

A MAJOR-PROJECT REPORT
ON
“Analysis of Twitter Data for Business Intelligence”

Submitted in partial fulfillment of the requirements of the degree of

“Bachelor of Computer Engineering”

by

Name	Division	Roll No
Ishmeet Arora	A	05
Apurva Chaudhari	A	11
Preeti Jaiswal	A	31

Project Guide: Prof. Sulochana Madachane



Department of Computer Engineering

K.C. College of Engineering and Management Studies And
Research, Thane (E)
University of Mumbai

2020-21

CERTIFICATE

Certified that the major-project work entitled “**Analysis of Twitter Data for Business Intelligence**” is a bonafide work carried out by

Name	Division	Roll No
Ishmeet Arora	A	05
Apurva Chaudhari	A	11
Preeti Jaiswal	A	31

The report has been approved as it satisfies the academic requirements in respect of major-project work prescribed for the course.

.....
Mrs. Sulochana Madachane
Project Guide

.....
Mr. Mandar Ganjapurkar
Head of the Department

.....
Dr. Vilas Nitnaware
Principal

Department of Computer Engineering
K.C. College of Engineering and Management Studies And
Research, Thane (E)
University of Mumbai

2020-21

Acknowledgement

No project is ever complete without the guidance of those expert who have already traded this past before and hence become master of it and as a result, our leader. So, we would like to take this opportunity to take all those individuals how have helped us in visualizing this project.

We would take this opportunity to thank our major project co-ordinator Mrs. Sulochana Madachane and other staff for their guidance in selecting this project and also for providing timely assistant to ourquery and guidance of this project.

We are very grateful to our Head of the Department Mr. Mandar Ganjapurkar for extending his help directly and indirectly through various channels in our project work.

We extend our sincerity appreciation to our Principal Dr. Vilas Nitnaware for providing us the opportunity to implement our project.

We are really thankful to all our Professors from K.C. College of Engineering and Management Studies and Research for their valuable inside and tip during the designing of the project. Their contributions have been valuable in so many ways that we find it difficult to acknowledge of them individual.

Thanking You.

Project Report Approval for B.E.

This project report entitled **Analysis of Twitter Data for Business Intelligence** by

Name	Division	Roll No
Ishmeet Arora	A	05
Apurva Chaudhari	A	11
Preeti Jaiswal	A	31

is approved for the degree of Bachelor of Engineering in Computer Engineering.

Project Guide:

Prof. Sulochana Madachane

Date: 07/07/2021

Place: Thane

Abstract

Understanding customers' feedback and knowing what your strengths and weaknesses are is key to any business. Nowadays, companies have access to a lot of information that could give them those insights: website reviews, chat interactions, conversations transcripts, social media comments.

Automated tools have emerged to facilitate this analysis, however most lack the capability of extracting the relationships between the reviews' rich expressions

Managers and marketers often resort to manually read through voluminous reviews to find the relationships. To address these challenges we have proposed to develop a system which will automatically analyze customer reviews, such as opinions in survey responses and social media conversations, allowing brands to learn what makes customers happy or frustrated, so that they can tailor products and services to meet their customers' needs.

TABLE OF CONTENT

Sr.No.	Topic	Page No.
	Acknowledgement	I
	Approval Sheet	II
	Abstract	III
	Table of Content	IV
	List of Figures	V
	List of Tables	VI
	Declaration	VII
1	Introduction	1
2	Review of Literature	2
3	Report on Present Investigation	3
	3.1 Requirement Analysis	4
	1. Scope	
	2. Feasibility Study	
	3. Hardware & Software Requirement	
	3.2 Problem Statement	6
	3.3 Project Design	7
	3.4 Methodology	8
	3.5 Implementation Plan	
	3.5.1 Semester VII	
	3.5.2 Semester VIII	
4	Conclusion	10
	References	

List of Figure

Figure No.	Name of Figure	Page No.
1	Project Design	7
2	Semester VII Plan	9
3	Semester VIII Plan	9

Declaration

We declare that this written submission represents our ideas in our own words and where other's ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Group Members:

Name of Students

Batch & Roll. No

Signature

Ishmeet Arora

A1 - 05



Apurva Chaudhari

A1 - 11



Preeti Jaiswal

A2 - 31



Date: 10/11/2021

1. Introduction

Nowadays everything is getting digital. With the boom in internet, there is a high competition in the industry to retain the customers, companies have the urge to analyze the feedback and evolve over time. Companies have access to a lot of information that could give them those insights: website reviews, chat interactions, conversations transcripts, social media comments.

Many organizations, large or small, gather customer feedback to improve their performance. But gathering feedback alone can't make much of a difference. We need to analyze our feedback to discover insights that inspire us to drive action at our organizations.

Actionable insights are meaningful findings that result from analyzing data. They make it clear what actions need to be taken or how one should think about an issue. Organizations use actionable insights to make data-informed decisions.

Actionable insights matter because one can use them to make strategic, well thought-out decisions. These decisions can drive positive outcomes specific to the business. This is because we derive the insight straight from the data such as the sales data or the customers' feedback. The large volume of review data makes it difficult to manually analyze customers' concerns. Hence, our objective is to quickly turn unstructured feedback into insights

2. Review of Literature

- **HTwitt: a Hadoop -based platform for analysis and visualization of streaming Twitter data-** They present Their framework HTwitt, built on top of the Hadoop ecosystem, which consists of a Map Reduce algorithm and a set of machine learning techniques embedded within a big data analytics platform, data preprocessing needs substantial manual effort; domain knowledge is required before the classification.
- **Product Features based Sentiment Analysis from Twitter** – This search discussed three different ways to extract(feature/opinion) pairs from each text including: Normal Tokenization, N-gram Modeling Extraction, and Noun Chunking Extraction.
- **A Survey on Deep Learning for Named Entity Recognition-** This paper includes a comprehensive review on existing deep learning techniques for NER. They first introduce NER resources, including tagged NER corpora and off-the-shelf NER tools. Then, they systematically categorize existing works and survey the most representative methods for recent applied techniques of deep learning in new NER problem settings and applications.
- **Machine translation using natural language processing** - This paper discusses the building of a deep neural network that functions as a part of end-to-end translation pipeline. It has three main parts which are preprocessing, creation of models and Running the model
- **Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents** - The study is focused on how the algorithm can be applied on number of documents. its discusses the working principle and steps which should be followed for implementation of TF-IDF .
- **DepecheMood++: a Bilingual Emotion Lexicon** - In this paper we learnt that, straight forward classifiers/regressors built on top of the proposed lexica and without additional features obtain good performances even on domain-specific tasks, and can provide more challenging baselines.
- **Survey on Parts of Speech Tagger Techniques** - This paper discusses the survey of different types of POS taggers designed by the researchers and organizations. Different taggers will have their own tag set.
- **Investigation of Sentiment Analysis Models for Natural Language Processing** - Learning from consumer reviews to classify polarity - In this project the aim is to investigate a sentiment analysis model that learns from consumer reviews from the yelp data set and then classifies these reviews as either positive or negative. It was found that the most successful model was the bag of valuable words model for feature extraction in combination with the naive bayes model for classification
- **Deep Learning for Hate Speech Detection in Tweets** - In this paper, we got introduced to the application of deep neural network architectures for the task of hate speech detection. We found them to significantly outperform the existing methods.
- **Extracting Business Intelligence from Online Product Reviews** - The project proposes to build a system which is capable of extracting business intelligence for a manufacturer, from online product reviews.

3. Report on Present Investigation

3.1 Requirement Analysis

3.1.1 Scope

Business Intelligence is data-driven decision-making that includes the generation, accumulation, analysis, and visualization of data in a bid to help managers make better business decisions. It involves processes and procedures that strengthen our capacity for collecting, sharing, and analyzing data and making informed decisions. Customer data is like a gold mine and Business Intelligence can help businesses leverage this data to increase customer engagement and satisfaction and resolve business inadequacies.

3.1.2 Feasibility Study

An analysis of the current mode of operation:

Progressive companies today say they want to be data-driven. Forrester reports that 74% of companies say this is a goal. Although only 29% of these companies are actually successful in actioning their analytics. It's clear that the missing link for companies wanting to drive business outcomes from their data is actionable insights.

A definition of requirements:

The Hardware Requirements are as follows,

Processor: Quad-Core (or more)

RAM: 8GB or above

The Software Requirement are as follows,

Language- Python

OS- Windows, Mac or Linux

Web Browser- Google Chrome

IDE- Jupyter Notebook

Libraries- Matplotlib, nltk, seaborn, re, pandas, numpy, sklearn, wordcloud, flask

An agreed upon course of action:

This system which will automatically analyze customer reviews, such as opinions in survey responses and social media conversations, allowing brands to learn what makes customers happy or frustrated, so that they can tailor products and services to meet their customers' needs.

First we will extract real time data from twitter using Tweepy library in python. We will store data in dataframes. We will pre process the data by removing null values, punctuation, converting emoticon to words, converting abbreviations into phrases.

Then we tokenize the data by converting sentences into tokens and we lemmatize these tokens using nltk lemmatizer. Then we start by looking at the most frequently used words by using n-grams concept. Then we create a wordcloud to visualize the data.

Then we use tf-idf vectorizer to convert words into numerical arrays. Then we use a classifier to predict the sentiment. We use libraries like SHAP or LIME to explain from feature impact.

The next step is to code your feedback data so we can extract meaningful insights in the next step. Coding is the process of labelling and organizing your data in such a way that you can then identify themes in the data, and the relationships between these themes. Create a code frame to group your codes. The coding frame is the organizational structure of all your codes. Based on the number of times a particular code occurs, you can now see the common themes in your feedback data.

The Steps to be followed for making this project –

Fetching Real Time Data

Preprocessing and Cleaning

Feature Extraction

Sentiment Analysis

Thematic Analysis

3.1.3 Hardware & Software Requirements

Hardware Requirement –

- Processor: Quad-Core (or more)
- RAM: 8GB or above

Software Requirement –

- Language- Python
- Web Browser- Google Chrome
- IDE- Jupyter Notebook
- Libraries- Matplotlib, nltk, seaborn, re, pandas, numpy, sklearn, wordcloud
- OS- Windows, Mac or Linux

3.2 Problem Statement

With millions of customers, it is almost impossible to manually review the overall customer sentiment about their product. That is where our problem is, we need to analyze the customer reviews about their products and help the organization improve upon their business strategies accordingly. These days people are very active on social media and are more vocal than ever before about their opinions on various products, brands, etc.

So our objective is, we will provide a platform where a few organization can analyze the performance of their mobile phones, laptops & earphones and use it to make important business decisions using the reviews by the customer on social media sites. The Companies can analyze online reviews of their products and compare them to their competition. Maybe their competitor released a new product that landed as a flop. Find out what aspects of the product performed most negatively and use it to their advantage.

Our Project can be used in social media monitoring, allowing businesses to gain insights about how customers feel about certain topics, and detect urgent issues in real time before they spiral out of control.

3.3 Project Design

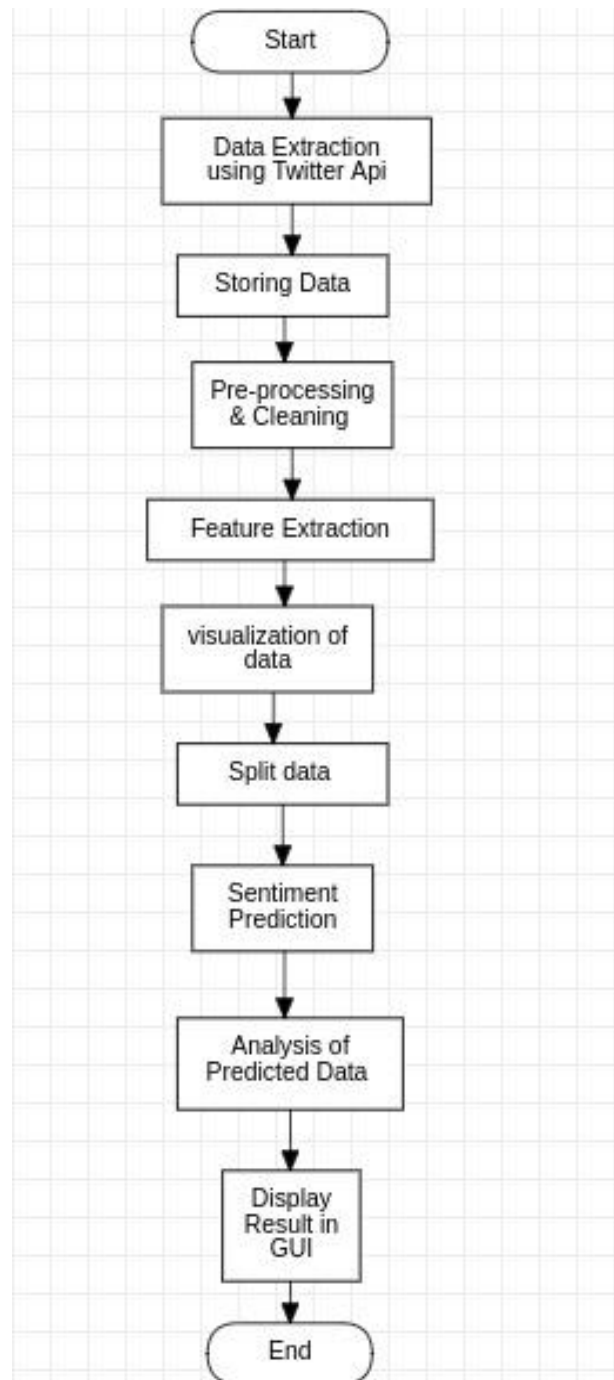


Fig. 1 – Project Design

Methodology

Dataset: Real time data

Steps:

1. Extracting real time reviews and storing them in the system.
2. Pre-processing and Cleaning data
3. Feature extraction using tokenisation and lemmatization
4. Visualization of data
5. Using Vectorizers and Classifiers for Sentiment Analysis of Data
6. Thematic Analysis of data which includes
 - a. We code our feedback data so we can extract meaningful insights in the next step.
 - b. Once it has all been coded, create a code frame to group your codes.
7. Based on the number of times a particular code occurs, you can now see the common themes in your feedback data.

3.4 Implementation Plan

3.4.1 Semester VII

	Jul-21	Aug-21	Sep-21	Oct-21
Requirement Gathering				
Study of Research Paper				
Innovation Ideas Identified				
Literature Review				
Problem Statement				

Fig. 2 – Semester VII Plan

3.4.2 Semester VIII

	Nov-21	Dec-21	Jan-21	Feb-21	Mar-21
Design					
Coding					
Testing					
Real Time Applicability					

Fig. 3 – Semester VIII Plan

4. Conclusion

Thus, we have created a platform where the organization can not only see the customer views on about their product but can also use the analysis to make important business strategies and improve their performances. Thus, business intelligence using Sentiment analysis can be of great help to organization and facilitate their growth.

References

- [1] Qaiser, Shahzad & Ali, Ramsha. (2018). Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents. International Journal of Computer Applications. 181. 10.5120/ijca2018917395.
- [2] S. G. Kanakaraddi and S. S. Nandyal, "Survey on Parts of Speech Tagger Techniques," 2018 International Conference on Current Trends towards Converging Technologies (ICCTCT), 2018, pp. 1-6, doi: 10.1109/ICCTCT.2018.8550884.
- [3] Jing Li, Aixin Sun, Jianglei Han, and Chenliang Li. (2020). A Survey on Deep Learning for Named Entity Recognition. <https://arxiv.org/pdf/1812.09449.pdf>
- [4] Middi, Venkata Sai Rishita & Raju, Middi & Harris, Tanvir Ahmed. (2019). Machine translation using natural language processing. MATEC Web of Conferences. 277. 02004. 10.1051/mateconf/201927702004.
- [5] Iftene, A. (2011). Sentimatrix– Multilingual Sentiment Analysis Service. ACL HLT
- [6] Demirbaga, Umit. "HTwitt: a Hadoop-Based Platform for Analysis and Visualization of Streaming Twitter Data." Neural Computing and Applications (2021): n. pag. Web.
- [7] of Computer Science IJCSIS, J. (2020). Product Features based Sentiment Analysis from Twitter. IJCSIS Vol. 18 No. 8 August Issue. <https://doi.org/10.5281/zenodo.4012460>
- [8] Torin Pitchers.(2018). Investigation of Sentiment Analysis Models for Natural Language Processing: learning from consumer reviews to classify polarity. https://www.academia.edu/36138536/Investigation_of_Sentiment_Analysis_Models_for_Natural_Language_Processing_learning_from_consumer_reviews_to_classify_polarity?email_work_card=view-paper
- [9] Demirbaga, U. (2021). HTwitt: a Hadoop-based platform for analysis and visualization of streaming Twitter data. Neural Computing and Applications. <https://doi.org/10.1007/s00521-021-06046-y>
- [10] Soundarya.V, & Rupa, Siddareddy & Khanna, Sristi & G.Swathi, & D.Manjula,. (2013). Extracting Business Intelligence from Online Product Reviews. International Journal on Soft Computing. 4. 10.5121/ijsc.2013.4302.
- [11] <https://towardsdatascience.com/step-by-step-twitter-sentiment-analysis-in-python-d6f650ade58d>
- [12] <https://towardsdatascience.com/extracting-data-from-twitter-using-python-5ab67bff553a>