

**Sentiment Analysis on E-Commerce Business in Bangladesh Perspective**

**BY**

**Khandker Ashik Mahmud**

**ID: 161-15-6889**

**AND**

**GM Taium Ahmed**

**ID: 161-15-6888**

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

Supervised By

**Md Zahid Hasan**

Assistant Professor

Department of CSE

Daffodil International University



**DAFFODIL INTERNATIONAL UNIVERSITY**

**DHAKA, BANGLADESH**

**DECEMBER 2019**

## APPROVAL

This Project titled “Sentiment Analysis on E-Commerce Business in Bangladesh Perspective”, submitted by Khandker Ashik Mahmud (ID:161-15-6889) and GM Taium Ahmed (ID:161-15-6888) 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 BSC in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 5th December, 2019.

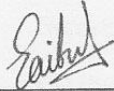
### BOARD OF EXAMINERS



---

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

**Chairman**



---

**Saiful Islam**  
**Senior Lecturer**  
Department of Computer Science and Engineering  
Faculty of Science & Information Technology  
Daffodil International University

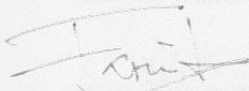
**Internal Examiner**



---

**Shaon Bhatta Shuvo**  
**Senior Lecturer**  
Department of Computer Science and Engineering  
Faculty of Science & Information Technology  
Daffodil International University

**Internal Examiner**



---

**Dr. Dewan Md. Farid**  
**Associate Professor**  
Department of Computer Science and Engineering  
United International University

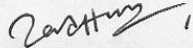
**External Examiner**

## DECLARATION

I am declare that, this project has been done by me under the supervision of **Md Zahid Hasan, Assistant Professor, Department of CSE, Daffodil International University.**

I 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 Zahid Hasan**

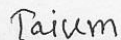
**Assistant Professor and Coordinator of MIS**

Department of Computer Science and Engineering

Faculty of Science & Information Technology

Daffodil International University

**Submitted by:**



---

**GM Taium Ahmed**

ID: 161-15-6888

Department of CSE,

Daffodil International University



---

**Khandker Ashik Mahmud**

ID: 161-15-6889

Department of CSE,

Daffodil International University

## ACKNOWLEDGEMENT

First of all, our heartiest thanks and gratefulness to Almighty Allah for His divine blessing that makes us capable to complete this project successfully.

We would like to thanks to our honorable teacher & project supervisor **Md Zahid Hasan, Assistant Professor, Department of CSE**, Daffodil International University for his endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior draft and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to **Dr. Syed Akhter Hossain**, Head, Department of CSE, for his kind help to finish our project and we are also thankful to all the other faculty and staff members of our department for their co-operation and help.

We must acknowledge with due respect the constant support and patients of our parents.

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

## **Abstract**

In this extend, we propose a framework that relegates scores showing positive or negative to English reviews additionally viewpoint base analyze on Bangla surveys. Numerous works have been done on opinion investigation, report clustering for aspect base in English. We are getting to do the same for Bengali dialect for E-commerce commerce audits. As there's no dataset of audits E-Commerce, we are utilizing web crawler to induce necessary reviews to form a dataset to use for this extend. To the most excellent of our information, there's many investigate on the aspect-based estimation examination (ABSA) of Bangla content. This will be depicted as being due to the require of available datasets for ABSA. In this paper, we offer one freely accessible datasets to perform the ABSA assignment in Bangla. We as well delineate a design approach for the subtask of viewpoint category extraction to evaluate our datasets.

**Key Words:** Sentiment Analysis, ABSA Bangla Sentiment analysis-Commerce reviews

# **TABLE OF CONTENTS**

<b>CONTENTS</b>	<b>PAGE</b>
Board of Examiners	I
Declaration	II
Acknowledgment	III
Abstract	IV
<b>CHAPTER 1: INTRODUCTION</b>	<b>1-4</b>
1.1 Introduction	1
1.2 Objectives	2
1.3 Motivation	2
1.4 Rational of the Study	3
1.6 Expected Outcome	3
1.7 Layout of the Report	3-4
<b>CHAPTER 2: BACKGROUND</b>	<b>5-7</b>
2.1 Introduction	5
2.2 Related Works	5-6
2.3 Research Summary	6-7
2.4 Challenges	7

<b>CHAPTER 3: RESEARCH METHODOLOGY</b>	<b>8-19</b>
3.1 Introduction	8
3.2 Data Collection	8-9
3.3 Data Formatting	10
3.4 Data Tagging	11
3.5 Data Classifier	11
3.6 Work Flow	12-13
3.7 Flow Chart	14
3.8 Machine Learning Algorithm	15-19
<b>CHAPTER 4: EXPERIMENTAL RESULTS AND DISCUSSION</b>	<b>20-27</b>
4.1 Introduction	20
4.2 Performance Evaluation	21-27
<b>CHAPTER 5: SUMMARY AND CONCLUSION</b>	<b>28</b>
5.1 Conclusion	28
5.2Future Works	28
<b>REFERENCES</b>	<b>29-30</b>

# LIST OF FIGURES

<b>FIGURES</b>	<b>PAGE NO</b>
3.1 -Workflow	12
3.2 -Flowchart of sentiment analysis	14
3.3- SVM	19
4.1- CNN	21
4.2 - Max Pooling	21
4.3 - Result of Model (Bangla)	23
4.4 - Countvectorization	24
4.5 - Confusion Matrix	25
4.6 - Result of Model (English)	27



# LIST OF TABLES

<b>TABLES</b>	<b>PAGE NO</b>
3.1 - English Reviews Dataset	8
3.2 - Bangla Review Dataset	10
4.1 - SVM	22
4.2 - RF	22
4.3 - KNN	23
4.4 - LR	26
4.5 - MNB	26
4.6 - BNB	26

## **Abbreviations**

NLP- Natural Language Processing

SVM – Support Vector Machine

CSV – Comma Separated Values

MNB - Multinomial Naïve Bayes

BNB - Bernoulli Naïve Bayes

ML - Machine Learning

ABSA - Aspect Base Sentiment Analysis

RF- Random Forest

KNN- K nearest Neighbor

CNN – Convolutional Neural Network

# CHAPTER 1

## INTRODUCTION

### 1.1 Introduction

World Wide Web (www) has ended up the foremost prevalent communication stages for open audits, suppositions, and comments. Presently a days we are utilizing innovation for everything. In our world where 2.5 quintillion bytes of information is produced each day[1],opinion examination has ended up a fundamental instrument for making sense of this information. Estimation analysis is characterized as the method of mining of information, sentence to anticipate the feeling of the sentence or perspective base analyze of the sentence through Natural Language preparing (NLP). The estimation examination make classification of content into three stage Positive, Negative or Impartial. For our work, we are centering on E-Commerce Trade of Bangladesh.

Sentiment is analyzed Three Level: First one is document level, second one is sentence level, and the another one is aspect level[15].The document level incorporates a estimation on an entity, and the work is to classify whether a complete archive communicates a positive or negative estimation. The work at the sentence level consider sentences and classifying each sentence positive, negative, or unbiased supposition. Not one or the other the record level nor the sentence level investigation find precisely what individuals preferred and did not just. Like the aspect-based sentiment analysis—(ABSA) aspect level performs a finer-grained examination that distinguishes the aspect of a given archive or sentence and the estimation communicated towards each aspect. This level of investigation is the foremost point by point form that's able of finding complex suppositions from surveys.

For our project, we have chosen to analyze the assumption of E-Commerce business in Bangladesh. Bengali is one of the most noteworthy talked dialect, positioned seventh within the world, but shockingly exceptionally few works had been done with perspective based Bengali dialect on estimation examination.

## 1.2 Objective

In our digital world we are using technology everywhere. In the field E-Commerce Business without technology e-commerce business is nothing. Our main objective is to classify Bangla comments aspect based and English is to find positive and negative. Our objective is to supply benchmark datasets for Bangla ABSA. Our datasets are planned for two major assignments of ABSA. These are perspective category extraction and the distinguishing proof of extremity for each viewpoint category. In this paper, we tested with the primary subtask, that's, the extraction of the perspective category. We connected three major steps to extricate the perspective category. Firstly, preprocessing was performed on the dataset. After this, we extricated highlights from the data and at long last performed classification utilizing a few well known classification models. We are going to collect the information of buyer conclusion on e-commerce business site in Bangladesh. At that point utilize common dialect handling and machine learning calculation on the information. At that point we'll examine the assumption around shopper supposition. That will help product development both sellers and buyers. So we describe our goal:

- English sentence classify positive and negative
- Bangla sentence aspect based

## 1.3 Motivation

Our work is mainly inspired by Md. Atikur Rahman and Emon Kumar Dey work. They classified restaurant and cricket comment into five aspect based sentiment analyze. In our work, we are going to implement SVM, Random Forest and Knn. And for English comment we will do Logistic Regression and naïve Bayes. We were interested to do something different. So we decided that we will try to do research on Artificial Intelligence (AI) and Machine Learning (ML) field and then we started to search for some ideas. As we are online shopping lover, one day we thought that we can do something on E-Commerce Business. We live in an age of modern science. Now we can order everything what we need to stay at our home. In future very few people will go out to buy their needs. So it will need to develop e-commerce business

site. For this purpose we will analysis about their sentiment. So that both buyers and sellers are understand easily which product and aslo which service is good or bad.

#### **1.4 Rational of the study**

There is no doubt that there are thousands of works done on Natural Language Processing (NLP). But there are only a few works done on Bangla aspect based analyze. So our work is a new approach using different algorithms and simulation. To develop more efficient classifier application in the field of E-Commerce Business we give out best effort to develop our own model.

#### **1.5 Expected Outcome**

In this section there is some points given that points was our min expected outcome. Expected outcome of this research based project is to build an algorithm or making a complete efficient procedure that will categorize comment with respect to the built model of trained dataset. We create the dataset and train the dataset through different types of algorithm. And then we will find the best accuracy of the model.

#### **1.6 Layout of the Report**

In our research paper we gather knowledge and share knowledge on our ability. In our paper introduction we described introduction, objective, motivation, rational of the study, expected outcome and layout of the report.

The second chapter we will describe Background and it contains introduction, related work, research summary and challenges.

The third chapter is research methodology. It contains research subject, data collection , data pre processing , different types of algorithm .

Chapter four provides the experimental results, performance evaluation and result discussion. Some experimental pictures are presents in this chapter to make realize the project.

Chapter five discussed with summery of the study, future work and conclusion. This chapter is responsible to show the whole project report adhering to recommendation. The chapter is closed by showing the limitations of our works that can be the future scope of others who want to work in this field.

## **CHAPTER 2**

### **Background Study**

#### **2.1 Introduction**

Sentiment analysis called opinion mining, is the field of ponder that investigate people's consule, audit, comment and feelings through protest such as items, administrations, organizations, people and other properties. It speaks to a colossal issue space..In this section that will contain related works, research summary and challenges about this research. In Related works part that will contain about research paper and their various works, their methods, and accuracy which are connected to our work. In research summary section we will give the summary of our related works. In challenges part, we will discuss what our challenge is.

#### **2.2 Related Works**

In this modern period of computerized world, we are more social in social media, web, online shopping etc. So that specifically or by implication online conclusions or surveys are picking up most consideration. But the genuine bargain is analysis or mining of conclusions. For the investigation SA (sentiment analysis) is important.

His are some related works from where we set up the thought of this subject. The dataset of the restaurant review, given by Ganu et al. [8], was utilized to form strides rating estimates. They had not orchestrated add up to ABSA dataset, as the point category was show but the comparing limit of that perspective was truant. The SemEval 2014 evaluation campaign [4] increased their dataset by counting three more ranges with the perspective category. They conveyed their dataset with four regions being contained for each study, that's, with the perspective term happening inside the sentences, the point term's limit, the viewpoint category, and the perspective category's limit. They as well given a laptop-review dataset and physically commented on with comparable substances as for the eatery dataset. These are the benchmark datasets that [9–12] examines have utilized for performing the ABSA task.The task was reiterated in SemEval 2015 [13], for which viewpoint

categories were the combination of the substance sort and an quality sort.. Multilingual datasets were discharged interior the SemEval 2016 workshop [9] on the seven spaces (eatery, flexible workstation, adaptable phone, and advanced camera, lodging, and appear passage) and in eight tongues A book-review dataset interior the Arabic lingo was given by [5]. They commented on book audits into 14 categories and 4 sorts of polarities, checking “Conflict”. In [6], the creator made an IT product-review dataset for the ABSA assignment, in which a incorporate up to 2200 surveys were contain.

### **2.3 Research Summary**

Sentiment analysis is the portion of natural language processing. And natural language is the portion of Machine Learning (ML).Here I will attempt to depict Bangla sentence viewpoint based analyze and English sentence level which is positive or negative. Convolutional Neural Network (CNN) is one of the most grounded systems in profound learning. It is a manufactured neural organize, which could be a known as feed-forward ANN. In a “feed-forward” arrange data streams right through the networks. CNN is exceptionally fruitful models in content classification.

Within the Text Classification Issue, we have a set of texts and their particular names. But we straightforwardly can't utilize text for our demonstrate. You would like to change over these content into some numbers or vectors of numbers. Bag-of-words model (Bow ) is the best way of extricating highlights from the content. Bow changes over content into the lattice of event of words inside a report. This show concerns approximately whether given words happened or not within the document.

CNN's classification precision is way better than any other content classification calculations. There are distinctive sorts of layers that are utilized in CNN such as Thick, Implanting, GlobalMaxPooling1D, Conv1D, Dropout, and LSTM.

When performing ABSA have two primary portion. The primary portion is to extricate the particular areas or perspectives specified within the opinioned audit. The moment portion is to distinguish the extremity (which is positive, negative, or impartial) for each viewpoint. For case, the taking after survey of a online shopping



uncovers two perspectives: product and price. Both viewpoints have a positive extremity.

“The service was excellent and the product also good.”

And for English information we to begin with do preprocessing the information and from the cleaned dataset, potential highlights are extricated and are changed over to numerical organize. The vectorization strategies are utilized to change over printed information to numerical organize.

## **2.4 Challenges**

The main challenges of this work is collecting and processing the dataset, dealing with the data set was too hard. To clean and normalize we used several steps and methods. After all training with many layers with different size of epoch took long time in our machine, so getting the final output we waited so much with keeping patience. There was not another dataset or resources regarding this paper domain. There was not enough work done before so we have to start from our own motivation.

## CHAPTER 3

### RESEARCH METHODOLOGY

#### 3.1 Introduction

In this section we are going to elaborate the workflow of our novel approach to classify sentence. There are some key point like data collection, processing, proposed model also described with relevant equation, graph, table and description. The chapter is being closed by giving the clarification of our project's measurable speculations and other than, giving the clear concept of the usage necessities.

#### 3.2 Data Collection

For our work, we used the comments of people from facebook.com and many E-Commerce website. No dataset of comment existed at that time, so we needed to create our own dataset. We used web crawler to get the source html script of the website of facebook.com. We pulled html script programmatically of all the post of Facebook, also we used regular expression to extract the comment from the source html script of the site. We saved the comment in a excel file.

Table 3.1 Sample of our English Comment Dataset:

Review	Liked
Wow... Fully satisfied.	1
Facing a one problem, There is no headphone.	0
Quick Delivery.	1
There is no huwai phone.	0
Special price shoeing high but during giving order it's not working....showing high...how can i get the special price.	1
If there is an indication of cash on delivery is available on the top right page of the product under delivery options.	0

Price Please.	1
Does it contain official warranty?	1
Warranty Policy EN... 1 Year Manufacturer Warranty.	1
Best in price.	1
Not Good.	0
I have purchsed three items but not working.	0
Working fine...Really appreciate for the lowest price ...good packaging and rapid delivery...Recommend both product and the seller.	1
Does not work... Very low quality.	0
Bad Product.	0
Works fine... Take care about packing more.	0
Highly recommended.	1
Original product... I am satisfied.	1
Good product as a price.	1
Good.	1
Product quality, shipping time, packaging etc... All good.	1
The product, amazing.	1
Service is also cute.	1

Table: 3.2. Sample For Bangla Comment Dataset

Text	Category	Polarity
সেবাকিন্তু, আমাদেরজন্যভয়ঙ্করছিল।	service	Negative
যাইহোক, পণ্যগুলোবেশভাল।	Product	Positive
তারাএতদ্রুতকিনছেনযে,তারাসাহায্যকরতেপারবেনা।	anecdotes/miscellaneous	Neutral
আমরাখুবহতাশছিলাম	anecdotes/miscellaneous	Negative
এটিসুনির্দিষ্টভাবেবিশেষআফগানিস্তান।	anecdotes/miscellaneous	Positive
এটিসুনির্দিষ্টভাবেবিশেষআফগানিস্তান।	Price	Positive
আমিসত্যিইএইজিনিসটাপছন্দকরেছি।	Product	Positive
এটিরমানএকটিযুগ্মমাননয়।	Price	Negative
যদিওসেবাটিদরিদ্র,মূল্যযুক্তিসংগত	Price	Positive
যদিওসেবাটিদরিদ্র,মূল্যযুক্তিসংগত।	Service	Negative
চমৎকারপণ্যতালিকা	Product	Positive

### 3.3 Data Formatting

We saved the comment along with their respective aspect and sentiment in our dataset. We separated the dataset for English and Bengali.

### **3.4 Data Tagging**

To train our models we needed manually tagged data. To make the process of data tagging easier we wrote a code that automated the whole data tagging process. It read the comment one by one and tagged those comment positive, negative or neutral and also its aspect classification. After that, we took all the month-by-month text files and comma separated the comment, aspect and tag. This process was also automated by writing a piece of code. We then saved all the data in one csv file, and this csv file worked as our main dataset.

### **3.5 Classifier Selection**

There are numerous machine-learning classifiers for content classification. After investigating, For English comment we chosen to utilize Multinomial Naive Bayes, Bernoulli naive Bayes, and Logistic Regression. Logistic Regression are parallel classifiers, they are way better suited in classifying extremity of a sentence. Since our work is to distinguish between positive and negative which is like twofold classification and Logistic Regression works superior for it. Multinomial Naive Bayes classification calculation tends to be a standard arrangement for opinion examination assignment. The fundamental thought of Naive Bayes procedure is to discover the probabilities of classes' allotted to writings by utilizing the joint probabilities of words and classes. . We needed to see how it performs against Logistic Regression and Naïve Bayes. And for Bengali comment we classify angle based and here we utilize SVM (support vector machine), RF (Random Forest), KNN (K Nearest Neighbor). Here we too utilize CNN (Convolutional Neural Network) and compare all those calculation which is way better.

### 3.6 Work Flow

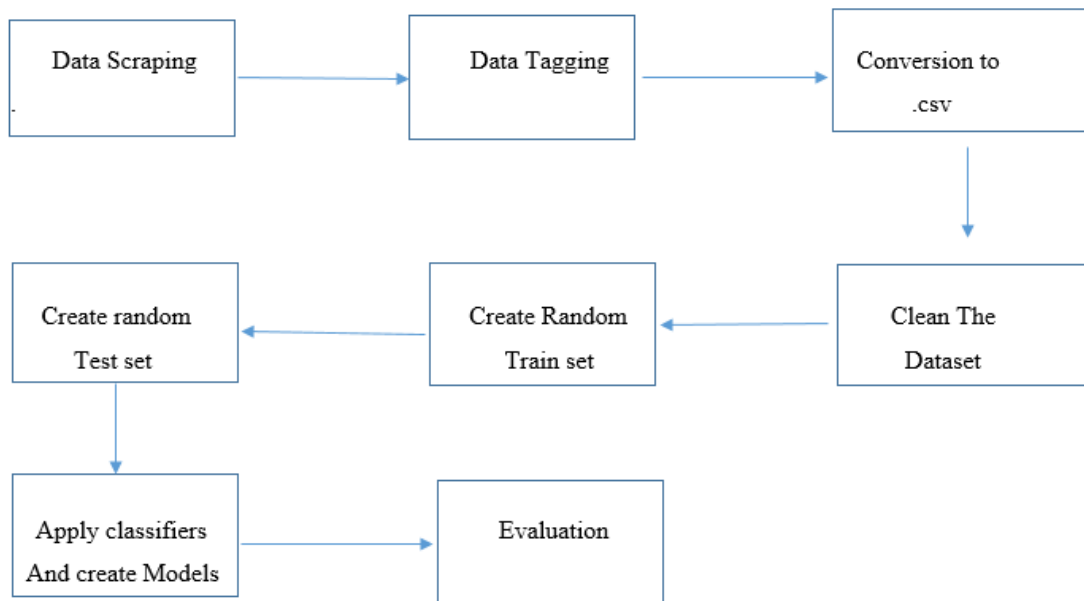


Figure 3.1: Workflow of our approach

Step 1 – Data Scrapping: We collect the data from facebook.com and E- commerce website through data scrapping.

Step 2- Data Tagging: To train our models we needed manually tagged data. To make the process of data tagging easier we wrote a code that automated the whole data tagging process.

Step 3- Conversion to .csv or .xlsx format: To read our dataset from code we make the dataset .csv or .xlsx format.

Step 4-Clean the dataset: IN our code use removing the Stopwords, numeric and special characters and Normalizing each review using the approach of stemming.

Step 5- Create Random Train set: We create random train set to train our model.

Step 6- Create Random Test Set: we create random test set what is the machine performance.

Step 7- Apply Classifier and create models: To train and validate our data for better accuracy we choose out model. There are hundreds of convolutional neural networks. To get better accuracy with our machine configuration we implement few model and finally one model was selected for final training and testing process.

Step 8-Evaluation: In this section, all the results have been discussed. After training and testing those process gave us few accuracy with validation loss and accuracy. We also calculated the confusion matrix and a table for showing the precision, recall and f1 measure.

### 3.7 Flow chart

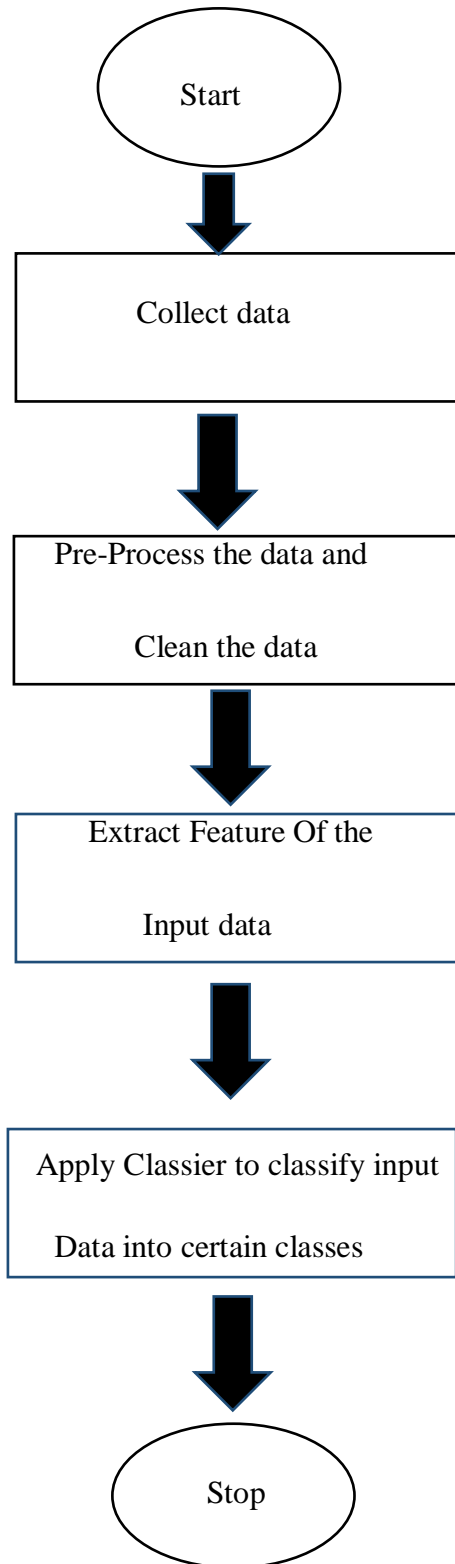


Figure 3.2: Flow Chart of Sentiment Analysis



### 3.8 Machine Learning Algorithm

#### 3.8.1 The Multinomial Naive Bayes' Classifier

Document classification issue is most utilized for MNB, that archive has a place to the category of sports, audits, innovation etc. The predictors utilized by the classifier are the recurrence of the words display within the document. Multinomial Naive Bayes classification algorithm manage a standard arrangement for sentiment analysis assignment. The fundamental thought of Naive Bayes procedure is to discover the probabilities of classes doled out to writings by utilizing the joint probabilities of words and classes.

The algorithm that's measurable show we'll be utilizing is the multinomial Naive Bayes' classifier, a part of the Gullible Bayes' classifier family. NB classifiers are probabilistic classifiers, meaning that they utilize the probabilities of watched results to return a sensible gauge of an obscure result. At a tall level, NB classifiers utilize Bayes' run the show with one credulous (or disentangling) suspicion: that highlights are autonomous from/uncorrelated with others highlights. (For content classification, an case of a highlight is the event of a word. In any case, within the common case, the word highlights is utilized freely, since choosing what constitutes a highlight in ML may be a point of its claim.) In spite of the fact that this suspicion almost autonomy is unreasonable, in any case, NB classifiers perform well in hone. For this reason, and for being fast and basic, they are utilized regularly in NLP.

At a high level, Bayes' rule says, if we know the effect given the cause, we can calculate the cause given the effect.

$$\Pr(\text{cause} | \text{effect}) = \frac{\Pr(\text{cause})\Pr(\text{effect} | \text{cause})}{\Pr(\text{effect})} \dots\dots\dots (1)$$

### 3.8.2 Support Vector Machine(SVM)

SVM (Support Vector Machine) could be a machine-learning algorithm. SVM may be a directed learning show, which analyzes information for classification and relapse analysis. SVM construct a demonstrate utilizing preparing algorithm that allocates unused cases into one or two categories. SVM isolated categories as wide as possible by making a crevice. Unused applications categories hole mapped into that same space or crevice on which side the application fall on. The issue is when information are not appropriately named supervised learning isn't conceivable. At that point we need to take after an unsupervised learning approach to analyze, by which ready to isolate the information into partitioned bunches. It takes after a clustering approach that's called back vector clustering [16] and is frequently utilized in mechanical applications either when information isn't named or when as it were a few information is named.

Outlier:: An outlier is an perception point that's far off from other observations. An observation that's well exterior of the expected extend of values in a think about or try, and which is regularly disposed of from the dataset.

Hyper plane: In geometry, a hyper plane may be a subspace of one measurement less than its surrounding space. In the event that a space is 3-dimensional at that point its hyper planes are the 2- dimensional planes, whereas in case the space is 2-dimensional, its hyper planes are the 1- dimensional lines. This idea can be utilized in any common space in which the concept of the measurement of a subspace is characterized.

Assume, we have three hyper-planes (A, B and C). Presently, we got to recognize the correct hyper-plane to classify star and circle. We have to be keep in mind a thumb run the show to recognize the correct hyper-plane.

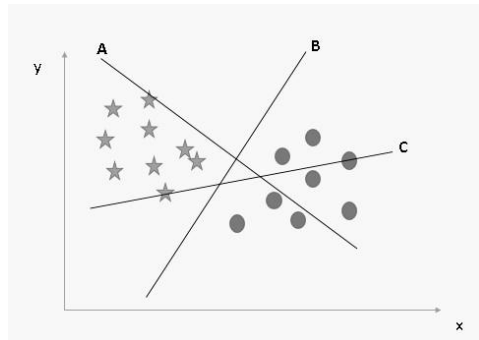


Figure 3.3: Hyper plane in scatter

Presently, here we have three hyper plane all are in scramble conceivable ways. Presently, how can we distinguish the correct hyper-plane from these?

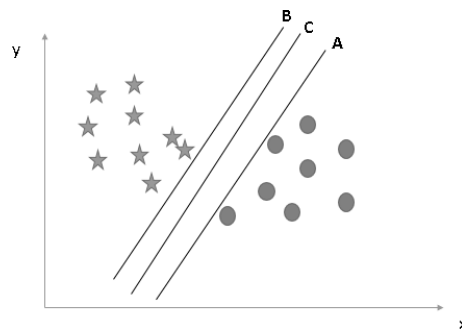


Figure 3.4: Creating Hyper Plan margin

Here, maximizing the separations between closest information point (either course) and hyper- plane will offer assistance us to choose the proper hyper-plane. This remove is called as Margin.

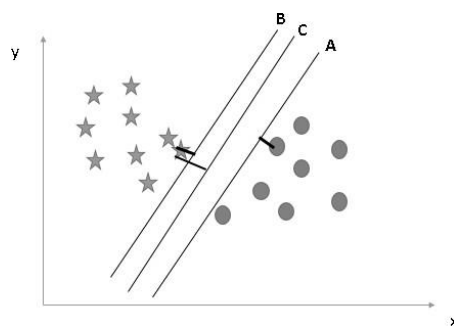


Figure 3.5: Finding right Hyper Plane

Over, you'll be able see that the edge for hyper-plane C is higher than A and B. Consequently, we title the proper hyper-plane as C. We have chosen the hyper-plane with higher. edge is strength [16], because if we select a hyper plane having moo edge at that point there's a tall chance is that the edge will be miss-classified.

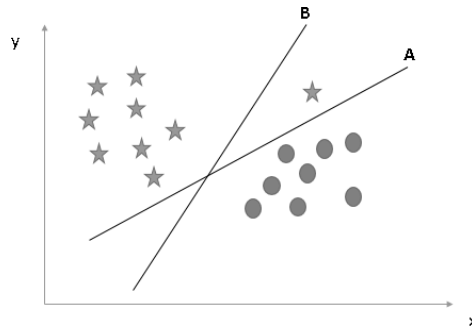


Figure 3.6: Correct Hyper Plane

Here, one might select B as the higher margin, but SVM selects the hyper plane, which classifies the classes accurately prior to maximizing margin. B has a classification error and A has classified all correctly. Therefore, the right hyper plane is A.

Below we cannot segregate two classes with a straight line, as one of the star is in the territory of other.



Figure 3.7: Hyper Plane with Outliers

SVM has a feature to ignore outliers and find the hyper plane that has maximum margin. Therefore, it can be said that, SVM is robust to outliers.

In the scenario below, a linear hyper plane between two classes is not possible.

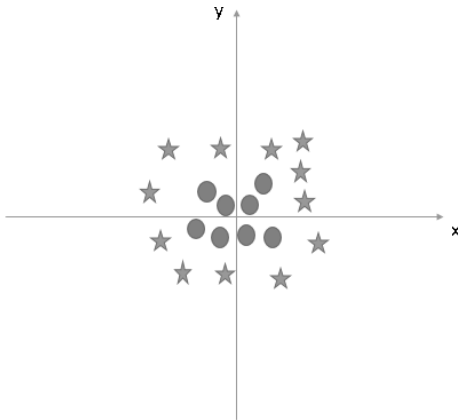


Figure 3.8: Hyper plane with maximum margin

This problem also can be solved by SVM. SVM introduces an additional feature  $z=x^2+y^2$ . If we plot the data points on axis  $x$  and  $z$ :

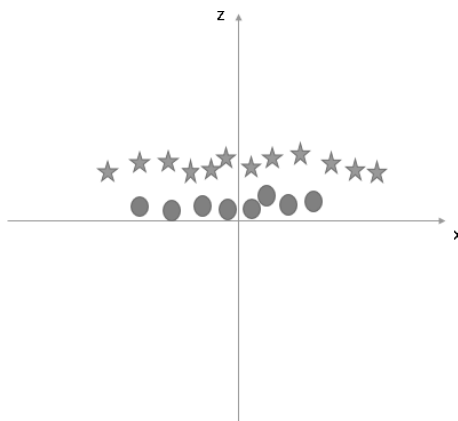


Figure 3.9: Solving with SVM

## CHAPTER 4

### EXPERIMENTAL RESULTS AND DISCUSSION

#### 4.1 Introduction

**Precision:** Within the field of data recovery, accuracy is the division of recovered archives that are pertinent to the inquiry:

$$Precision = \frac{tp}{tp+fp} \dots\dots\dots (3)$$

Precision is utilized with recall, the percent of all pertinent reports that's returned by the look. The two measures are sometimes used together within the F1 Score (or f-measure) to supply a single estimation for a framework. Note that the meaning and utilization of "precision" within the field of data recovery varies from the definition of accuracy and precision inside other branches of science and innovation.

**Recall:** Recall is the piece of important occasions that have been recovered over the full sum of important instances. High recall implies that an algorithm returned most of the important result.

$$Recall = \frac{tp}{tp+fn} \dots\dots\dots (4)$$

**F-measure:** f-score could be a degree of test's exactness by considering both precision and recall. it may be a consonant normal of precision and recall.

$$F - score = 2 * \frac{precision*recall}{precision+recall} \dots\dots\dots (5)$$

**Accuracy:** accuracy refers to the recognition of the measured value to a known esteem.

$$accuracy = \frac{tp+tn}{tp+tn+fp+fn} \dots\dots\dots (6)$$

## 4.2 Performance Evaluation

### 4.2.1 For Bangla ABSA dataset

Inside the preprocessing arrange, each Bangla report was talked to as a “bag of words”. We associated ordinary preprocessing steps for the evaluation. Firstly, accentuations and end words were cleared from each of the comments. After this, we cleared the digits from our dataset, since we found that digits were not fundamental for the perspective category. At final, we tokenized each Bangla word from our dataset. Hence, a dictionary of Bangla words was orchestrated after preprocessing.[15] We made a include framework for which each overview was talked to by a vector of that vocabulary. Interior the planning organize, removed consolidate sets were orchestrated by the well-known facilitated machine learning calculations. Since this was a multi-label classification issue, we organized our models by setting up multi-labelabandon.

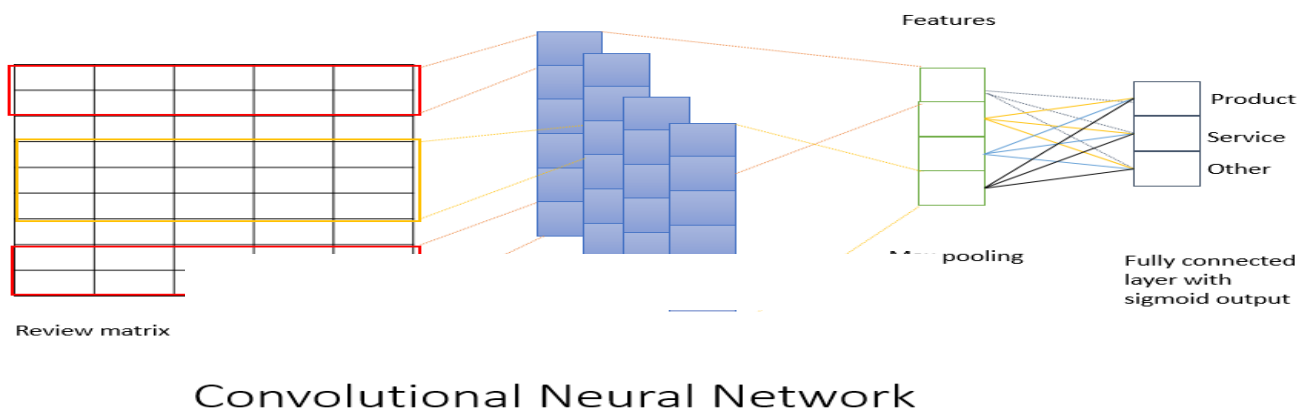


Figure 4.1 Convolution Neural Network

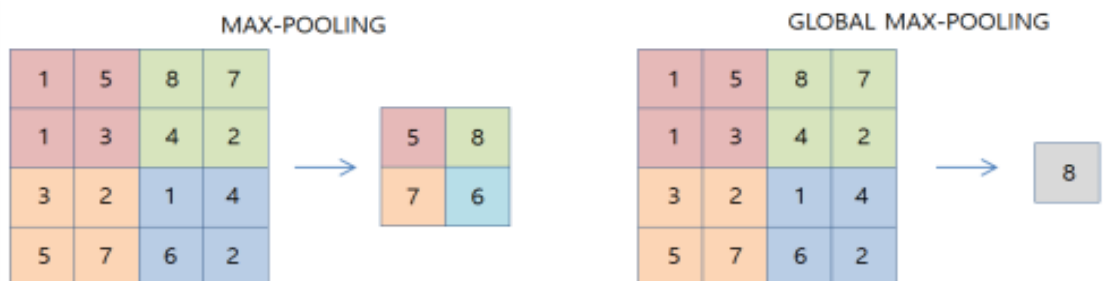


Figure 4.2 The difference of max-pooling and global max pooling

We utilized coordinate SVC inside the support vector machine (SVM) execution. The taking after machine learning calculations were utilized:

I. Support Vector Machine (SVM)

II. Random forest (RF)

III. K-nearest neighbor (KNN)

After the arranging was completed, our proposed Bangla test dataset was executed on the arranged illustrate. We are going see that utilizing the SVM, we gotten the first lifted exactness rate for both of the datasets. Both datasets showed up a moo study and F1-score. Appears up the in common precision of the models utilizing our datasets. The normal nature of the datasets is the reason behind the lower execution of the models for both datasets. People share their conclusion with their person judgment. In this way, the collection of suppositions interior the datasets is much more prominent. On the other hand, point extraction may be a multi-label classification issue. One's conclusion might have different point of view categories. Standard classifiers miss a few of these point of see categories.

Table 4.1 Support Vector Machine

Model	Precision	Recall	F1-Score
SVM	0.70	0.23	0.33
	0.60	0.38	0.47

Table 4.2 Random Forest

Model	Precision	Recall	F1-Score
RF	0.60	0.37	0.45
	0.59	0.60	0.60



Table 4.3 KNN

Model	Precision	Recall	F1-Score
KNN	0.42	0.52	0.46
	0.43	0.40	0.41

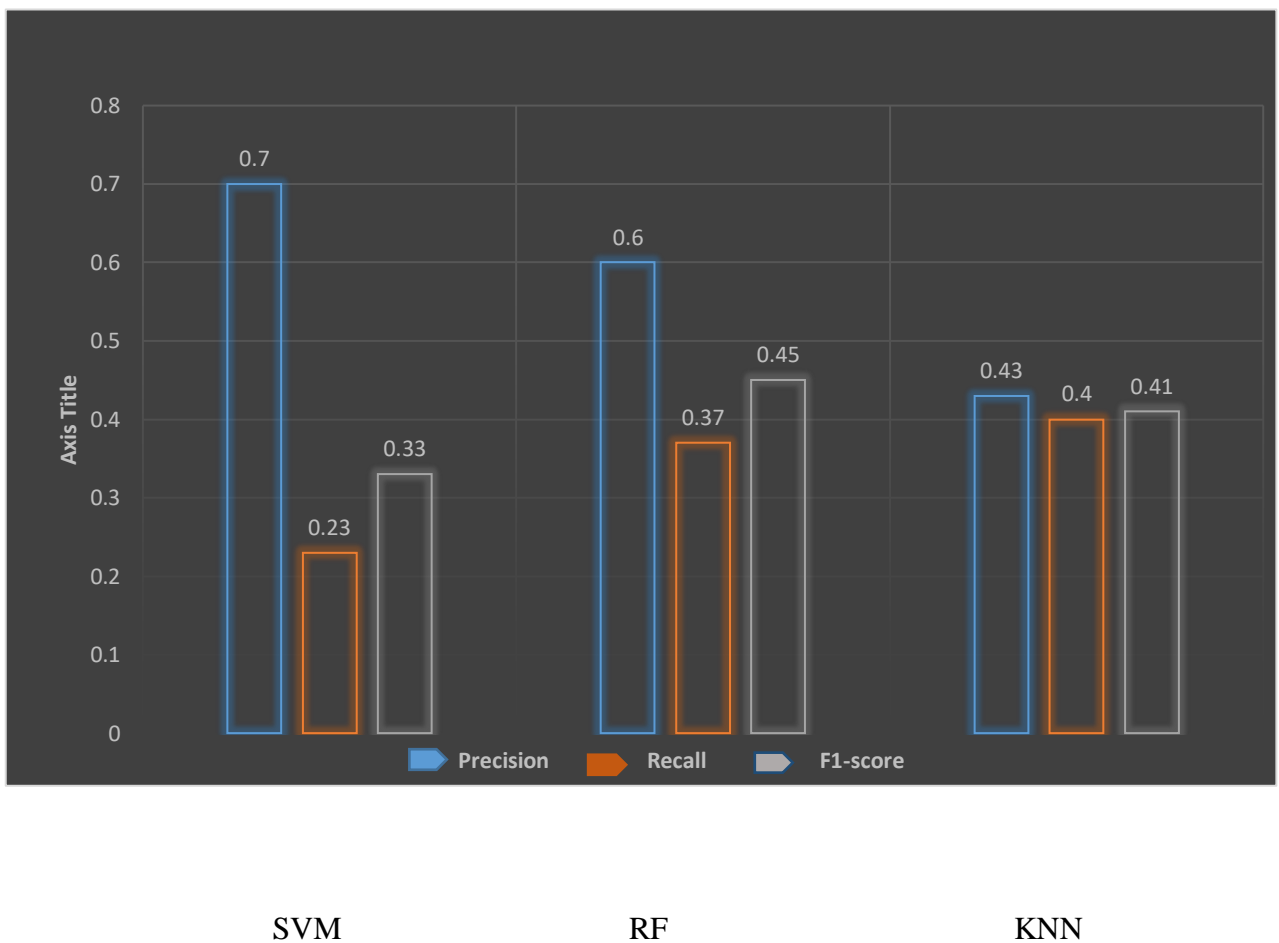


Figure 4.3.:The result of three models of our datasets

These comes about can be progressed on the off chance that we prepare and prepare the datasets in a more advanced way. In this work, we have taken all of the lexicon as highlights for the assessment after evacuating accentuation, halt words, and digits. A few state-of-the-art methodsfor data pick up can be connected to the dataset some

time recently classification and after the preprocessing steps to accomplish superior comes about.

#### 4.2.2 For English Reviews Dataset

Inside the Text Classification Issue, we have a set of writings and their specific names. But we clearly can't utilize content for our illustrate. You'd like to alter over these substance into a few numbers or vectors of numbers. To construct a demonstrate to foresee in the event that audit is positive or negative, taking after steps are performed.Importing Dataset

- Preprocessing Dataset
- Vectorization
- Training and Classification
- Analysis Conclusion

Bag-of-words show (Bow ) is perfect way">the most perfect way of removing highlights from the substance. BoW changes over substance into the grid of occasion of words interior a report. This appear concerns roughly whether given words happened or not inside the report.

Example: There are three archives:

Doc 1: This is good Doc 2: This is bad. Doc 3: This is awesomeNow, you can create a matrix of document and words by counting the occurrence of words in the given document. This matrix is known as Document-Term Matrix (DTM).

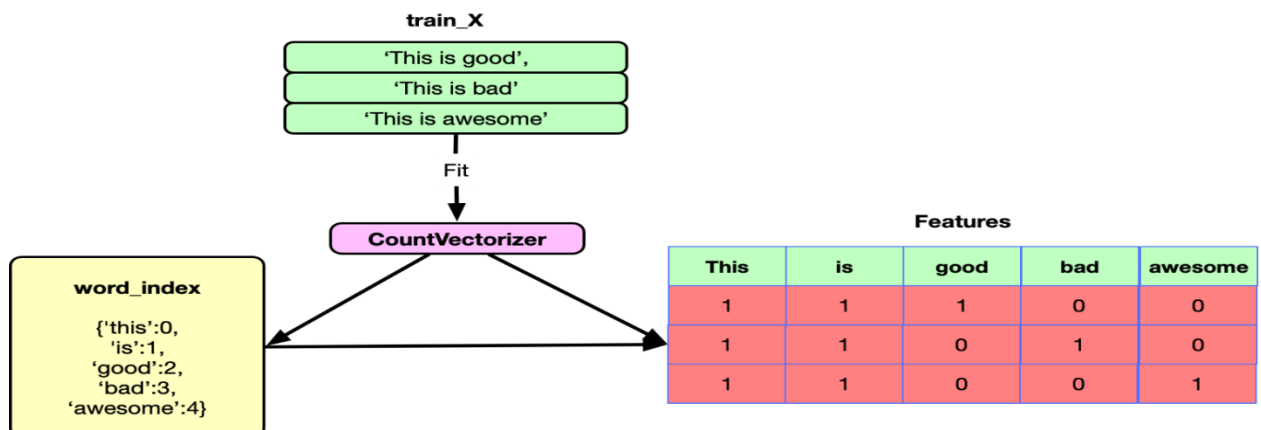


Figure 4.4: Countvectorization

### 4.2.3 Confusion Matrix

When we get the information, after information cleaning, pre-processing, the primary step we do is to nourish it to an extraordinary show and of course, get output in probabilities [17]. Confusion Matrix may be an execution estimation for machine learning classification. It is very helpful for calculating Recall, Precision, Accuracy.

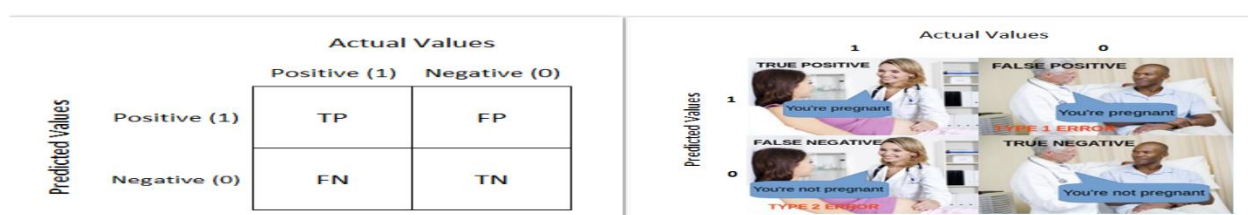


Figure 4.5: Confusion Matrix

#### True Positive

You predicted positive and it's true.

#### True Negative

You predicted negative and it's true.

#### False Positive

You predicted positive and it's false.

#### False Negative

You predicted negative and it's false.

TABLE 4.1: Logistic Regression:

N=301	Predicted Yes	Predicted No	
Actual Yes	TP -116	FP – 26	142
Actual No	FN – 46	TN – 113	159
	162	139	

Accuracy is 76.08 %

Precision is 0.81

Recall is 0.71

TABLE 4.2: MultinomialNB:

N=301	Predicted Yes	Predicted No	
Actual Yes	TP -102	FP – 40	142
Actual No	FN – 40	TN – 119	159
	142	159	

Accuracy is 73.42 %

Precision is 0.75

Recall is 0.75

TABLE 4.3: BernoulliNB

N=301	Predicted Yes	Predicted No	
Actual Yes	TP -98	FP – 44	142
Actual No	FN – 31	TN – 128	159
	129	172	

Accuracy is 75.08 %

Precision is 0.74

Recall is 0.81

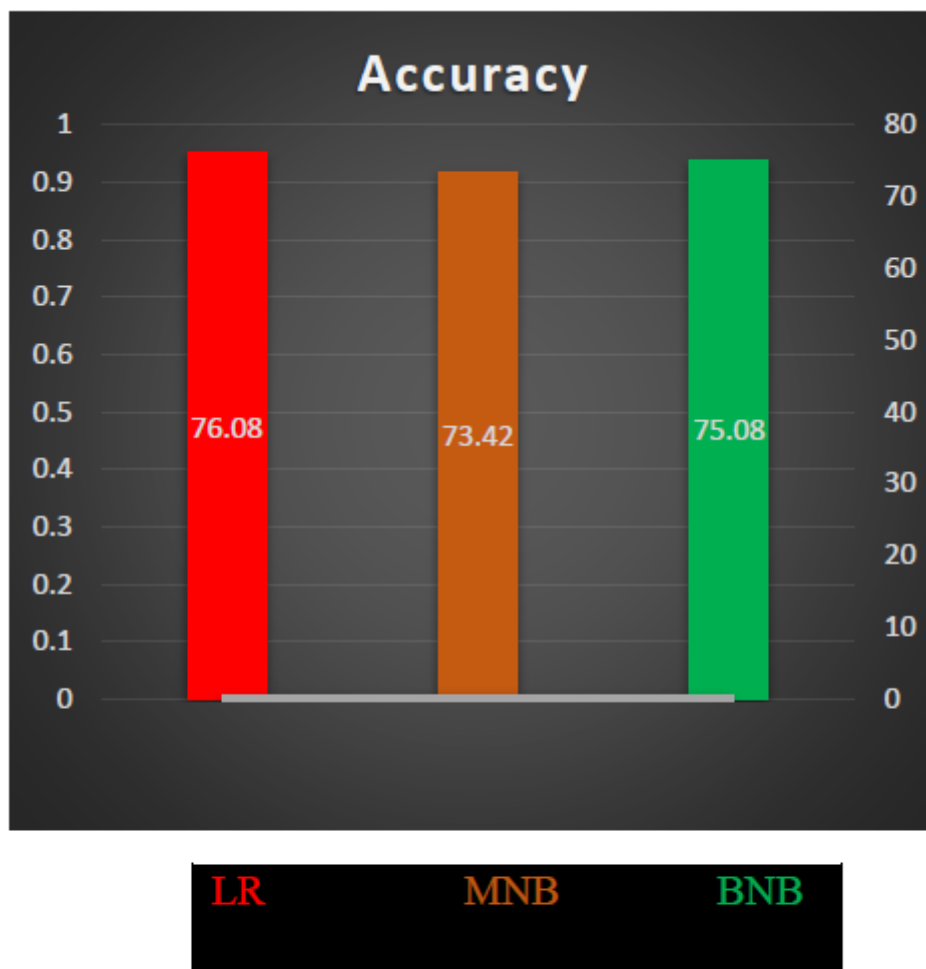


Figure 4.6: The result of three models of our datasets

## **CHAPTER 5**

### **Conclusion And Future work**

#### **5.1 Conclusion**

In our proposition work, we attempted to make a demonstrate which can foresee on the off chance that a survey is positive or negative for English and angle base Bangla sentence. We made the show with a few months surveys from a Bengali E-Commerce site and Facebook page. There was no dataset with the E-Commerce website reviews. Hence, we had to gather the information by scratching their site. After collecting the information, we labeled them physically. Two datasets are given one is ABSA of Bangla content and another one is English comment audits. ABSA datasets have been outlined to perform two assignments covering perspective category extraction and the recognizable proof of extremity for that viewpoint category. We too report pattern comes about to assess the errand of perspective category extraction.

#### **5.2 Future Works**

As future plans, we point to improve our work by counting assist spaces such as cars, mobiles, and portable workstations. We are working on more progressed strategies for the ABSA of Bangla content utilizing our datasets to realize way better performance.

## REFERENCES

- [1] Sentiment Analysis /MonkeyLearn, available at <<  
<https://monkeylearn.com/sentiment-analysis/>>> last accessed on 07-07-2019 at 1.00 AM
- [2] Trusov, M.; Bucklin, R.E.; Pauwels, K. Effects of word-of-mouth versus traditional marketing: Findings from an internet social networking site. *J. Mark.* 2009, 73, 90–102.
- [3] Jeyapriya, A.; Selvi, C.K. Extracting Aspects and Mining Opinions in Product Reviews Using Supervised Learning Algorithm. In Proceedings of the 2015 2nd International Conference on Electronics and Communication Systems (ICECS), Coimbatore, India, 26–27 February 2015.
- [4] Pontiki, M.; Galanis, D.; Pavlopoulos, J.; Papageorgiou, H.; Androutsopoulos, I.; Manandhar, S. SemEval-2014. Aspect Based Sentiment Analysis. Available online: <http://www.aclweb.org/anthology/S14-2004>
- [5] Al-Smadi, M.; Qawasmeh, O.; Talafha, B.; Quwaider, M. Human Annotated Arabic Dataset of Book Reviews for Aspect Based Sentiment Analysis. In Proceedings of the 2015 3rd International Conference on Future Internet of Things and Cloud (FiCloud), Rome, Italy, 24–26 August 2015.
- [6] Tamchyna, A.; Fiala, O.; Veselovská, K. Czech Aspect-Based Sentiment Analysis: A New Dataset and Preliminary Results. Available online: <https://pdfs.semanticscholar.org/cbd8/7f4201c427db33783b1890bca65f5bf99d2c>.
- [7] Apidianaki, M.; Tannier, X.; Richart, C. Datasets for Aspect-Based Sentiment Analysis in French. Available online: [http://www.lrec-conf.org/proceedings/lrec2016/pdf/61\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2016/pdf/61_Paper.pdf)
- [8] . Gayatree, G.; Elhadad, N.; Marian, A. Beyond the Stars: Improving Rating Predictions Using Review Text Content. Available online: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.150.140&rep=rep1&type=pdf>
- [9] Kiritchenko, S.; Zhu, X.; Cherry, C.; Mohammad, S. NRC-Canada-2014: Detecting Aspects and Sentiment in Customer Reviews. Available online: <http://www.aclweb.org/anthology/S14-2076>
- [10] Kiritchenko, S.; Zhu, X.; Cherry, C.; Mohammad, S. Supervised and Unsupervised Aspect Category Detection for Sentiment Analysis with Co-occurrence Data. In *IEEE Transactions on Cybernetics*; IEEE: Piscataway, NJ, USA, 2017.
- [11] Soujanya, P.; Cambria, E.; Gelbukh, A. Aspect extraction for opinion mining with a deep convolutional neural network. *Knowl.-Based Syst.* 2016, 108, 42–49.
- [12] Pengfei, L.; Joty, S.; Meng, H. Fine-Grained Opinion Mining with Recurrent Neural Networks and Word Embeddings. Available online: <http://www.aclweb.org/anthology/D15-1168>

- [13] Pontiki, M.; Galanis, D.; Papageorgiou, H.; Manandhar, S.; Androutsopoulos, I. Semeval-2015 Task 12: Aspect Based Sentiment Analysis. Available online: <http://www.aclweb.org/anthology/S15-2082>
- [14] Mohammad Samman Hossain, Israt Jahan Jui, Afia Zahin Suzana, Sentiment Analysis for Bengali Newspaper Headlines. Available at <<[http://dspace.bracu.ac.bd/xmlui/bitstream/handle/10361/8246/13301040%2c%2013301120%2c%2013101289%20\\_%20CSE.pdf?sequence=1&isAllowed=y](http://dspace.bracu.ac.bd/xmlui/bitstream/handle/10361/8246/13301040%2c%2013301120%2c%2013101289%20_%20CSE.pdf?sequence=1&isAllowed=y)>>
- [15] Md. Atikur Rahman, Emon Kumar Dey, Datasets for Aspect-Based Sentiment Analysis in Bangla and Its Baseline Evaluation.. <https://www.mdpi.com/2306-5729/3/2/15>
- [16] Support vector machine/Analytics vidhya, available at <<<https://www.analyticsvidhya.com/blog/2015/10/understaing-support-vector-machine-example-code/>>> last accessed on 12-10-2019 at 12.00 AM.
- [17] Confution Matrix/ Towards data Science, available at <<<https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62>>>. last accessed on 27-11-2019 at 1.00 AM.



## Sentiment Analysis on E-Commerce Business in Bangladesh Perspective

### ORIGINALITY REPORT

<b>23%</b> SIMILARITY INDEX	<b>19%</b> INTERNET SOURCES	<b>15%</b> PUBLICATIONS	<b>8%</b> STUDENT PAPERS
--------------------------------	--------------------------------	----------------------------	-----------------------------

### PRIMARY SOURCES

<b>1</b>	<b>www.mdpi.com</b> Internet Source	<b>9%</b>
<b>2</b>	<b>www.datacamp.com</b> Internet Source	<b>3%</b>
<b>3</b>	<b>Md. Rahman, Emon Kumar Dey. "Datasets for Aspect-Based Sentiment Analysis in Bangla and Its Baseline Evaluation", Data, 2018</b> Publication	<b>3%</b>
<b>4</b>	<b>www.analyticsvidhya.com</b> Internet Source	<b>1%</b>
<b>5</b>	<b>Submitted to National Institute of Technology, Patna</b> Student Paper	<b>1%</b>
<b>6</b>	<b>Submitted to University of Sheffield</b> Student Paper	<b>1%</b>
<b>7</b>	<b>G Krishna Mohan, N Yoshitha, M L.N.Lavanya, A Krishna Priya. "Assessment and Analysis of Software Reliability Using Machine Learning</b>	<b>1%</b>