VISVESVARAYA TECHNOLOGICAL UNIVERSITY

"Jnana Sangama", Belgavi-590 018, Karnataka, India

An Internship Report
On

"MACHINE LEARNING WITH PYTHON"

Submitted in Partial Fulfillment of the requirement for the award of the degree of

BACHELOR OF ENGINEERING
IN
COMPUTER SCIENCE AND ENGINEERING

Submitted By

Name: MANJESH KUMAR K P
USN : 1SJ18CS127

Under the guidance of

Internal Guide
Prof. Srinath G M
Assistant Professor
Dept. Of CSE, SJCIT

External Guide
Mr. Kiran
Product Manager
Compsoft Technologies

S J C INSTITUTE OF TECHNOLOGY
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
CHIKKABALLAPUR-562101

2021-2022

||Jai Sri Gurudev||

Sri Adichunchanagiri Shikshana Trust®

S.J.C INSTITUTE OF TECHNOLOGY, Chickballapur - 562101

Department of Computer Science and Engineering

CERTIFICATE

This is to certify that the Internship work entitled "MACHINE LEARNING WITH PYTHON" carried out by MANJESH KUMAR K P bearing USN:1SJ18CS127 a bonafide student of Sri Jagadguru Chandrashekaranatha Institute of Technology in partial fulfilment for the award of **Bachelor of Engineering in Computer Science and Engineering** of **Visvesvaraya Technological University, Belgaum** during the year **2021-22**. It is certificated that all corrections / suggestions indicated for internal assessment have been incorporated in the report deposited in the departmental library. The Internship report has been approved as it satisfies the academic requirements in respect of Internship work prescribed for the said Degree.

Signature of Guide	Signature of HOD	Signature of Principal
Prof. Srinath G M	**Dr. Manjunath Kumar B H**	**Dr. G T Raju**
Assistant Professor	Professor & HOD,	Principal, SJCIT,
Dept. of CSE,SJCIT	Dept. of CSE,SJCIT	Chickballapur

External Examiners:

Name of the Examiners	Signature with Date
1. Dr Murthy SVN	23/2/22
2. RASHMI K. A	23/2/2022

COMPANY CERTIFICATE

Compsoft Technologies

Providing a Complete Suite of IT Solutions

Certificate ID - 1CST21SA023

Date - 24/10/2021

Certificate of Internship

This is to Certify that **Mr Manjesh Kumar K P (1SJ18CS127)** Student of "Visvesvaraya Technological University" has completed a one-month Internship on **"Machine Learning with Python(Research-Based)"** with a grade of "A" in our Company from 5th September 2021 to 10th October 2021.

He was very much interested to be a part of our research on **Sentimental Analysis**. During the internship, He demonstrated good leadership skills with a self -motivated attitude to learn new things. His performance exceeded expectations and was able to complete the project successfully on time, We wish him all the best in his future endeavours.

Warm regards,

Dhanush S

(Project Manager, CST)

Compsoft Technologies
No. 363, 19th main road,
1st Block Rajajinagar Bangalore- 560010

www.compstechnologies.com

services@compstechnologies.com

*Search Engine Optimisation *Development *ML & Research
*Branding and Design *Content Writing *Embedded Systems and IOT

DECLARATION

I, **MANJESH KUMAR K P**, student of VIII semester B.E in Computer science & Engineering at S J C Institute of Technology, Chickballapur, hereby declare that the Internship work entitled "MACHINE LEARNING WITH PYTHON" has been independently carried out by me under the supervision of **Prof. Srinath G M,** Assistant Professor, Department of CSE , and the coordinator **Narendra Babu C** Assistant Professor, submitted in partial fulfillment of the course requirement for the award of degree in **Bachelor of Engineering** in **Computer Science & Engineering** of **Visveswaraya Technological University, Belgavi** during the year 2021-2022. I further declare that the report has not been submitted to any other University for the award of any other degree.

PLACE:CHICKBALLAPUR
Date:13/05/2022

NAME :MANJESH KUMAR K P
USN: 1SJ18CS127

ABSTRACT

Fundamentally, a sentiment refers to the reflection of emotions of people. Today's world stands on the strings of emotions. People express happiness, sadness, love, hatred etc. through some actions. Division of emotions i.e positive, neutral and negative, is called sentiment analysis. Nowadays there is a sentiment rich data in the form of tweets, status updates, blog posts, reviews, comments, forums for discussion etc. If we efficiently work upon this bucket full of sentiment rich data, it gives way in apprehending the opinions, views or perspective of masses in a specific functional area. Moreover, the result of this analysis will aid people in taking suitable actions or corrective measures for their growth. This effort of ours is like a drop in the ocean to try to analyze the reviews posted by people at four different websites (airlinequality.com, Amazon, Yelp, and IMDB). Further, the reviews are processed and analyzed using machine learning procedures, algorithms and other related aspects. Finally, the conclusion is derived by finding the polarity of a particular review whether it is poor, average or excellent for Airlines dataset and 0 or 1 for the other three datasets. The entire task was performed using Python.

ACKNOWLEDGEMENT

With reverential pranam, we express my sincere gratitude and salutations to the feet of his holiness **Byravaikya Padmabhushana Sri Sri Sri Dr. Balagangadharanatha Maha Swamiji,** & his holiness **Jagadguru Sri Sri Sri Dr. Nirmalanandanatha Swamiji** of Sri Adichunchanagiri Mutt for their unlimited blessings. First and foremost we wish to express my deep sincere feelings of gratitude to our institution, **Sri Jagadguru Chandrashekaranatha Swamiji Institute of Technology.** For providing me an opportunities for completing my internship work successfully.

I extend deep sense of sincere gratitude to **Dr. G T Raju, Principal, S J C Institute of Technology, Chickballapur**, for providing an opportunity to complete the Internship Work.

I extend special in-depth, heartfelt, and sincere gratitude to our HOD **Dr. Manjunath Kumar , Professor and Head of the Department, Computer Science and Engineering, S J C Institute of Technology, Chickballapur,** for his constant support and valuable guidance of the Internship Work.

I convey our sincere thanks to Internship Internal Guide **Girish B G, Assistant Professor, Department of Computer Science and Engineering, S J C Institute of Technology,** for his constant support, valuable guidance and suggestions of the Internship Work.

I am thankful to Internship External Guide **Dhanush S , Product Manager, Compsoft Technologies, Banglore,** for providing valuable guidance and encouragement of the Internship Work.

I also feel immense pleasure to express deep and profound gratitude to our Internship Coordinator **Narendra Babu , Assistant Professor, Department of Computer Science and Engineering, S J C Institute of Technology,** for his guidance and suggestions of the Internship Work.

Finally, I would like to thank all faculty members of Department of Computer Science and Engineering, S J C Institute of Technology, Chickballapur for their support.

I also thank all those who extended their support and co-operation while bringing out this Internship Report.

MANJESH KUMAR K P
(1SJ18CS127)

CONTENTS

LIST OF FIGURES

CHAPTER - 1

COMPANY PROFILE

1.1 History of the Organization

Compsoft Technologies is founded by a group of Enterprise Architects having over two decades of experience in software architecture, design and development of missioncritical systems for some of the Fortune 500 companies. We hire some of the best talents in the market to deliver quality software on your aggressive milestone dates. Being in theindustry and having hands on experience, we fully understand the entire Software Development Life Cycle and we only hire resources who meet our high standards. All of our resources go through our rigorous interview process based on your requirements and we only select the candidates who not only technically strong but also they are fully dedicated to deliver on your promise, the success of your organization.

Our resources are expert in designing and developing applications using Agile and Scrum methodologies. Whatever your software development methodologies may be, our resources have experience in broad areas and they can pull any project successfully.

We work hard to enhance continuously our reputation for accessibility, professionalism, performance, and the depth and quality of our long-term consultative relationships with our clients. We endeavor to be valued as an industry leader in client satisfaction, quality performance and reputation. All activities will be conducted to the highest ethical and professional standards.

To help our clients achieve their objectives by serving as their manpower consulting firm. Compsoft Technologies has one-to-one relationships with a number of clients, helping them benefit from all of the technologies available to them and build a better solution that exceeds client's expectation. It is our goal that offers a full range ofsoftware, consulting, support, automation combined with a wide range of technologies that enable clients to consider how they could achieve their objectives.

1.1.1 Objectives

We are committed to going the extra mile to bring success to the clients consistently. We are dedicated to delivering the right people, solutions, and services to the clients that theyrequire to meet their technology challenges and business goals.

Delivering the most efficient and the best solution to our clients to every client leveraging leading technologies & industry best practices.

1.1.2 Operation of the Organization

The race for digital transformation is on. In this globally connected on-demand world with rapid advancements in internet technologies, businesses worldwide are under constant pressure to add innovative real-time capabilities to their applications to respond to market opportunities.

Every business worldwide is building event-driven, real-time applications – from financial services, transportation, and energy, to retail, healthcare, and Gaming companies. Our endeavor is to make it easy to develop innovative real-time applications and efficient to operate them in production

We have a proven record of building highly scalable, world-class consulting processes that offer tremendous business advantages to our clients in the form of huge cost-benefits, definitive results and consistent project deliveries across the globe.

We prominently strive to improve your business by delivering the full range of competencies including operational performance, developing and applying business strategies to improve financial reports, defining strategic goals and measure and manage those goals along with measuring and managing them.

1.1 Major Milestones

Compsoft Technologies is an Information Technology manpower consulting and product development firm specialized in bringing businesses and technologies together under the same umbrella. Our in-depth knowledge in developing mission critical systems for Fortune 500 companies has earned us the reputations in the marketplace. Whether you are buildingcomplex systems for your customers or migrating your legacy systems to cutting edge technologies, our resources are fully trained and equipped with the knowledge required toperform the job right, from the very first day. We serve on a wide variety of our clients including Banking, Accounting, Insurance, Healthcare, Retail, Trucking & Transportation. We have proven record of evaluating the best candidates for your requirement and stand by on the quality throughout the project implementation.

In today's competitive marketplace, it is important to bring the businesses and technologies together to deliver on your promise. More than ever, Compsoft Technologies is committed to deliver on our promise so that you can deliver on yours, the success of your organization.

1.1 Structure of the Organization

CST, core strength lies in our super energetic and gigantic team, forming an excellent blend of IT minds along with a creative bent. Taking ownership of not only one's own task but also creating an enduring & contented customer is what every individual works together with. Our endeavour is to continuously improve and deliver maximum– enriching products & solutions. That's what we mean – doing IT better driving business transformation, digitally – Applying disruptive technologies for business transformations. CST brings great advantage to business space by bridging gaps, simplifying businesses and elevating competitive benefits by providing technology-based business solutions.

Leveraging the enormous talent of our passionate and proven individuals. We are hugely a customer-centric organization that is bent upon consummating the needs of the customers beyond their expectations. We successfully host a consortium of experienced

professionals who work in synergy in order to gain an edge over the market. we look at ourselves as a team where we co-create with them.

Having delivered successful projects we pride ourselves on being a sought-after mobile application development. Through the years, and have been successfully deliveringvalue to our customers. We truly believe that our customer's success is our success We don't look at ourselves as a vendor for their projects instead. You would be excited to hear some of our stories and know to what extent we have gone in the interest of the success of our customers. and we work hard to make that happen. This philosophy and execution have resulted in a long-term partnership with most of our clients.

1.1 Services Offered

It is believed that service and quality is the key to success, Enabling business success driven by technology. Harnessing the power of technology, we create a measurable difference for our clients across various industries & multiple geographies.

1 Development.

We develop responsive, functional and super-fast websites. We keep User Experience in mind while creating websites. A website should load quickly and should be accessible even on a small view-port and slow internet connection.

2 Branding and Design.

We offer professional Graphic design, Brochure design & Logo design. We are experts in crafting visual content to convey the right message to the customers. We also design custom wraps for your products(also known as package designing).

3 Search Engine Optimization

We help you manage your SEO campaign more efficiently and effectively. We help you gain market share by leveraging our expertise. our holistic approach to identify anything that may be hurting your traffic or rankings and show you just how to outrank the competition.

4 Content Writing

We provide content writing services for blogs and product descriptions, our team helps you generate content to Increase your Brand Recall. We can amplify your marketing needs & help you reach your potential customers.

5 Research

We equip business leaders with indispensable insights, advice and tools to achieve their goals, our main area of research is in sentimental analysis, having published multiple papers on the same, we are in the process of creating a virtual bot that is intended to use our sentimental analysis data to provide real time replies.

6 Embedded System and IOT

CST works with Consumer Electronics, Lighting, Home Automation, Metering, Sensor-Technology, Home Appliance and Medical Device companies to help them create smart and connected products.

Through its integrated Embedded and IoT services, Techno soft helps build intelligent & connected devices that can be remotely monitored and controlled while leveraging edge and cloud computing for a host of intelligent applications and analytics.

7 Machine Learning

Machine learning is one of the most essential tools for data science, and allows companies, organizations, and governments to harness the power of big data. When it comes to finding work in this field, the most important requisite is hands-on experience.

8 Full Stack Web Development

Full stack web development is the practice of working on both the front-end and back- end of a program. Full Stack is a layer of software or web development consisting of the front-end and the back-end portions of an application. Front-end is what the users will see or interact with on your application. Back-end part is what users do not see, such as application's logic, database, server, etc. A full-stack web developer is comfortable working with both back-end and front-end technologies which make a website or application function properly.

CHAPTER – 2

ABOUT THE DEPARTMENT

2.1 Specific Functionalities of the Department

Our department of tech support majorly focused on manage, maintain and repair IT systems.

The Special functionalities include

- Understanding the work to be completed.
- Planning the assigned activities in more detail if needed
- Completing assigned work within the budget, timeline and quality expectations
- Informing the project manager of issues, scope changes, risk and quality concerns
- Proactively communicating status and managing expectation

2.2 Process Adopted

The department aims to first understand the user requirements. Further on, a basic structure of the product that needs to be built is drawn and understood. Eventually, the technologies that would best help in developing the product are understood. If the product has database requirements, the schema and the database design are worked upon. The department believes in "Think before you code"- the requirements and logics are first understood over a paper and then are moved to a code form. Agile processes generally promote a disciplined project management process that encourages frequent inspection and adaptation, a leadership philosophy that encourages teamwork, self-organization and accountability, a set of engineering best practices intended to allow for rapid delivery of high-quality software, and a business approach that aligns development with customer needs and company goals. Agile development refers to any development process that is aligned with the concepts of the Agile Manifesto. The Manifesto was developed by a group fourteen leading figures in the software industry, and reflects their experience of what approaches do and do not work for software development.

2.3 Testing

Testing was done according to the Corporate Standards. As each component was being built, Unit testing was performed in order to check if the desired functionality is obtained. Each component in turn is tested with multiple test cases to verify if it is properly working. These unit tested components are integrated with the existing built components and then integration testing is performed. Here again, multiple test cases are run to ensure the newly built component runs in co-ordination with the existing components. Unit and Integration testing are iteratively performed until the complete product is built. Once the complete product is built, it is again tested against multiple test cases and all the functionalities.

The product could be working fine in the developer's environment but might not necessarily work well in all other environments that the users could be using. Hence, the product is also tested under multiple environments (Various operating systems and devices). At every step, if a flaw is observed, the component is rebuilt to fix the bugs. This way, testing is done hierarchically and iteratively.

2.4 Structure of the Department

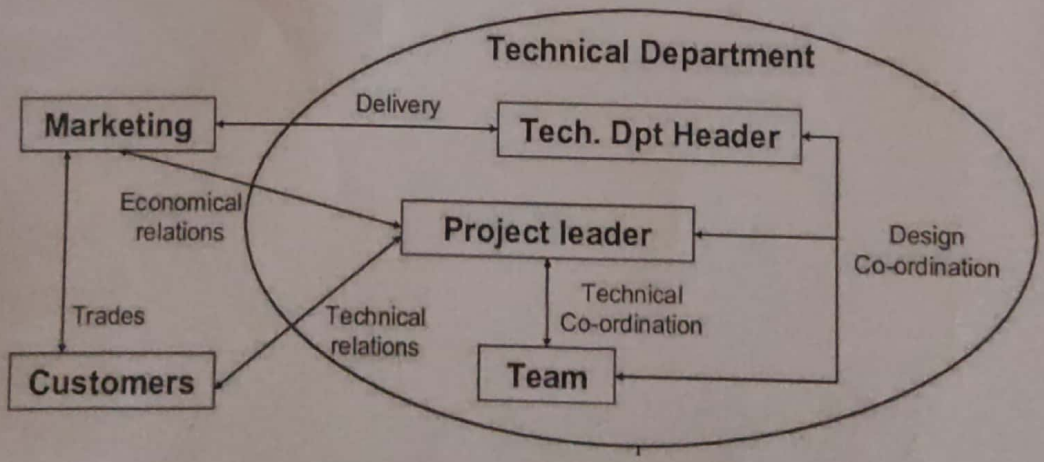

Figure 2.4.1 Structure of the Department

2.5 Roles and Responsibilities of Individuals

Since the internship was online, to ensure easy onboarding of interns, the company had additional individuals who took care of the smooth run of online training.

- Operation and Strategy Head- Ensured there were no difficulties for interns while onboarding. Best of mentors and doubt clarifying sessions were arranged too.
- Technical Lead- Ensured the technicalities of online training to be smooth. Best platforms were arranged for our meetings and trainings.
- Mentors- They have helped us to understand the concepts, gave us tasks to get practical take a ways and clarified doubts to the best.
- Interns- Worked through the tasks given either individually or in a group

CHAPTER – 3

TASK PERFORMED

Internship was on Machine Learning with Python.

Training Program

- Initially trained on the following:
  - Python Programming
  - Machine Learning Algorithms
- Working on various python libraries for implementing project.
- Searching and collection of dataset related to the twitter sentiment analysis.
- Analyzing and understanding features of dataset and its purpose of usage.
- Performing pre-processing of raw data into clean data set.
- Visualize and split data into training and testing dataset.
- Applying ML algorithms to dataset to predict the positive or negative tweet.

CHAPTER – 4

REFLECTION NOTES

4.1 Experience

As per our experience during the internship, Compsoft Technologies follows a good work culture and it has friendly employees, starting from the staff level to the management level. The trainers are well versed in their fields and they treat everyone equally. There is no distinguishing between fresher graduates and corporates and everyone is respected equally. There is a lot of teamwork followed in every task, be it hard or easy and there is a very calm and friendly atmosphere maintained at all times. There is a lot of scope for self-improvement due to the great communication and support that can be found. Interns have been treated and taught well and all our doubts and concerns regarding the training or the companies have been properly answered. All in all, Compsoft Technologies was a great place for a fresher to start career and also for a corporate to boost his/her career. It has been a great experience to be an intern in such a reputed organization.

4.2 Technical Outcomes

4.2.1 System Requirements and Specification

HARDWARE REQUIREMENTS:

- Processor : x86 or x64
- Hard Disk : 500 GB or more.
- Ram : 512 MB(minimum) , 1 GB(recommended)

SOFTWARE REQUIREMENTS:

- Operating System : Windows or Linux
- Development Environment : Anaconda Navigator (Jupiter Notebook or Spyder)

CHAPTER - 5

System Analysis and Design

5.1.1 Existing System

The existing system was developed using Decision tree which showed 63% accuracy and k-nearest neighbors algorithm algorithm showed 38% accuracy.when this accuracy is compared with the proposed sytem that is done by using naïve bayes classifier has shown 94% accuracy

5.1.2 Disadvantages of the Existing System

- The accuracy of desion tree and the k-nearest neighbors algorithm is very less when compared to the naïve bayes algorithm.

5.1.3 Proposed System

The proposed system is developed in Google coolaboratory platform using python to achieve greater accuracy than desion tree and the k-nearest neighbors algorithm and display all the positive and negative words in wordcloud format and also an Count of words is also displayed and confusion matrix with actual values and predicted values displayed in proposed system.

5.1.4 Advantages of the Proposed System

- Naïve bayes classifier is easier to implement, interpret, and very efficient to train.
- It doesn't require as much training data.

CHAPTER – 6

Implementation

6.1 Modules

Data Preprocessing:

Step 1: The first step in the machine learning process is to prepare the data. This includesimporting all the packages that will help us organize and visualize the data

```
import pandas as pd
import numpyas np
import matplotlib.pyplot as mp
```

Step 2: After importing all the necessary packages, we need to load the dataset. We use the help of pandas to load the data set

```
train = pd.read_csv('train.csv')
test = pd.read_csv('test.csv')
```

Step 3: We need to get the positive tweets of dataset. This is done as follows:

```
train[train['label'] == 0].head(10)
```

Step 4: We need to get the negative tweets of dataset. This is done as follows:

```
train[train['label'] == 1].head(10)
```

Step 5: We need to get the positive and negative count of tweets in dataset. This is done as follows:

```
train['label'].value_counts()
```

positive tweets count = 29720

negative tweets count =2242

2. Data Visualization:

The sentimental words are categorized into positive and negative words, visually represented using wordcloud. In this we mainly focus on sentimental words confusion matrix, hashtag counts and ROC Curve used to determine Accuracy using Naïve bayes classifier.

3. Applying ML Algorithm

- Convert your data set to .csv file.
- Import dataset into Jupyter Notebook.
- Get the positive tweets.
- Get the negative tweets.
- Split data into training and testing data .
- Apply the Naïve Bayes Classifier Algorithm.
- calculate accuracy of model.

Python Code

```
import re #for regular expressions
import nltk #for text manipulation
import string
import warnings
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
pos_tweets = nltk.FreqDist(ht_regular)
df1 = pd.DataFrame({'Hashtag': list(pos_tweets.keys()),'Count':list(pos_tweets.values())})

#selecting top 20 most frequent hashtags
df1 = df1.nlargest(columns="Count",n=20)
plt.figure(figsize=(16,5)
```

```
from sklearn.feature_extraction.text
 import TfidfVectorizer, CountVectorizer
 import gensim

#removing words whose length is less than 3

 combine['tidy_tweet'] = combine['tidy_tweet'].apply(lambda x: ' '.join([w for w in x.split() if le
 n(w)>3]))
# to display all words
 all_words = ' '.join([text for text in combine['tidy_tweet']])
 from wordcloud import WordCloud
 wordcloud = WordCloud(width=800,height=500,random_state=21,max_font_size=110).gener
 ate(all_words)
 plt.figure(figsize=(10,7))
 plt.imshow(wordcloud, interpolation="bilinear")
 plt.axis('off')
 plt.show()

 # Naive Bayes Classifier
 from sklearn.naive_bayes import MultinomialNB  # Naive Bayes Classifier
 model_naive = MultinomialNB().fit(X_train, y_train)
 predicted_naive = model_naive.predict(X_test)
```

6.2 : Screen Shots

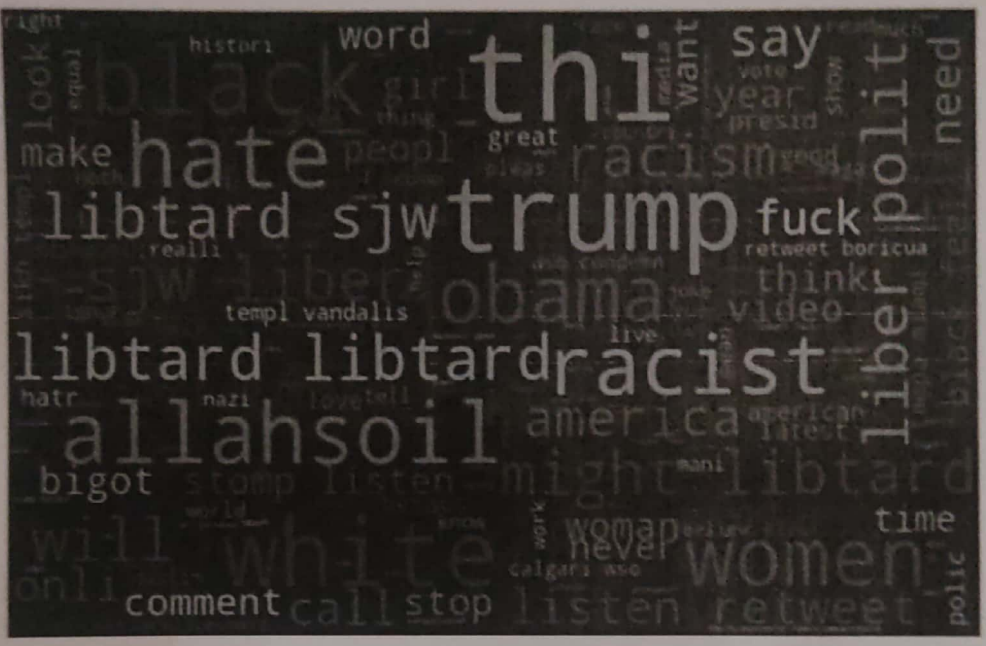

Figure 6.2.1 POSITIVE WORDS

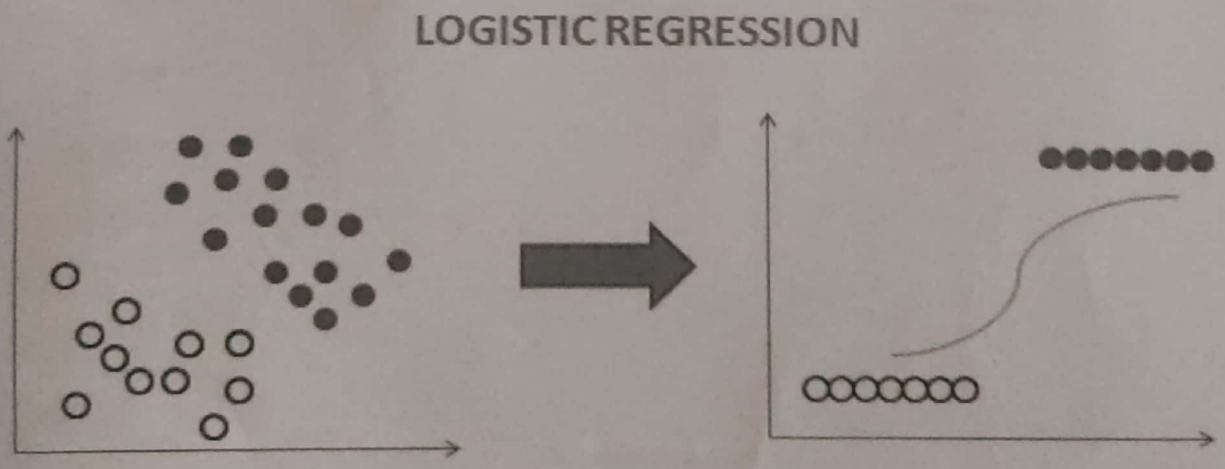

Figure 6.2.2 LOGISTIC REGRESSION

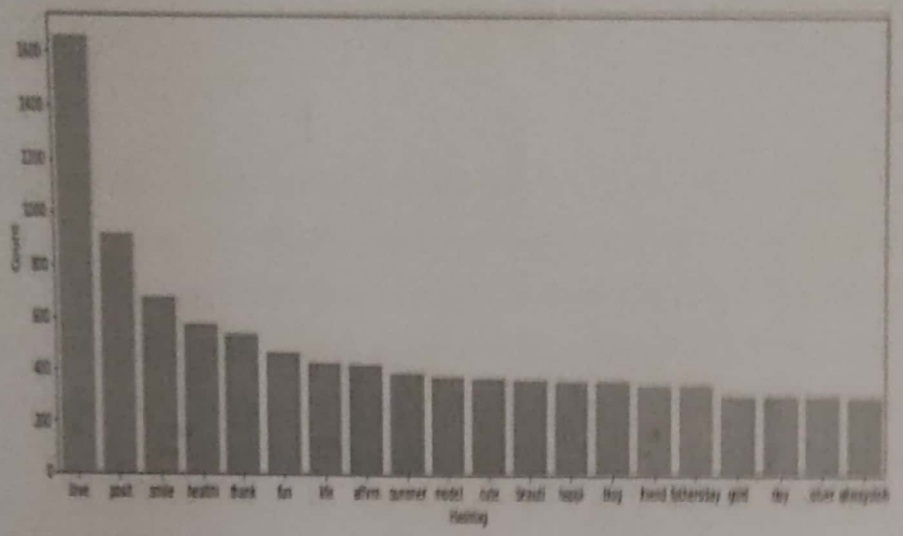

Figure 6.2.3 TOP 20 HASHTAGS WORDS

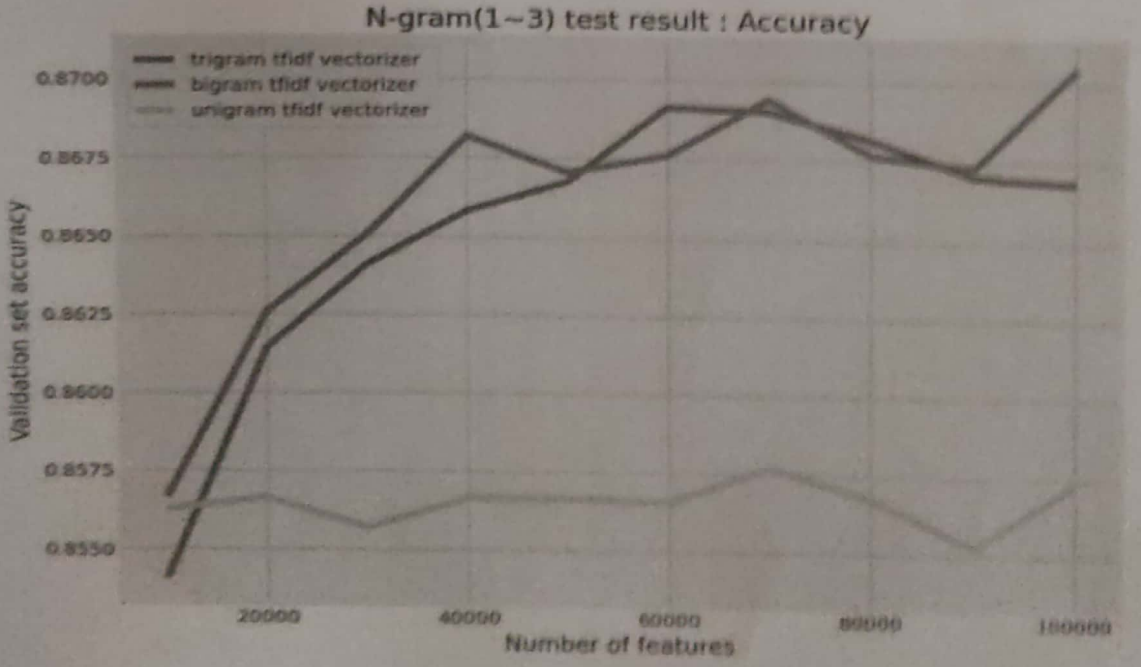

Figure 6.2.4: NUMBER OF FEATURES

CHAPTER – 7

CONCLUSION

Proved the important phenomenon correct that no algorithm can be ranked as best for sentiment analysis because these algorithms are domain specific. Some would work best for one type of data set and others would work best for some other type. Another important factor discovered was that the analysis of tweets is comparatively easier as compared to reviews due to the constraint of 140 character imposed by Twitter. The user has to express his/her entire emotion within these 140 characters whereas in reviews the user has one full comment section which can contain multiple paragraphs too. And it is quite obvious that analysis for sentiments in short texts would be simpler than paragraphs. This too is the reason for decreased accuracies of our algorithms. Finally, the best and worst algorithms in terms of accuracy for every dataset is specified in the table. The field of sentiment analysis has a very wide scope of research and work. It helps to find the overall polarity of a huge amount of dataset in no time and the result can be used for further analysis, for growth, improvement, and betterment of the that particular domain or sector.

BIBLIOGRAPHY

[1] Source: https://www.brightlocal.com/learn/local-consumerreview-survey/

[2] Source:https://www.inc.com/craig-bloem/84-percent-of-people-trustonline-reviews-as-much-.html

[3] Y. Mejova, "Sentiment analysis: An overview," Comprehensive exam paper, available on http://www. cs. uiowa. edu/ ymejova/publications/CompsYelenaMejova. pdf [2010-02-03], 2009.

[4] Fig 2 Source:http://msb.embopress.org/content/12/7/878.

[5] Neethu M. S., Rajasree R Sentiment Analysis in Twitter using Machine Learning Techniques,4th ICCCNT 2013,July 2013, Tiruchengode, India.

[6] G. Vinodhini and R. Chandrasekaran, "Sentiment analysis and opinion mining: A survey," International Journal, vol. 2, no. 6, 2012.

[7] Z. Niu, Z. Yin, and X. Kong, "Sentiment classification for microblog by machine learning," in Computational and Information Sciences (ICCIS), 2012 Fourth International Conference on, pp. 286289, IEEE, 2012.

[8] P. D. Turney, "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews," in Proceedings of the 40th annual meeting on association for computational linguistics, pp. 417424, Association for Computational Linguistics,2002.

[9] From Group to Individual Labels using Deep Features, Kotzias et. al,. KDD 2015.

[10] Thomas Kluyver, Benjamin Ragan-Kelley, Fernando Prez, Brian Granger, Matthias Bussonnier, Jonathan Frederic, Kyle Kelley, Jessica Hamrick, Jason Grout, Sylvain Corlay, Paul Ivanov, Damin Avila, Safia Abdalla, Carol Willing, Jupyter Development Team,cover Jupyter Notebooks a publishing format for reproducible computational workflows,87 - 90,,ElPub conference