with me for a crime someone els...
8	@DavidWohl @realDonaldTrump You are so right! ...
9	At 14 she's cutting your hair because she's an...
10	I've never seen ASAP angry lol https://t.co/WO...
11	@SocialPowerOne1 Ffs i said banging my head ag...
12	Say what you like about @peta and @PETAUK - th...
13	When catapiller calls it end world calls it b...
14	Spoke to an elderly lady this week who was bot...
15	@celianomas @willow1411 @angry_fish @CarolinaM...
16	Speechless!! Angry!!!\n https://t.co/GSA5w8aXrk
17	im so fucking angry i might just cry
18	@KateHoeyMP @DCBMEP @labourleave I think you'l...
19	@TheAdvocateMag @Amanda_Kerri Acting emotional...
20	@richardcalhoun If by a storm you mean a bunch...
21	And while our rallies may seem like an "ambush...
22	I'm so angry with this one what has become of ...
23	this was me when i texted an old friend that i...
24	@Sleazebuster A hoard of angry pensioners risi...
25	im still hesitant to talk much publicly bc im ...
26	New post (Royal FURY: How the Queen's Morocco ...
27	@that_iso_xav @_fruitsalad_ @souljaboy You s...
28	@KateHoeyMP @labourleave Not nearly as angry a...
29	@rob_kelsall Most catholic heads did not want ...
..	..
70	@PhotonWolfsky You're angry because it was wel...
71	Oh no I wanna be an angry boy and a good boy n...
72	I went to see On The Basis Of Sex and I loved ...
73	@Abysmal_Unicorn @TheBeatles_J @Dale_Teale @Pl...
74	I can not fullv express how anorv mv 71\$. stri...

Fig 9.1EXTRACTION OF TWITTER DATA

9.2 Splitting of Words

It Split the strings of the dataframe into individual words by splitting after every space. It finds a unique word and count how many times it appears



The image shows a Python 3.7.2 Shell window on the left and a Squeezed Output Viewer window on the right. The shell window displays the following code and output:

```
Python 3.7.2 (tags/v3.7.2:9a3ffc0492, Dec 23 2018)
Type "help", "copyright", "credits" or "license()"
>>>
RESTART: C:\Users\HP\Downloads\Twitter-Sentiment
LIST OF TWEETS (DATAFRAME):
Squeezed text (62 lines).
GET EACH UNIQUE KEYWORD FROM MY DATAFRAME
Squeezed text (62 lines).
TERM FREQUENCIES
Squeezed text (64 lines).
Remove words which frequencies are less than 10
Squeezed text (64 lines).
drop any obvious signs of these words being :(
Squeezed text (64 lines).
>>>
```

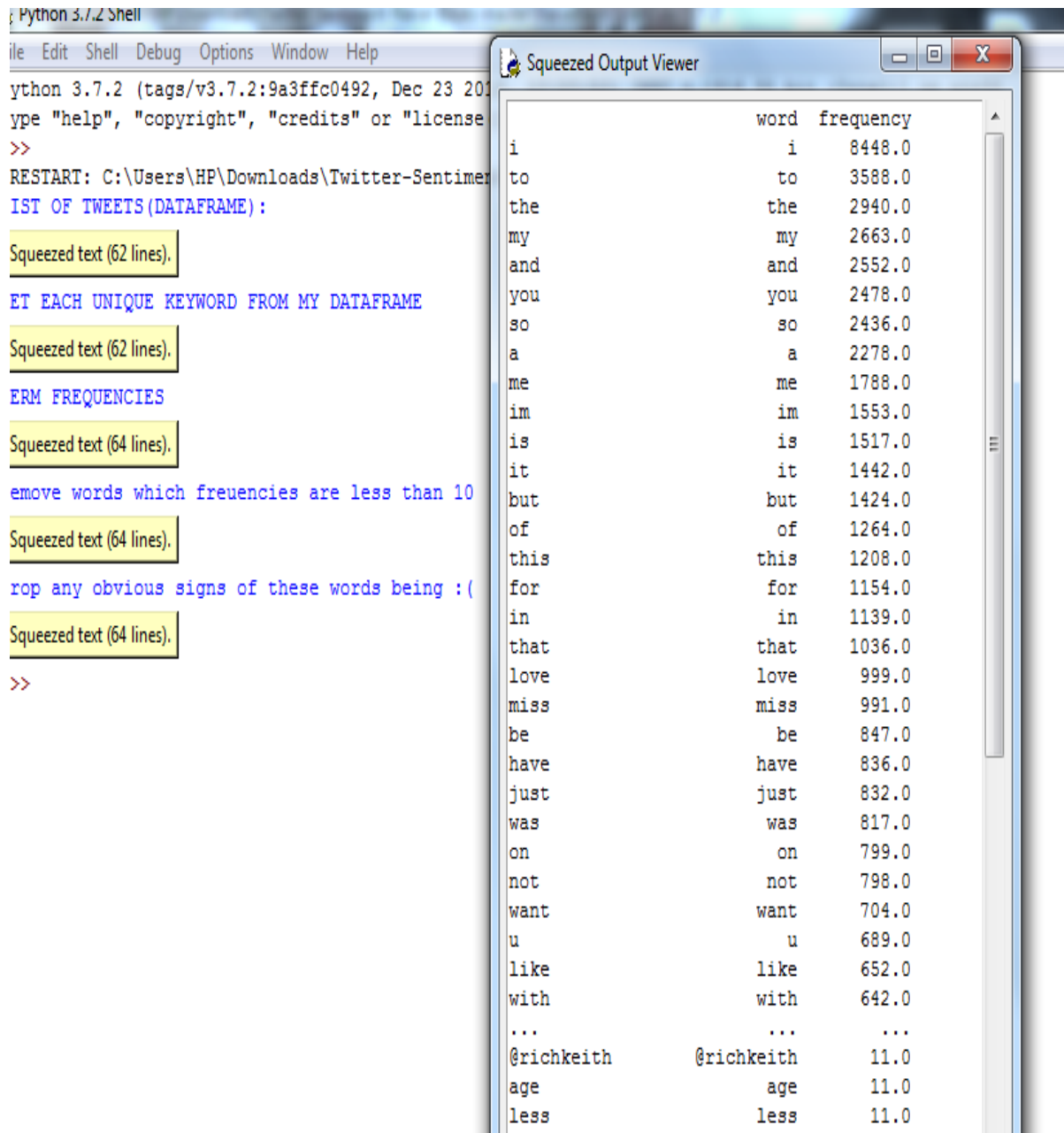
The Squeezed Output Viewer window displays the following word frequency data:

:	14157
:	10581
i	8448
https://t	4142
to	3588
the	2940
my	2663
and	2552
you	2478
so	2436
a	2278
me	1788
im	1553
is	1517
it	1442
but	1424
:	1344
of	1264
this	1208
for	1154
in	1139
that	1036
love	999
miss	991
be	847
have	836
just	832
was	817
:	810
on	799
...	
co/e3xpmfash8	1
@bts_trading	1
co/wfcyqru6mj	1
@adrianachechik	1
psychology	1
@saikrishankumar	1

Fig 9.2 SPLITTED WORDS

9.3 Dropping of special characters

It will remove the word without any meaning that is “nan”, some special characters such as :-), :-(, !!!, # etc...



```
Python 3.7.2 Shell
File Edit Shell Debug Options Window Help
Python 3.7.2 (tags/v3.7.2:9a3ffc0492, Dec 23 2018)
Type "help", "copyright", "credits" or "license()"
>>
RESTART: C:\Users\HP\Downloads\Twitter-Sentiment-Analysis\Twitter-Sentiment-Analysis
LIST OF TWEETS (DATAFRAME):
Squeezed text (62 lines).
ET EACH UNIQUE KEYWORD FROM MY DATAFRAME
Squeezed text (62 lines).
TERM FREQUENCIES
Squeezed text (64 lines).
remove words which frequencies are less than 10
Squeezed text (64 lines).
drop any obvious signs of these words being :(
Squeezed text (64 lines).
>>
```

word	frequency
i	8448.0
to	3588.0
the	2940.0
my	2663.0
and	2552.0
you	2478.0
so	2436.0
a	2278.0
me	1788.0
im	1553.0
is	1517.0
it	1442.0
but	1424.0
of	1264.0
this	1208.0
for	1154.0
in	1139.0
that	1036.0
love	999.0
miss	991.0
be	847.0
have	836.0
just	832.0
was	817.0
on	799.0
not	798.0
want	704.0
u	689.0
like	652.0
with	642.0
...	...
@richkeith	@richkeith 11.0
age	age 11.0
less	less 11.0
...	...

Fig 9.3 DROPPING SPECIAL CHARACTERS

9.4 TERM FREQUENCY

For each dataframes of words, the frequency is calculated by iterating through all of the tweets in it.

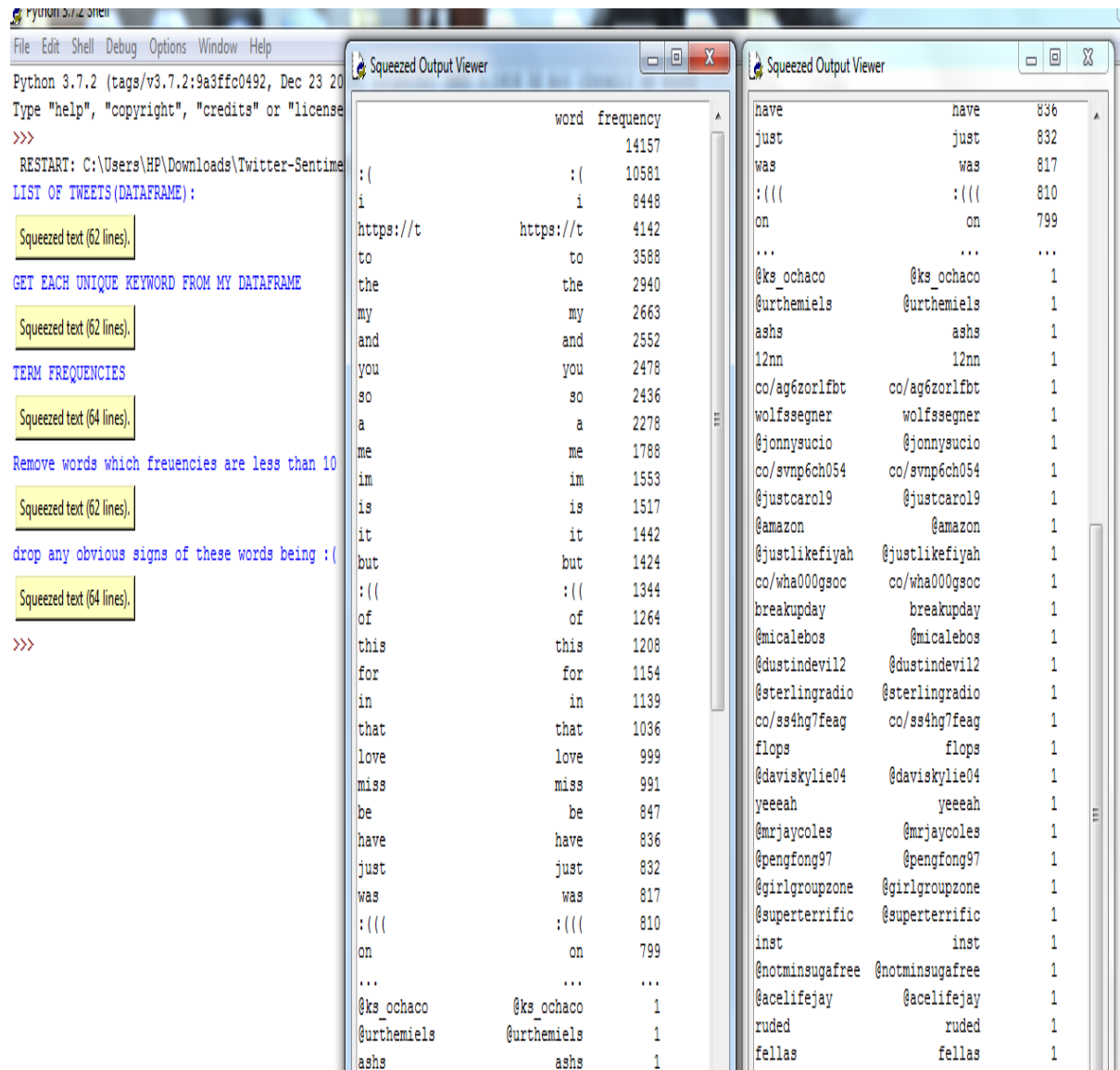
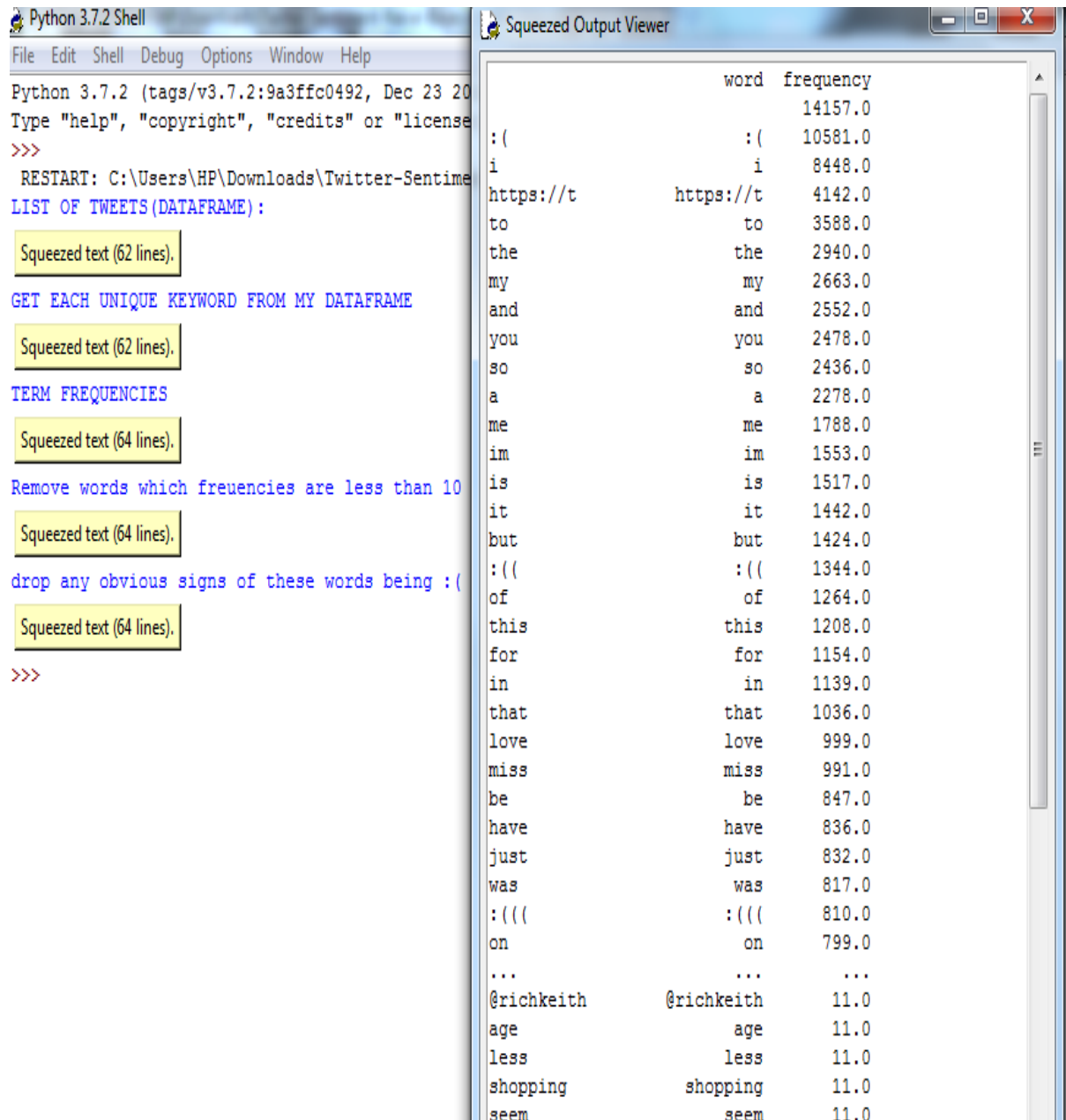


Fig 9.4 TERM FREQUENCIES

9.5 Remove the words with the Frequency lesser than 10

It will remove the words from the bag whose Frequency is lesser than 10 and it will store the words with the frequency greater than 10 in a new csv file.



The screenshot shows a Python 3.7.2 Shell window on the left and a Squeezed Output Viewer window on the right. The shell window contains the following code and output:

```
Python 3.7.2 (tags/v3.7.2:9a3ffc0492, Dec 23 2018)
Type "help", "copyright", "credits" or "license()"
>>>
RESTART: C:\Users\HP\Downloads\Twitter-Sentimental-analysis\Twitter-Sentimental-analysis
LIST OF TWEETS (DATAFRAME):
Squeezed text (62 lines).
GET EACH UNIQUE KEYWORD FROM MY DATAFRAME
Squeezed text (62 lines).
TERM FREQUENCIES
Squeezed text (64 lines).
Remove words which frequencies are less than 10
Squeezed text (64 lines).
drop any obvious signs of these words being :(
Squeezed text (64 lines).
>>>
```

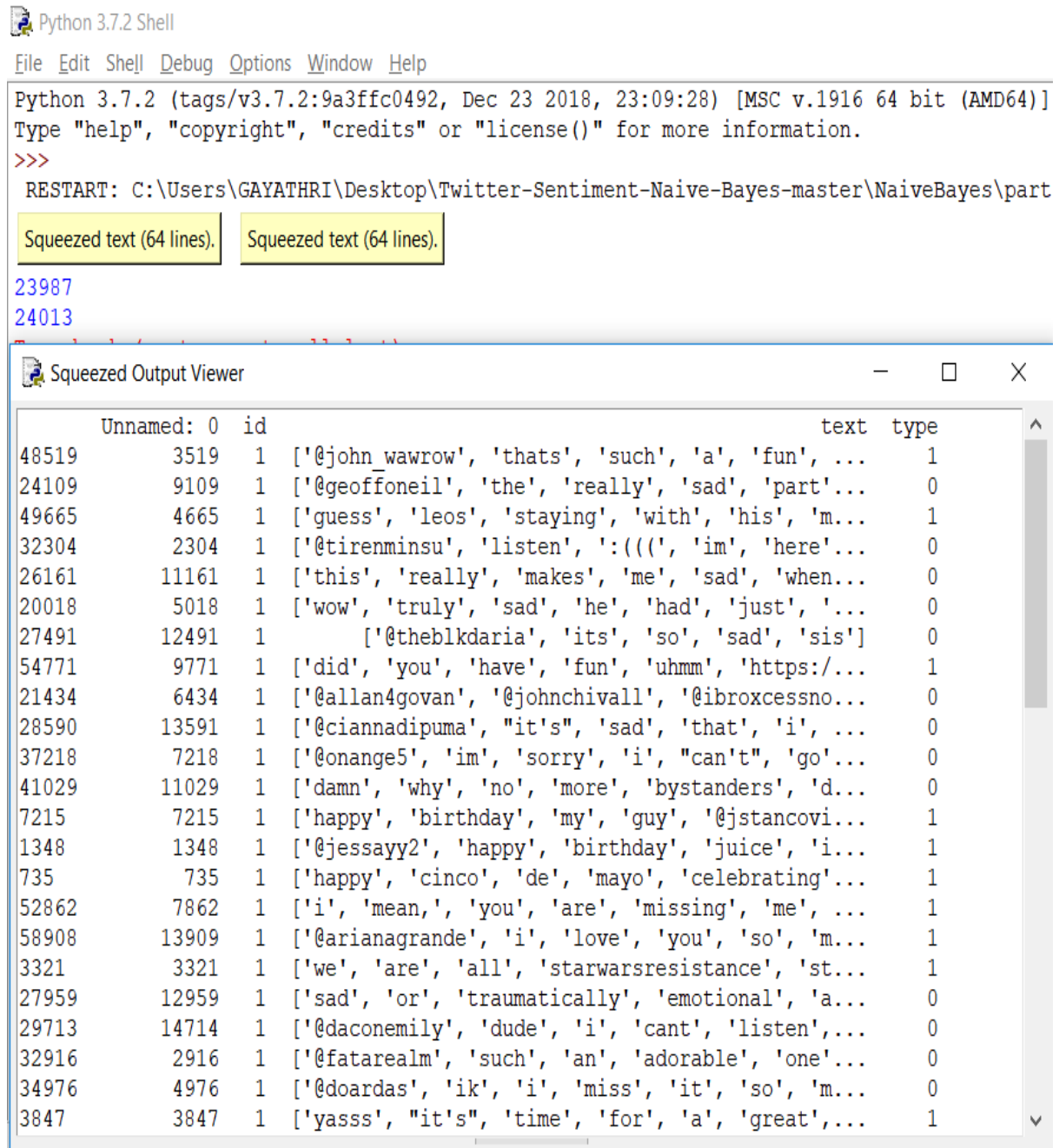
The Squeezed Output Viewer window displays a table of word frequencies:

word	frequency
:	14157.0
:(10581.0
i	8448.0
https://t	4142.0
to	3588.0
the	2940.0
my	2663.0
and	2552.0
you	2478.0
so	2436.0
a	2278.0
me	1788.0
im	1553.0
is	1517.0
it	1442.0
but	1424.0
:((1344.0
of	1264.0
this	1208.0
for	1154.0
in	1139.0
that	1036.0
love	999.0
miss	991.0
be	847.0
have	836.0
just	832.0
was	817.0
:(((810.0
on	799.0
...	...
@richkeith	11.0
age	11.0
less	11.0
shopping	11.0
seem	11.0

Fig 9.5 FREQUENCIES GREATER THAN 10

9.6 Finding the Positive instance and Negative instance

By finding positive instance and negative instance, the type 0 represent Negative words and type 1 represents the Positive Words.



The image shows a Python 3.7.2 Shell window and a Squeezed Output Viewer window. The shell window displays the command prompt and the execution of a script. The output viewer shows a table of sentiment analysis results.

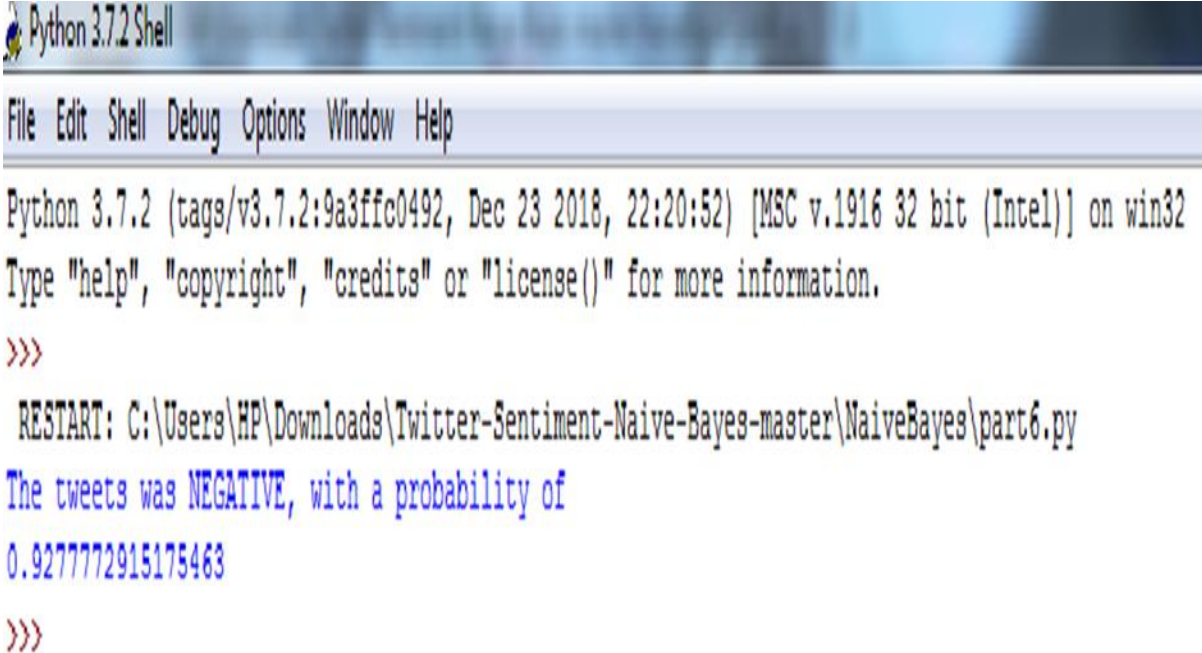
```
Python 3.7.2 (tags/v3.7.2:9a3ffc0492, Dec 23 2018, 23:09:28) [MSC v.1916 64 bit (AMD64)]
Type "help", "copyright", "credits" or "license()" for more information.
>>>
RESTART: C:\Users\GAYATHRI\Desktop\Twitter-Sentiment-Naive-Bayes-master\NaiveBayes\part
Squeezed text (64 lines). Squeezed text (64 lines).
23987
24013
```

Unnamed: 0	id	text	type
48519	3519	1 ['@john_wawrow', 'thats', 'such', 'a', 'fun', ...	1
24109	9109	1 ['@geoffoneil', 'the', 'really', 'sad', 'part'...	0
49665	4665	1 ['guess', 'leos', 'staying', 'with', 'his', 'm...	1
32304	2304	1 ['@tirenminsu', 'listen', ':((', 'im', 'here'...	0
26161	11161	1 ['this', 'really', 'makes', 'me', 'sad', 'when...	0
20018	5018	1 ['wow', 'truly', 'sad', 'he', 'had', 'just', '...	0
27491	12491	1 ['@theblkdaria', 'its', 'so', 'sad', 'sis']	0
54771	9771	1 ['did', 'you', 'have', 'fun', 'uhmm', 'https:/...	1
21434	6434	1 ['@allan4govon', '@johnchivall', '@ibroxcessno...	0
28590	13591	1 ['@ciannadipuma', "it's", 'sad', 'that', 'i', ...	0
37218	7218	1 ['@onange5', 'im', 'sorry', 'i', "can't", 'go'...	0
41029	11029	1 ['damn', 'why', 'no', 'more', 'bystanders', 'd...	0
7215	7215	1 ['happy', 'birthday', 'my', 'guy', '@jstancovi...	1
1348	1348	1 ['@jessay2', 'happy', 'birthday', 'juice', 'i...	1
735	735	1 ['happy', 'cinco', 'de', 'mayo', 'celebrating'...	1
52862	7862	1 ['i', 'mean', 'you', 'are', 'missing', 'me', ...	1
58908	13909	1 ['@arianagrande', 'i', 'love', 'you', 'so', 'm...	1
3321	3321	1 ['we', 'are', 'all', 'starwarsresistance', 'st...	1
27959	12959	1 ['sad', 'or', 'traumatically', 'emotional', 'a...	0
29713	14714	1 ['@daconemily', 'dude', 'i', 'cant', 'listen',...	0
32916	2916	1 ['@fatarealm', 'such', 'an', 'adorable', 'one'...	0
34976	4976	1 ['@doardas', 'ik', 'i', 'miss', 'it', 'so', 'm...	0
3847	3847	1 ['yasss', "it's", 'time', 'for', 'a', 'great',...	1

Fig 9.6 Positive and Negative Instances

9.7 Sentimental Analysis

By using the naïve bayesit find the probability of Negative tweets and positive tweets. We can predict that the tweets were NEGATIVE with the probability of 0.9277772915175463.

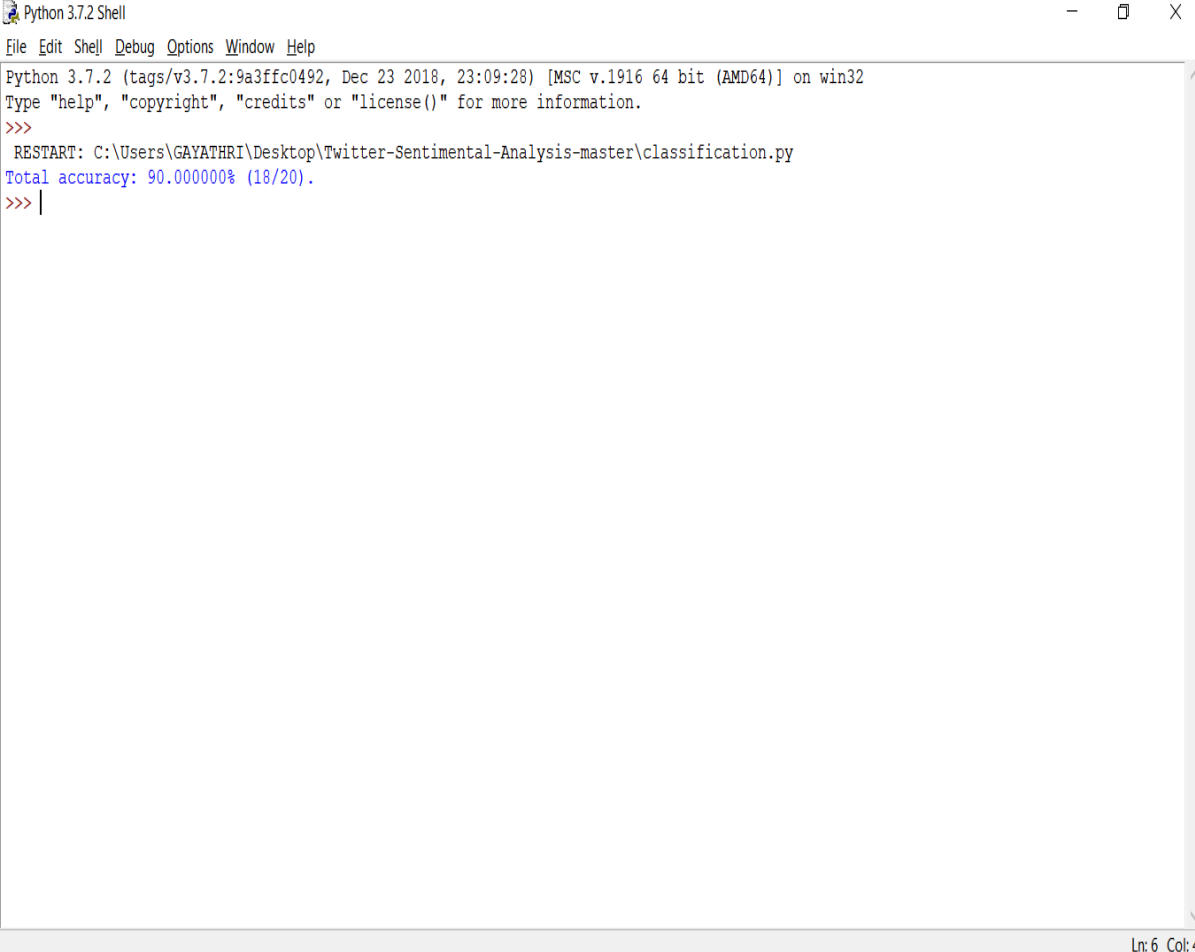


```
Python 3.7.2 Shell
File Edit Shell Debug Options Window Help
Python 3.7.2 (tags/v3.7.2:9a3ffc0492, Dec 23 2018, 22:20:52) [MSC v.1916 32 bit (Intel)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
>>>
RESTART: C:\Users\HP\Downloads\Twitter-Sentiment-Naive-Bayes-master\NaiveBayes\part6.py
The tweets was NEGATIVE, with a probability of
0.9277772915175463
>>>
```

Fig 9.7 SENTIMENT ANALYSIS

9.8 Finding the Accuracy

By using the Training data and the Test data the accuracy is calculated. The training data which consist of unique tweet id and then positive tweets and the negative tweets in the separate training data set.



```
Python 3.7.2 Shell
File Edit Shell Debug Options Window Help
Python 3.7.2 (tags/v3.7.2:9a3ffc0492, Dec 23 2018, 23:09:28) [MSC v.1916 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
>>>
RESTART: C:\Users\GAYATHRI\Desktop\Twitter-Sentimental-Analysis-master\classification.py
Total accuracy: 90.000000% (18/20).
>>> |
```

The screenshot shows a Python 3.7.2 Shell window. The title bar reads "Python 3.7.2 Shell". The menu bar includes "File", "Edit", "Shell", "Debug", "Options", "Window", and "Help". The main text area displays the following content: "Python 3.7.2 (tags/v3.7.2:9a3ffc0492, Dec 23 2018, 23:09:28) [MSC v.1916 64 bit (AMD64)] on win32", "Type 'help', 'copyright', 'credits' or 'license()' for more information.", a prompt ">>>", the command "RESTART: C:\Users\GAYATHRI\Desktop\Twitter-Sentimental-Analysis-master\classification.py", the output "Total accuracy: 90.000000% (18/20).", and another prompt ">>> |". The status bar at the bottom right indicates "Ln: 6 Col: 4".

Fig 9.8 finding the accuracy

CHAPTER 10

CONCLUSION AND FUTURE WORK

10.1 FUTURE WORK

In this project, we tried to show the basic way of classifying tweets into positive or negative category using Naive Bayes as baseline. We could further improve our classifier by trying to extract more features from the tweets, trying different kinds of features, tuning the parameters of the naïve Bayes classifier, or trying another classifier all together. In the future work, we plan to utilize the extracted information from applicants' resumes to dynamically generate user profiles to be further used for recommending jobs to job seekers.

10.2 CONCLUSION

A resume won't mean what we're all familiar with: a brief written account of your personal, educational, and professional experience that's prepared by you, the applicant. Instead, recruiters are quickly adapting to **social recruiting** practices—thereby acquiring information about you by examining your social footprints. This social data about you will eventually be acquired and curated automatically throughout the web. Recruiters and hiring managers will use intuitive search software to navigate millions of distinguished candidates instead of seeing individual social media sources personally. However, recruiters are currently overusing networks such as Facebook, LinkedIn, and Twitter many recruiters are still learning how to target their social media audiences. The analytics of these studies will be tested and retested in time, and eventually automated. Knowing just a little bit about the science behind these methods goes a long way in helping you visualize the future of recruitment.

Twitter allows you to place your 140 characters before the recruiter, hiring manager and CEO who make a large percentage of hiring decisions. When used the right way, Twitter can play an important role in the recruiting and screening phases of your hiring process. While Facebook is usually used for personal relationships and LinkedIn for professional networking, Twitter may be somewhat of both. After an interview, the smartest managers look at Twitter to validate their thoughts. Take a look at potential a candidate's profile, it can give some insight into his or her character. For example, if the person posts inappropriate jokes or pictures, poor judgment may be shown. **"You'll get a feel for their personality, and it can help you decide if the person is a good fit for your company."**

Through this project we provide user's personality which is predicted from text written on Twitter. Users' degree of reflection between the real world and the Twitterworld was measured by the users' way of developing friendship. Finally, it provides the degree of reflection and the probability of the polarity.

CHAPTER 11

REFERENCES

- [1] K. Shin and J. Lee, "A job applicants' résumé verification method using a social network analysis Using Facebook like as LinkedIn for a recruiting," 2017 11th International Conference on Software, Knowledge, Information Management and Applications (SKIMA), Malabe, 2017, pp. 1-5.[doi:10.1109/skima.2017.8294124](https://doi.org/10.1109/skima.2017.8294124)
- [2] B. Y. Pratama and R. Sarno, "Personality classification based on Twitter text using Naive Bayes, KNN and SVM," 2015 International Conference on Data and Software Engineering (ICoDSE), Yogyakarta, 2015, pp. 170-174.
- [3] A. Zaroor, M. Maree and M. Sabha, "JRC: A Job Post and Resume Classification System for Online Recruitment," 2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI), Boston, MA, 2017, pp. 780-787.
- [4] A. Teixeira and R. M. S. Laureano, "Data extraction and preparation to perform a sentiment analysis using open source tools: The example of a Facebook fashion brand page," 2017 12th Iberian Conference on Information Systems and Technologies (CISTI), Lisbon, 2017, pp. 1-6.
- [5] R. Jiamthapthaksin and T. H. Aung, "User preferences profiling based on user behaviors on Facebook page categories," 2017 9th International Conference on Knowledge and Smart Technology (KST), Chonburi, 2017, pp. 248-253.
- [6] Cantador, Iván & Fernandez-Tobias, I &Bellogín, Alejandro & Kosinski, Michal & Stillwell, David. (2013). Relating Personality Types with User Preferences Multiple Entertainment Domains. CEUR Workshop Proceedings. 997.
- [7] H. Parveen and S. Pandey, "Sentiment analysis on Twitter Data-set using Naive Bayes algorithm," 2016 2nd International Conference on Applied and

Theoretical Computing and Communication Technology (iCATccT), Bangalore, 2016, pp. 416-419.

[8] M. Kumar and A. Bala, "Analyzing Twitter sentiments through big data," *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, New Delhi, 2016, pp. 2628-2631.

[9] Sumner, C., Byers, A., Boochever, R., & Park, G. J. (2012). Predicting Dark Triad Personality Traits from Twitter Usage and a Linguistic Analysis of Tweets. 2012 11th International Conference on Machine Learning and Applications. doi:10.1109/icmla.2012.218

[10] Alireza Souri, Shafigheh Hosseinpour and Amir Masoud Rahmani, "Personality classification based on profiles of social networks' users and the five-factor model of personality", 2018 Human-centric Computing and Information Sciences.(Springer)

[11] Xiao-Feng Zhong, Shi-Ze Guo, Liang Gao, Hong Shan and Di Xue, "A General Personality Prediction Framework Based on Facebook Profiles", ICMLC 2018 Proceedings of the 2018 10th International Conference on Machine Learning and Computing, pages 269-275.

[12] Gabriel Yakub N.N. Adi, Michael Harley Tandio, Veronica Ong, and Derwin Suhartono, "Optimization for Automatic Personality Recognition on Twitter in Bahasa Indonesia", 2018 3rd International Conference on Computer Science and Computational Intelligence, Procedia Computer Science, Volume 135, 2018, Pages 473-480

[13] Tadesse, M. M., Lin, H., Xu, B., & Yang, L. (2018). Personality Predictions Based on User Behaviour on the Facebook Social Media Platform. IEEE Access, 1–1. doi:10.1109/access.2018.2876502

[14] “Why Twitter Might Be Your Best Recruiting Tool” by Stephanie Vozza through an article in ‘Entrepreneur India’.

(<https://www.entrepreneur.com/article/226960>)

[15] “5 Personality Traits: Using Twitter to Evaluate Candidates”, a blog written by Jordan Taylor in ‘Entelo blog’.(<https://blog.entelo.com/5-personality-traits-using-twitter-to-a-evaluate-candidates>)

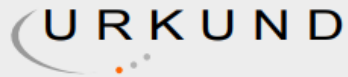
[16] “Get Hired from a Tweet? I Think So”, by David Smooke in Hiring Success Journal. (<https://www.smartrecruiters.com/blog/get-hired-from-a-tweet-i-think-so/>)

[17] Twitter Analytics Using R by Ranjan Baitha in Academia.edu.

(https://www.academia.edu/15083048/Twitter_Analytics_Using_R)

CHAPTER 12

PLAGIARISM



Urkund Analysis Result

Analysed Document:	reportcontent (2).docx (D51220465)
Submitted:	4/29/2019 7:40:00 AM
Submitted By:	srjan1998@gmail.com
Significance:	6 %

Sources included in the report:

https://www.onlineprivacyfoundation.org/research_/Sumner_Predicting_Dark_Triad_Traits_from_Twitter_Usage_V5.pdf
<https://hcis-journal.springeropen.com/articles/10.1186/s13673-018-0147-4>
<https://www.hindawi.com/journals/jcnc/2016/8069672/citations/>
<https://dblp.uni-trier.de/db/conf/icmlc2/icmlc2018>
<http://derwinsuhartono.com/research-activities>

Instances where selected sources appear:

9