

Bachelor of Science in Computer Science & Engineering



Sentiment Analysis of Product Reviews of Online Shop

by

Md. Jahedul Alam Rifat

ID: 1604121

Department of Computer Science & Engineering

Chittagong University of Engineering & Technology (CUET)

Chattogram-4349, Bangladesh.

July, 2021

Table of Contents

List of Figures	ii
1 Introduction	1
1.1 Motivation	2
2 Background and Present State	2
2.1 Research Gap	3
3 Specific Objectives and Possible Outcomes	4
4 Impact Identification	4
5 Outline of Methodology	4
5.1 Collecting Dataset	5
5.2 Preprocessing the Dataset	5
5.3 Building The Model	6
6 Applications	7
7 Required Resources and Tools	7
7.1 Required Resources	7
7.2 Required Tools	7
8 Cost Estimation	8
8.1 Time Management	8

List of Figures

5.1	Overview of this work	5
5.2	Flowchart for Pre-processing	6
5.3	Flowchart for Building the Model	7
8.4	Gantt chart for this work	9

1 Introduction

Facebook is one of the most globally used social sites. In this pandemic situation people are tend to become more addicted to this. Some of us have started doing small business using this kind of social networks. And Almost all the big companies have a Facebook page to provide services or as a customer service point.

People of our country, Bangladesh, are not apart from this. We have also an online based economy growing up. People using Social sites like Facebook as their side income source. Some are also investing much and using as their main income source. So, Many Transaction are being done regularly. Some finds good product whether other finds product being cheaper but damaged. Some of us also being cheated as the owner shows something else and provide almost a different thing. Sometimes the color of the ordered product gets different to the product delivered. In these cases, we complain about this problems in the comment section of the page. Although, It's not the fact that we always befooled in the online shopping. We get authentic products also. At that time we post a positive feedback about the product or the delivery process and the time taken for the delivery. When someone visits a page searching for his desired product, he goes after the review before buying. In case, if the product is very popular, there can be lots of reviews. That much review can't be read or it will be very time consuming. But if he doesn't read any review then there might be a chance not getting the right product.

This work is related to such Facebook page which provides gadgets using online services. We are to analyze the customer reviews and comments. After analyzing, using training data we will build a model that can evaluate a sentimental output of that review. Thus, a new buyer can easily get the message about the product that he is intended to buy.

1.1 Motivation

Whether it's a gadget, clothing, food or anything of our daily needs, we can find it online today. People are becoming more addictive to buy online rather than going market in person. So they order their needs in any online sites. Later they are likely to post a review about their deals with that shop and the product they got. As a result, Millions of reviews are being generated all over internet. Some of the reviews are positive, whereas some are negative. Some have delivery issue also. As there can be a lot of review for only one product, an user cannot go through all of them. So We can analyze each of those reviews for them and categorize the reviews into some groups. That can be positive or negative according to the review text. So that, Future buyer may find it easier to get information easily about a product he wants.

2 Background and Present State

Rajkumar et al.[1] Satuluri Vanaja and Meena Belwal [2] analyzed two Machine Language approaches namingly Naïve Bayes (NB) and Support Vector Machine(SVM) for performing Sentiment Analysis on reviews of a specific product. In those approaches, they had collected the dataset from Amazon, which included reviews regarding Laptops, Cameras, Mobiles, Tablets, video surveillance and TVs. The dataset of[1] consisted 4783 negative and 2006 positive words with sentiment scores intended for every sentence were evaluated. Applying the ML approaches, NB got 98.170% accuracy and SVM got 93.54% accuracy for Reviews related to cameras. Both the concluded that, the NB was more accurate than the SVM algorithm

Bagus Setya and Riyanarto Sarno [3] had tried to extract the features adopting a Word Sense Disambiguation (WSD) technique. They reviewed the sentences by assigning the correct sense of the word using that technique. They claim that the performance of their suggested semantic characteristics has improved by 10.9%, 9.2%, and 10.6% of precision, recall, and F-Measure, respectively after using 10-fold cross-validation.

LI YANG [4] applied sentiment analysis on the reviews taken from dangdang.com, a famous Chinese e-commerce website. They had around 100000 orders of magnitude, which are being used in the field of Chinese sentiment analysis. According to him, there are some methods such as Word2Vec, FastText, and Glove, the traditional machine learning method are used to obtain the word vector features of the text automatically but those need human intervention, vectorization of the text to extract the emotional features of the structured data from the input text and then use the traditional machine learning model to classify the sentiment text feature. That's why, he preferred Deep Learning as he had a huge dataset also. Navneet Kaur [5] tried to do sentiment analysis using the tweets of a multinational firm called Samsung Electronics Ltd. They tried to add neutral tweets , along with the positive and negative tweets .

According to Xiao [6] In the training stage, the LSTM method outperforms Naive Bayes and logistic regression algorithms, and it provides a credible reference for product review analysis. They applied LSTM with two other algorithms on at least 100,000 reviews of the digital and electronic products on Jingdong Mall,China. Sheela and SP [7] tried to predict the rating from the reviews if any of the customer missed to input the rating after commenting about a product.

Wiga Maulana and Aulia Nurwanti [8] had used the combination of K-means and the algorithm of naive Bayes classifier. K-means is used to distinguish the class review comments of a product into a positive comment class or a negative comment class. While the naive Bayes classifier algorithm is used to derive the accuracy value of positive comment data and negative product reviews. In terms of accuracy and computation, the Naive Bayes classifier method is thought to be a potentially effective way for classifying data that differs from previous classification methods.

2.1 Research Gap

Sentiment Analysis is one of the most common research topic in the field of NLP. There are also some recent researches for product reviews as mentioned in the above section. Some of the recent works [1, 2, 5, 9]on sentiment analysis

of product reviews have evaluated two (positive or negative) or three outputs including neutral reviews. Bagus and Sarno [3] tried for the semantic analysis of a review to extract the correct meaning. So we will try to expand the work by evaluating more than three reactions and also predict a rating analyzing the user reviews.

3 Specific Objectives and Possible Outcomes

We are to complete this following objectives:

- To extract Reviews from Facebook Page
- To predict a rating, analyzing the reviews of a product
- To perform Sentiment Analysis of the Review (positive, negative, sarcastic, angry, admiration)

4 Impact Identification

This work of analyzing sentiment can be impactful as-

- Customers can easily get a brief overview of a product Viewing the rating and setimental output
- Customers may find information about the shop whether the owners are customer friendly or not
- Ratings can represent the delivery issues i.e timely,well packaged or not to a new buyer of the shop

5 Outline of Methodology

The main goal of this project is to build a model for analyzing the sentiment of product reviews. For this, We have to collect data from Facebook. Facebook offers a graph API of its own. Using that we can gather comments and reviews

from Facebook page. Then we have to manually check for unrelated comments or redundancy. And also label the data. Then we have to separately analyze every sentence and determine the emotion and negation. We have to further check if that sentence hold mixed emotions. Generally, we notice some Sarcastic comments regarding products or the delivery services now-a-days. For this reason, we have to analyze the sentences individually and then preprocess them. After that, we are to send the pre-processed data for training the model. Finally, the model is being tested with some random comments. Then it can be evaluated for real time environment.

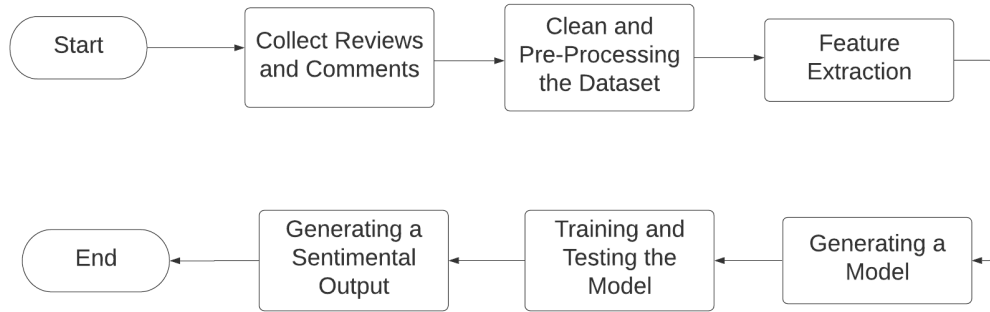


Figure 5.1: Overview of this work

5.1 Collecting Dataset

Facebook offers a graph API for collecting post or comments via access token. It can be applied to collect posts, comments or posting a status using bots. We can simply apply for an access token and gather reviews and comments of the posts

5.2 Preprocessing the Dataset

After the collection of the dataset, we have an important task to complete next. That is, Pre-processing the dataset for further use. The dataset may contain unrelavent data, pictures or links. We are to remove those. Then we have to analyze the sentences separately. There can be three parts in a sentence,

- Words

- Question mark or Exclamation sign
- Emoticon

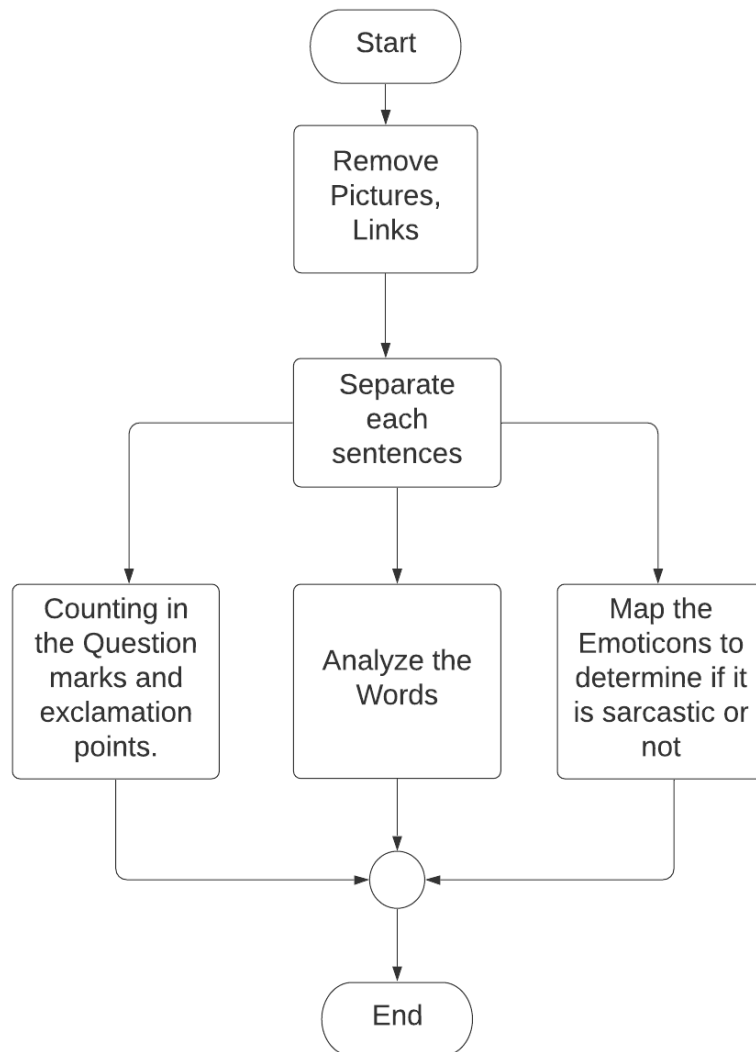


Figure 5.2: Flowchart for Pre-processing

5.3 Building The Model

After Preprocessing the dataset, we may have to label the data. Then, we can use that to build and training a model. After the completion of training, we may input random sample reviews to test the model and compare the outputs.

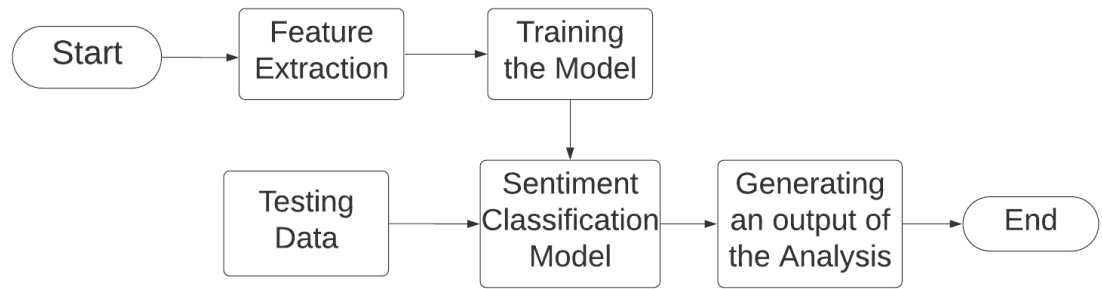


Figure 5.3: Flowchart for Building the Model

6 Applications

This Model can be used to analyze the sentiments of a review in real time environment. Thus, other viewer or buyer can easily get the information regarding the product . They may find if the seller is trusted or not. For any of this, They won't have to consume their time in reading reviews.

7 Required Resources and Tools

7.1 Required Resources

- Operating System : Windows 10.
- Python
- PyCharm IDE for Python
- Libraries: Tensorflow, Keras etc
- Lucid-For making Flowchart or diagram

7.2 Required Tools

- Personal Computer

8 Cost Estimation

a. Cost of Materials :

- Personal Computer Tk 64000

Total	Tk. 64000
-------	-----------

b. Drafting & Binding :

- Paper Tk 500
- Drafting Tk 1000
- Printing Tk 500
- Binding Tk 500

Total	Tk. 2500
-------	----------

c. Internet Browsing	Tk. 2000
----------------------	----------

d. Miscellaneous	Tk. 500
------------------	---------

Grand Total	Tk. 69000
-------------	-----------

8.1 Time Management

This is the estimated timeline for this thesis work.

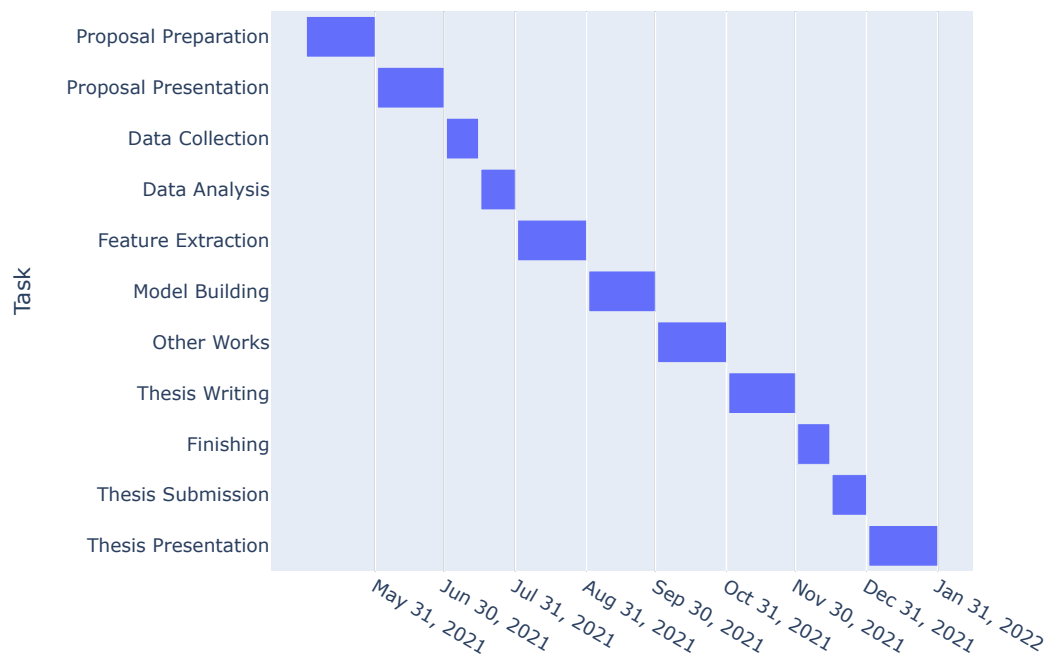


Figure 8.4: Gantt chart for this work

CSE Undergraduate Studies (CUGS) Committee

Reference :

Meeting No :

Resolution No :

Date :

Signature of the Student

Signature of the Supervisor

Signature of the Head of the Department

References

- [1] R. S. Jagdale, V. S. Shirsat and S. N. Deshmukh, ‘Sentiment analysis on product reviews using machine learning techniques,’ in *Cognitive Informatics and Soft Computing*, Springer, 2019, pp. 639–647 (cit. on pp. 2, 3).
- [2] S. Vanaja and M. Belwal, ‘Aspect-level sentiment analysis on e-commerce data,’ in *2018 International Conference on Inventive Research in Computing Applications (ICIRCA)*, IEEE, 2018, pp. 1275–1279 (cit. on pp. 2, 3).
- [3] B. S. Rintyarna, R. Sarno and C. Fatichah, ‘Semantic features for optimizing supervised approach of sentiment analysis on product reviews,’ *Computers*, vol. 8, 3 2019, ISSN: 2073431X. DOI: 10.3390/computers8030055 (cit. on pp. 2, 4).
- [4] L. Yang, Y. Li, J. Wang and R. S. Sherratt, ‘Sentiment analysis for e-commerce product reviews in chinese based on sentiment lexicon and deep learning,’ *IEEE Access*, vol. 8, pp. 23 522–23 530, 2020 (cit. on p. 3).
- [5] R. K. Bakshi, N. Kaur, R. Kaur and G. Kaur, ‘Opinion mining and sentiment analysis,’ in *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, 2016, pp. 452–455 (cit. on p. 3).
- [6] S. Xiao, H. Wang, Z. Ling, L. Wang and Z. Tang, ‘Sentiment analysis for product reviews based on deep learning,’ in *Journal of Physics: Conference Series*, IOP Publishing, vol. 1651, 2020, p. 012 103 (cit. on p. 3).
- [7] S. Sheela, ‘Sentiment analysis and prediction of online reviews with empty ratings,’ *International Journal of Applied Engineering Research*, vol. 13, no. 14, pp. 11 532–11 539, 2018 (cit. on p. 3).
- [8] T. Hariguna, W. M. Baihaqi and A. Nurwanti, ‘Sentiment analysis of product reviews as a customer recommendation using the naive bayes classifier algorithm,’ *International Journal of Informatics and Information Systems*, vol. 2, no. 2, pp. 48–55, 2019 (cit. on p. 3).
- [9] R. S. Jagdale, V. S. Shirsat and S. N. Deshmukh, ‘Sentiment analysis on product reviews using machine learning techniques,’ in *Cognitive Informatics and Soft Computing*, Springer, 2019, pp. 639–647 (cit. on p. 3).