SENTIMENT ANALYSIS ON BANGLADESHI E-COMMERCE REVIEW DATA

BY

NOWSHIN SHARMILY
ID: 161-15-7081

RAKIB UDDIN
ID: 161-15-6871

AND
ASHADUJJAMAN TUSHAR
ID: 161-15-6862

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering.

Supervised By

Md. Sadekur Rahman
Assistant Professor
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY
DHAKA, BANGLADESH
OCTOBER 2020

APPROVAL

This Project/internship titled **"Sentiment Analysis on Bangladeshi E-commerce Review Data"**, submitted by Rakib Uddin, ID No: 161-15-6871, Nowshin Sharmily, ID No: 161-15-7081 and Ashadujjaman Tushar, ID No: 161-15-6865 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 07 October 2020.
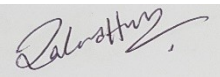
BOARD OF EXAMINERS

Dr. Syed Akhter Hossain   **Chairman**
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
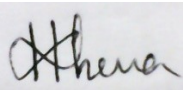Daffodil International University

Md. Zahid Hasan   **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Most. Hasna Hena   **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

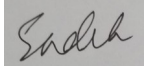**Dr. Mohammad Shorif Uddin**   **External Examiner**
Professor
Department of Computer Science and Engineering
Jahangirnagar University

DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Md. Sadekur Rahman, Assistant Professor, Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

SUPERVISED BY:

MD. SADEKUR RAHMAN
Assistant professor
Department of CSE
Daffodil International University
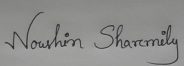
CO-SUPERVISED BY:

MR. AHMED AL MAROUF
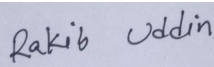**Lecturer**
Department of CSE
Daffodil International University

SUBMITTED BY:

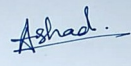**NOWSHIN SHARMILY**
ID: 161-15-7081
Department of CSE

RAKIB UDDIN
ID: 161-15-6871
Department of CSE

ASHADUJJAMAN TUSHAR
ID: 161-15-6862
Department of CSE

ACKNOWLEDGMENT

First, we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year Thesis successfully.

We really grateful and wish our profound our indebtedness to **Supervisor Md. Sadekur Rahman, Assistant Professor**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Artificial Intelligence*" to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stage have made it possible to complete this project.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, we must acknowledge with due respect the constant support and patience of our parents.

ABSTRACT

Today's world is easily depended on technologies. Nowhere is found running fast & without digital compartments. Even peoples' directly approaches are covered through the internet which made our lives not only easier but also accruable. Online shopping platform is growing faster among the whole world as well as our digitally developing country. Many e-commerce shopping sites are taking place of our day to day marketplace as we can purchase any available material or service staying at home.

Despite of the benefits there are also critics of e-commerce since the customers leave a review about the service of products in text review form. Sentiment analysis of online reviews has been researched ago but on Bengali language there are not found very promising work. So we found a scope of research on this topic to classify the sentiment behind Bangladeshi e commerce sites' reviews and deliver a modeled output. This will help the new customers to acknowledge a true description of old buyers and also the sites can upgrade their service or product through customer satisfaction. We used three different algorithms to sort our data to see which one gives better accuracy to human sentiments and classify the reviews as excellent, better, good, poor or very bad.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURES:

LIST OF TABLES

TABLES

LIST OF ABBREVIATION

DIU – Daffodil International University

CSE – Computer Science and Engineering

AI – Artificial Intelligence

ML – Machine Learning

NLP – Natural Language Processing

NLTK – Natural Language Toolkit

SVM – Support Vector Machine

LSTM – Language Short Long Term

CHAPTER 1
INTRODUCTION

1.1 Introduction

Web based business has become a mainstream business type everywhere throughout the world as well as Bangladesh. Merchant and client getting more intrigued by internet shopping day by day. After a fruitful shopping client leaves a review for that specific item. This survey isn't just a method for what is at the forefront of a client's thoughts, yet in addition to communicating what was his experience and about the quality of that item. Through performing notion investigation on item review a potential customer can know ahead of time what past purchasers think about that item/administrations before getting that administration or item, while organizations can break down criticism on various items or administrations presented by clients on gain information about which items or administrations to sell more which ought to be improved.

Sentiment analysis is the programmed extraction of conclusions, feelings, and assessments from texts. There has been a lot of exploration led in the region of sentiment classification, the majority of which is executed on the English language, however work in Bangla language is limited to just blogs and news corpus, such as in [1]. Products review data contrasts from blogs and news corpus in that it's substance is commonly littler and loud. We were unable to discover any research paper which explicitly centers around the issue of separating client assessments and perspectives from Bangla product review data, so there is an opportunity for research in this space for the Bengla language.

In this research work, we expect to automatically extricate sentiment or supposition from Bangladeshi online business product review data. Which would be advantageous to a wide scope of real world applications. Users can recognize the kind of criticism for a specific item or service so as to settle on an educated choice on their buy. By applying the AI calculations with the training dataset then it is conceivable to arrange absolutely unknown data. We use Multinomial Naive Bayes, Logistic Regression and Long short-term memory (LSTM) and do a comparative analysis on the presentation of these AI calculations by trying different combination sets of features.

1.2 Motivation

There is such a large amount of research work on product review data. In any case, above all there is basically no work on product review data in the point of view of Bangladeshi online business site information. This base gives us the motivation to chip away at it. The most imperative second that obviously roused us that these are basically no work had been performed principally based on Bangladesh online business dataset.

Sentiment analysis is a process to automatically extricate sentiment or opinion from online business product review data. Our prepared model will assist possible clients with settling on an informed decision on their purchase and organizations to improve their items or services.

1.3 Rationale of the Study

Our research will support organizations or the proprietor of the business to make a few strides for the betterment of their products and services. Also, clients can gain educated information before buying or taking products and services.

The most important thing that really inspired us is that there is practically no research has been completed based on Bangladeshi e-commerce product review dataset. We have chosen it for this reason. We think this will be a unique work from others.

1.4 Research Question

We have chosen some questions which are being addressed stepwise in our research work.

1 What is sentiment analysis and who are the beneficiaries of sentiment analysis?
2 How might we recognize customers' sentiment from their review?
3 Which algorithm performs better to predict customers' sentiment?

1.5 Expected Outcome

The model's performance will be tried by applying various sorts of machine learning algorithms on our dataset. The work would test how precise calculations respond to our dataset.

About ascertaining the absolute component of the dataset and all out the number of Excellent, Good, Normal, Bad and Terrible reviews. The research work will convey a magnificent aftereffect of the review of desired online business item review data by utilizing the Model. Regardless of whether it is Excellent, Good, Normal, Bad or Terrible, clients and proprietors of the online business will be profited by expecting the key purposes of the average assessment which have been given by the past clients of that website.

1.6 Report Layout

Chapter 1 Discusses about our thesis motivation, Rationale of the Study, Research Question and

Expected Outcome.

Chapter 2 Introduce the Background study of our research. It furthermore gives us the realities

of related works. Difficulties additionally are mentioned right here.

Chapter 3 Discusses about the technique of our research work. Details works of data collection, data processing, machine learning and NLTK technique. Here additionally mentioned about the data collection processes.

Chapter 4 Discuss details about the result and discuss our project with experiment and result.

Chapter 5 Discuss our research with future scope that can be implemented and conduct the research work.

CHAPTER 2

BACKGROUND

2.1 Introduction

Sentiment analysis is one of a vast researchable area for modern technology service monitoring and online audience reaction visualization. Many researches has been done basically on international language preference till now. For example Aljoharah had worked on data processing by amazon reviews where maximum contents are in english language. As it is most popular and used as international language, there are so many research works using various data mining algorithms applied on English language to analyze sentiments from all over the world.

Since there is not that much work being done on focusing Bangla language data as well as Bangladeshi data, our project is focused on addition of Bangla language and Bangladeshi text data on countries' e-commerce platforms. For this we used Long Short-Term Memory (LSTM) algorithm which gives maximum output analysis on multilingual operations to help AI recognize more of human reactions and sort them to categorize different sentiments on different services.

2.2 Related works

Sentiment analysis related works are doing more specifically than before. As an era of broad internet connection the sector of possibility is increasing through research. Here we are pointing out some of the research that works on typical data sourcing or characterizing sentiment analysis.

In 2019 Aljoharah worked in a research paper where they visualized on amazon books reviews to find out highest rated books by testing 1000 customer reviews. They also checked out packed bubbles, stacked bars, linear charts and word-cloud to check additional information using raw data processing methods which can help upgrade service or product. Their challenges are for more specifying the result of product characterization[1].

Shaika and Wasifa first time in 2014, visualized bengali microblogs posts for sentiment analysis using NLTK Python toolkit. They used huge data collection. Preprocessing them for lexicon, emoticon, stemming, word n gram,negation and POS taggings. Their classification result accuracy winner is svm which acquired 93% success using a semi supervised bootstrapping methodology[2].

Another one on amazon product review was done in 2018 by Tanjim, Nudrat and Faisal visualizing a larger scale than before. They worked on pool based active learning,feature extraction and text classification to categorize in three different products such as electronics, cell phone and musical instruments which consisted of about 48500 reviews. They found out that the Support Vector Machine brings the most accurate results which is about 90%. Their future work is to apply PCA to process fully automated data labeling[3].

Geetika and Divakar in 2014 used machine learning approaching to twitter data semantic analysis to check out results in positive or negative approaches by analysing labeled datasets.The result was 89.9% accuracy gaining by using wordnet than 88.2% accuracy of naive bias testing[4].

Ana in 2017 researched sentiment analysis in tripadvisor to find out information from user opinions by using sentistrength, bing, syuzhet and coreNLP analysis methods. Their works faced several challenges due to tripadvisors specific content based opinions. However sentistrength detected highest 57.48% negative reviews among all four methods.[5]

Rabia has done the research on precise tweet classification and sentiment analysis to extract knowledge from tweets and classify them based on knowledge perception. Their huge data collection with knowledge generator and enhancer performed is to classify data more accurately. Also synonym binder and filter engine has shown efficiency in tweet testing and verification[6].

2012 sentiment analysis on social media by federico is on basic operation analyzation. 1000 facebook comments are reviewed to analyze a specific indian company to characterize its user interference. Supervised and unsupervised clustering, machine translation engine ,geo differentiation machine, natural language search, semantic search are visualized to gain approximate classification by social media reviews[7].

In 2013, Umesh worked on research of sentiment analysis on hollywood movies where maxent method resulted more effectively than naive bias and positive sentiments were higher in number. The accuracy level between positive, negative and cognitive sentiments show that

a higher number of analysis possibilities can be achieved through micro blogging and tweets which helps to understand human sentimental behaviour over cultural segments[8].

2.3 Research Summary

After these works we found out space of working on specific data characterization according to regional user interference. Sentiment analysis for bengali language category and emoticon, lexicons are far tried as of now. Enough bangla words data collection, sorting keyword sentiments and making an outcome of these can review texts which can help understanding product outcomes and human behavioral information of Bangladeshi people.

Table 2.1- Research Summary

No	Author	Year	Language	Domain	Algorithms	Accuracy
1	Almjawel, Aljoharah, et al.	2019	English	E-Commerce		
2	Chowdhury Shaika and Wasifa Chowdhury	2014	Bangla	Microblog/ Social Media	SVM MaxEnt	69% 67%
3	Haque, Tanjim Ul, Nudrat Nawal Saber and Faisal Muhammad Shah	2018	English	E-Commerce	Linear SVM Multinomial Naive Bayes Random Forest	93.57 90.28 92.72
4	Gautam, Geetika and Divakar Yadav	2014	English	Microblog/ Social Media	Naive Bayes Maximum Entropy Support Vector Machine	88.2 83.8 85.5
5	Valdivia, Ana, M. Victoria Luzón and Francisco Herrera	2017	English	E-Commerce	Core NLP Bing SentiStrength	94.12 91.18 57.84
6	Batool, Rabia, et al.	2013	English	Microblog/ Social Media	Knowledge Enhancer & Synonym Binder Knowledge Generator	89 34
7	Neri, Federico, et al.	2012	English	Microblog/ Social Media	Supervised Machine Learning Unsupervised Method	87 77
8	Hodeghatta, Umesh Rao.	2013	English	Microblog/ Movie Review	Naive Bayes(Unigram) Naive Bayes(Bigram) MaxEnt(Unigram)	79 64 84

2.4 Scope of the Problem

There are multiple problems in this work while sorting out the keywords. Users can use slang or unusual typing like shortest or wrongly spelled words. As there is no specific pattern yet considerable, this might be a hard thing to understand the sentiment behind it. Some people may use neutral approaches which may get hard to characterize to a positive or negative counting.

2.5 Challenges

The new changing and discovering words and emoticons has to be updated regularly. Also trying to categorize patterns of slang or unusual word sentiments is important to acquire more accurate results.

CHAPTER 3
RESEARCH METHODOLOGY

3.1 Introduction

Since reviews are textual and differing; an intuitive portrayal would be steady to users through summing up opinions and giving criticism. As text mining is scarcely selective from numeric data as text needs to data pre processing and we additionally use Bangla data so finding the classifications is another level challenge for completing our research work.

As a matter of first importance, we have pre-processed our dataset and changed the text content into vector structure, after that we have applied those data into the various classification algorithms.

Classification algorithms used as supervised machine learning algorithms. We have utilized multiclass classification in our proposed model to decide if the review is excellent, good, normal, bad or terrible.

3.2 Research Subject and Instrumentation

We are working on the premises of users' sentiment and opinion of Bangladeshi e-commerce site. For this we have got used review data from a Bangladeshi e-commerce site. We collected our required data using a scraping algorithm. But there are Bangla and English data mixed so we also have to partition Bangla and English data separately for measuring the performance individually.

We are working at the premises of clients' sentiment and assessment of Bangladeshi e-commerce business. For this we have collected review data from a Bangladeshi web based business website. We gathered our necessary data utilizing a scraping algorithm. Be that as it may, there are Bangla and English information blended so we likewise need to partition Bangla and English data independently for estimating the analysis separately.

From that point forward, we utilized NLTK (natural language toolkit) for evacuating stop words, tokenizing and furthermore expelling punctuation. Be that as it may, here we have utilized the sci-kit learning library which executed with python programming language.

Presently it's turning into an absolutely mainstream and valuable apparatus for analyzing textual data. The benefit of sci-kit learning is that we can import various sorts of libraries which incorporate explicit algorithm visualization tools. Here in our research all AI methods were executed by sci-kit learning library.

3.3 Data Collection Procedure

Data collection for our research is likewise a next level challenge. There is such a significant number of databases on product reviews however none of them are explicitly Bangladesh based. So we have to gather our information all alone. There are such huge numbers of all shapes and sizes web based business organizations in Bangladesh however for our information assortment we choose "Daraz Bangladesh" as they have a large number of clients just as a large number of reviews.

"Daraz Bangladesh" is a Bangladesh based e-commerce business, it's clients and sellers are all from Bangladesh. Here clients buy an item from the site and leave a review after getting their items. For considering review of an item, the client needs to have an account and should need to buy that item. So the entirety of the data is so exact. Clients without an account or buy can't make a review. We have gathered data from various categories of items, for example, electronics, home machines, fashion style things, etc.

3.4 Statistical Analysis

There is countless review data on the "Daraz Bangladesh" site. From there we have chosen 850 distinct English data and 600 distinct Bangla data for our system training dataset. Assortment of various reviews will give better outcomes, lastly we have a legitimate data review for our training dataset which is brimming with text.

3.5 Implementation Requirement

For finishing our research work we have kept up a couple of steps. Those all are related with one another. As we worked with text data so we need to tidy up our data before fitting them in our supervised algorithms. There were likewise some more significant strides in our exploration work. Figure 3.4.1 demonstrating the entirety of the means of our research procedure.
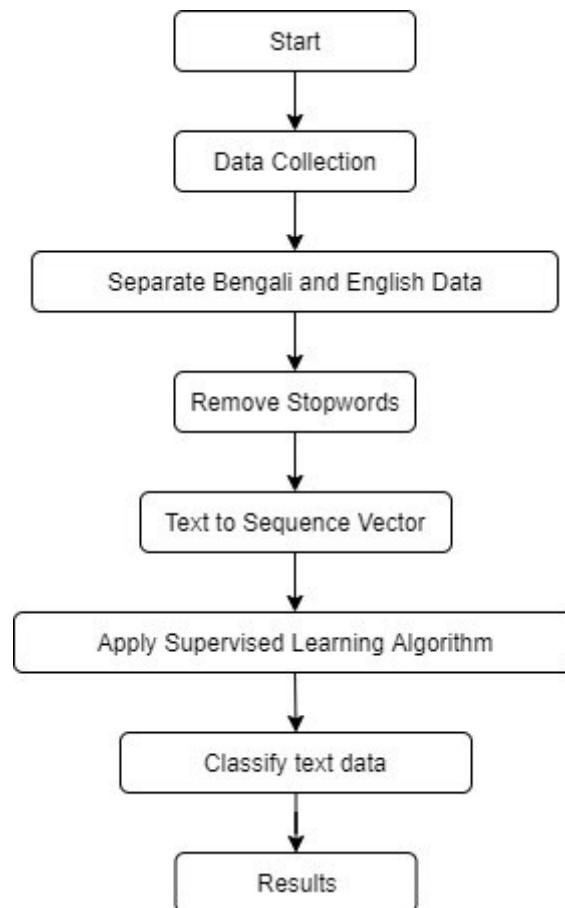
Figure 3.1- Research Model

3.5.1 Data Collection

For our research work data assortment was another level challenge as there are no open-source databases accessible on e-commerce product review dependent on Bangladeshi marketplace. We have to collect data in solitude. We used a scraping algorithm for gathering data from a Bangladeshi site called "Daraz Bangladesh". For scraping we used BeautifulSoup framework of Python programming language which gathers information from html code of a site. In any case, in "Daraz Bangladesh" site review data was wrapped by Json data format. So we need to accomplish some extra work on data collection.

3.5.2 Data Pre-processing

After data collection there were two sorts of information in our dataset. As Bangla/Bengali is native language of Bangladesh and English is International language. So in our dataset there were additionally two kinds of data which are Bangla and English. Separating them was another challenge for us. We used the TextBlob framework of Python programming language for separate Bangla and English data.

There were also other types of data, one of them is data with Bangla motive yet composed with the English word which aren't meaningful like "Items ta khub e valo chilo. Amar khub e pochondo hoyeche", and another type is Bangla-English blend data like "I like this shirt. Khub e fast delivery peyechi. Measurement o perfect". This sort of data confuses the system. So we have to expel that sort of data from our dataset.

3.5.3 Remove Punctuation and Capital Letter

In text data there are such a large number of punctuation and each line begins with a capital letter, which likewise confuses the system as the system thinks small letter one and capital letter on distinct. Be that as it may, the meaning of both are the equivalent. Also, punctuation doesn't add meaning to a sentence that is similar constantly. So we removed punctuation and capital letters from our training dataset.

Before removing punctuation and capital letter it looks like:

"This shirt is perfect. It's the same color as the photo displayed by the seller. And delivery was fast. Got it within 2 days. And this shirt feels premium. Thanks Seller for your great service. Keep it up. Hope to shop more with you guys."

After removing punctuation and capital letter it looks like:

```
import re
text = "This shirt is perfect. It's the same color as the photo displayed by the seller. And delivery was fast. Got it within 2 days. And t
letters_only = re.sub("[^a-zA-Z]", " ", text)
words = letters_only.lower()
print(words)
```

```
this shirt is perfect  it s the same color as the photo displayed by the seller  and delivery was fast  got it within  days  and this
shirt feels premium  thanks seller for your great service  keep it up  hope to shop more with you guys
```

Figure 3.2- Sk-Learn used for removing punctuation and capital letter

Here we used NLTK for removing punctuation and capital letters. After removing them we will remove stop words from our dataset as a part of data pre-processing.

3.5.4 Tokenization

For understanding the significance of the sentence in our dataset we have used sentence tokenizer. To distinguish the aggregate sum of word amount we used here word tokenizer of Natural Language Tool-Kit.

"this shirt is perfect its the same color as the photo displayed by the seller and delivery was fast got it within 2 days and this shirt feels premium thanks seller for your great service keep it up hope to shop more with you guys"

After word tokenizing it looks like:

```
import re
text = "This shirt is perfect. It's the same color as the photo displayed by the seller. And delivery was fast. Got it within 2 days. And t
letters_only = re.sub("[^a-zA-Z]", " ", text)
words = letters_only.lower().split()
print(words)

['this', 'shirt', 'is', 'perfect', 'it', 's', 'the', 'same', 'color', 'as', 'the', 'photo', 'displayed', 'by', 'the', 'seller', 'and',
'delivery', 'was', 'fast', 'got', 'it', 'within', 'days', 'and', 'this', 'shirt', 'feels', 'premium', 'thanks', 'seller', 'for', 'your',
'great', 'service', 'keep', 'it', 'up', 'hope', 'to', 'shop', 'more', 'with', 'you', 'guys']
```

Figure 3.3- Sk-Learn used for tokenizing

Here we used Sci-kit and NLTK for separating each sentence and word from one another. Presently those words are ready for using them on our training dataset. Also, our system will change those words into vectors.

3.5.5 Applying Algorithm

Our research depends on Supervised Learning algorithm in additionally it is a Classification Model problem. There are various kinds of tools for applying algorithms and procedures that are also unique. We used the Sci-Kit learn library which is quite different and popular from others.

Different algorithms additionally show different results. In this way, selecting the proper algorithm is additionally an indispensable undertaking. Here we have chosen the Multinomial Naive Bayes Algorithm, Logistic Regression and Long Short-Term Memory (LSTM). By using these Supervised algorithms, we made a classifier model.

3.5.6 Accuracy

We used Multinomial Naive Bayes Algorithm, Logistic Regression and Long Short-Term Memory (LSTM) algorithms for our research work. And tried to find out accuracy with different algorithms.

CHAPTER 4

EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Introduction

For increasing output results, our research project relies upon the selection of appropriate Data and the ideal Model. For constructing a Model using textual content dataset and finding the best possible outcome, the language pre-preparing part is exceptionally vital.

4.2 Experimental Results

To accomplish the expected result, we have used three different classification algorithms. The Long Short-Term Memory (LSTM) gives us the best accuracy which is around 74.07% for English Data and 68.92% for Bangla Data. In Table 4.2.1 represents the accuracy of Model performance for English and Bangla Data.

Table 4.1- Performance measurement accuracy of English and Bangla data using different algorithms

Number	Algorithms	Result for English Data	Result for Bangla Data
01	Multinomial Naive Bayes	73.6%	65.5%
02	Logistic Regression	74.1%	66.9%
03	Long Short-Term Memory (LSTM)	74.07%	68.92%

4.3 Descriptive Analysis

For measuring the overall performance, we have used cross validation. Cross validation is such a type of technique by cross validation it's possible to measure any kind of prediction model and also it can validate the result. As our goal is to predict the class whether it is Excellent, Good, Normal, Bad or Terrible that's why cross validation is much better effective way.

Table 4.2- English review text and the result

No	English Review/Data	Result
1	Very good sound quality and looking. I appreciate.	4
2	Sound Quality is not Good & Both the Headphone not working together	0
3	I ordered the black one but they gave me white other then that all ok	2
4	Low quality	1
5	Product is good at this price! I am satisfied	4
6	So far it's good. Color reproduction is ok. But the problem is making quality. It's not so good. Otherwise It's ok in this budget segment.	3
7	Buttons are stiff and hard to click. I don't know if it's a design error by the manufacturers or if I got a faulty one. Quite disappointed though.	2
8	Poor quality. broke with light handling.	0
9	Product is the same as the photo. Nice Case & Color. But color isn't permanent. As per product price color should be permanent.	3
10	The product does not work properly it needs to be set up.	1
11	Rubber are missing as before	2
12	Product build quality & sound quality is very good	4
13	Product Doesn't match with the displayed product.	1
14	Very disappointing. When I pressurized the glass got broken.	0
15	Average product. Pice should be decreased. But I'm Satisfied.	3

Table 4.3- Bengali review text and the result

No	Bengali Review/Data	Result
1	অসাধারণ। এই দামের মধ্যে যে এত ভালো প্রোডাক্ট পাব আশাই করিনি!	4
2	দেখতে তো খারাপ না। দাম অনুযায়ী মোটামুটি ভালোই। কিন্তু বাটনগুলো কাজ করতে চায় না!	2
3	একদম বাজে। লাল টা চেয়েছি নীল টা দিয়েছে।	0
4	ডেলিভারি পেয়েছি খুবই দ্রুত। প্রোডাক্ট কোয়ালিটি খুবই ভালো। প্যাকেজিং টা ছিলো জাস্ট ওয়াও। বাট কন্ট্রোলিং পজিশন আর সিস্টেমটা অতো ভালো না। খেলতে সুবিধা না।	3
5	তেমন একটা আশানুরূপ নয়	1
6	এই দামের মদ্দে অনেক ভাল কাবার। চারজিং সাইটের ফিনিসিং টা আর একটু ভাল করলে ভাল হয়।	2
7	ছবিতে যেরকম আছে হুবহু ওরকম পেয়েছি। প্রোডাক্ট কোয়ালিটিও ভালো এবং প্যাকেজিং সুন্দর ছিলো। ঢাকার ভিতর ডেলিভারী পেতে ৫ দিন লেগেছে। ধন্যবাদ।	4
8	ভালো প্রডাক্ট। তবে সাইজটা একটু ছোট	3
9	বাজে আর খারাপের একটা সীমা রেখা আছে কিন্তু এটার সেটা বলবার মতো রুচি নাই।	0
10	খুব বেশি দিন ইউজ করতে পারি নাই	1
11	সবই ঠিক আছে শুধু ভাইব্রেশন মোটর নাই!	3
12	খুবই হতাশাজনক। আপনাদের পেশাদারীত্ব দেখে খুবই হতাশ হলাম।	0
13	বেশ ভালো। আমি সন্তুষ্ট।	4
14	সস্তার তিন অবস্থা। খুব বেশি একটা ভালো না।	1
15	ভালো লাগছে। লোকাল মার্কেটের চেয়ে প্রাইস বেশি।	2

CHAPTER 5
SUMMARY, CONCLUSION, RECOMMENDATION AND IMPLICATION FOR FUTURE RESEARCH

In This part of our Report we have depicted some significant parts of our research which are the brief summary of research, recommendation, implication of the research work and and at long last the Future Scope of this Research. As a matter of fact, part by part all the segment has been portrayed underneath:

5.1 Summary of the Study

To complete our Research Work, we have concentrated how the Machine learning procedures are used for getting the solution of different sorts of AI issues. There are two types of Machine learning forms that are Supervised Machine Learning and Unsupervised Machine learning. This Study makes us realize that our research is about Supervised Machine learning and it is additionally a Classification Model Problem.

For doing research based on Machine learning we need a decent number of data. In any case, there is no open-source database dependent on Bangladeshi online marketplace or in Bangla language. So we have to collect our database in solitude. For getting our required database we used a scraping technique using BeautifulSoup framework of Python Programming Language.

We also have learned Machine Learning, Model Classification and the related Algorithms. As our training dataset is all literary textual documentation so we concentrated on how the Natural Language Processing is used for the procedure of textual documentation.

There are numerous tools and techniques which are appropriate for this area. We have additionally discovered which procedure is better for completing our research work. We have used various types of tools to finish our research work. We have used a few kinds of tools and techniques to achieve our goal. We profoundly found out about the BeautifulSoup framework, NLTK and Sci-pack learning library and Python programming language. In

addition, we took in the Cross Validation for the better Performance Measuring and furthermore how to import this to the Sci-Kit learn libraries.

5.2 Conclusions

Machine Learning is generally an excellent area for research. Based on our nation's marketplace data like user's opinion classify from e-commerce review dataset has not yet been done. We tried to build up a model which can predict Bangla and English textual information which is given by a user. By this, the organization or the owner of the business can take some steps for the betterment of their products and services. What's more, customers can gain informed knowledge before purchasing or taking products and services. Basically, we used all the reviews from real life e-commerce web-page which has been given by verified and valid clients. This review expresses what they feel about the products or services that are taken by them.

The important thing is that all of the data which we collected are legitimate and verified because the people who are buying the products or services, gave the review of what they feel about the product or service. We tried to describe all the working strategies, working methodology, models and furthermore the method with table and Figure. We have a plan to make this research finished in extra requirements additionally. All the Procedures have been huge in total so it required a lot of time to understand and put in force this in our studies. For collecting the data from the legitimate source, we went with the hassle. We have encountered a couple of various issues that were in the start of our research.

5.3 Recommendations

Despite the fact that we mentioned some related works but the measure of work is not very many. Also, there is no research work like us based on Bangla review data or Bangladeshi review data. We understood all their research procedures and work style after that we began to set our own research goal. After a hard exertion by doing the entirety of the work bit by bit at long last, we are at a phase that can be said it is our expected research goal. Along these lines, for making this sort of research work it needs huge work for guiding us through the right path of research. We have experienced some different issues that were inside the beginning of our research. We have also stuck with the Data collection, Separate Bangla and English Data and removing data which are written with english alphabet but expressing Bangla sentiment.

5.4 Implication for Further Study

From this research work, it could be done Latent Dirichlet Allocation (LDA) topic modeling by using which organizations and owners of e-commerce businesses can work for the improvement of their products or services. Additionally, customers can gain informed knowledge before buying or taking a product or service. We will be able to rank the e-commerce sellers or sites by means of their review data. We already built a Web Crawler for this research work. Using this crawler and by modifying this we can also collect a review dataset from many Single Vendor e-commerce sites and Multi Vendor e-commerce sites of Bangladesh. After collecting this data, it will be stored in a database. Then the further procedure is to pre-process that data and categories all the data using our proposed model. Then compare it to other e-commerce sites' data so it's possible to make a sorted rank. What's more, hope users will be capable of using this system by using Website and Android Mobile Application.

REFERENCES

[1] Almjawel, Aljoharah, et al. "Sentiment Analysis and Visualization of Amazon Books' Reviews." 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS). IEEE, 2019.

[2] Chowdhury, Shaika, and Wasifa Chowdhury. "Performing sentiment analysis in Bangla microblog posts." 2014 International Conference on Informatics, Electronics & Vision (ICIEV). IEEE, 2014.

[3] Haque, Tanjim Ul, Nudrat Nawal Saber, and Faisal Muhammad Shah. "Sentiment analysis on large scale Amazon product reviews." 2018 IEEE International Conference on Innovative Research and Development (ICIRD). IEEE, 2018.

[4] Gautam, Geetika, and Divakar Yadav. "Sentiment analysis of twitter data using machine learning approaches and semantic analysis." 2014 Seventh International Conference on Contemporary Computing (IC3). IEEE, 2014.

[5] Valdivia, Ana, M. Victoria Luzón, and Francisco Herrera. "Sentiment analysis in tripadvisor." IEEE Intelligent Systems 32.4 (2017): 72-77.

[6] Batool, Rabia, et al. "Precise tweet classification and sentiment analysis." 2013 IEEE/ACIS 12th International Conference on Computer and Information Science (ICIS). IEEE, 2013.

[7] Neri, Federico, et al. "Sentiment Analysis on Social Media." ASONAM 12 (2012): 919-926.

[8] Hodeghatta, Umesh Rao. "Sentiment analysis of Hollywood movies on Twitter." 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2013). IEEE, 2013.

Sentiment Analysis on Bangladeshi E-Commerce Review