Heaven's Light is Our Guide

DEPARTMENT OF ELECTRONICS & TELECOMMUNICATION ENGINEERING

Rajshahi University of Engineering & Technology,
Bangladesh

Sentiment Analysis and Product Review Classification in E-commerce Platform

Author

Mahmud Hasan Munna

1504031

Supervised By

A.S.M. Badrudduza

Assistant Professor

RUET

December 2020

Acknowledgment

Firstly, I would like to express my gratitude to my supervisor A.S.M. Badrudduza for his insightful discussions on my research work. He kept faith in me and allowed me to choose the field in which I interested in. I am grateful to him for his valuable guidance throughout my research with his patients. Without her knowledge, supervision, kind guideline, remarkable feedback, and encouragement, it was not possible to complete this research work. She has helped me to improve the lacking and kept me going with the finest improvement.

I would also like to express my gratitude to my parents, siblings, family members, and friends who supported me in this journey thoroughly.

Lastly, I am thankful and grateful to The Almighty Allah for giving me the ability to complete my research properly. His word from the Holly Quran kept me on the right path when I was losing motivation.

Mahmud Hasan Munna

Heaven's Light is Our Guide

DEPARTMENT OF ELECTRONICS & TELECOMMUNICATION ENGINEERING

Rajshahi University of Engineering & Technology, Bangladesh

CERTIFICATE

This is to certify that the thesis entitled "Sentiment Analysis and Product Review Classification in E-commerce Platform" by Mahmud Hasan Munna, Roll No. 1504031 has been carried out under my supervision. To the best of my knowledge, this thesis work is an original one and was not submitted anywhere for any degree or diploma.

Supervisor

....
(A. S. M. Badrudduza)
Assistant Professor
Dept. of ETE
RUET

External

....

Head

....
(Dr. Md Munjure Mowla)
Ascociate Professor
Dept. of ETE
RUET

Abstract

Online shopping is becoming one of the most demanding everyday needs, nowadays. These days people are feeling comfortable shopping online. The number of its customers is increasing day by day as well as raising some problems. About 2 billion people buy regular staffs from online shops where the customers of the online shop are about 3 million in Bangladesh. The major problem is that the customers can not choose the quality-full product by reading every review of an online product. Besides, the product reviews are helpful to improve the services of an e-commerce site but required huge manpower and time. We have focused on Bangla text and aimed to solve these problems by the application of Machine Learning(ML), Deep Neural Network (DNN) and Natural Language Processing (NLP). In this study, we have proposed two deep learning and four ML models: one DNN and two ML models are for sentiment analysis and the others are for Product Review Classification intended to improve both the quality and services.We have done The sentiment analysis targeting the benefit of the customers and The purpose of Product Review Classification is for the benefit of e-commerce service providers. Significantly, our proposed(DNN) models result in high accuracy: 0.84 and 0.69 for both Sentiment Analysis and Product Review Classification, respectively. Undoubtedly, these models can help the customers to choose the right product and the service provider to improve their services.

Key Words: Sentiment Analysis, Online Product Review Classification, E-commerce, Bangla NLP, Deep Learning, Machine Learning

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Online shopping is very popular nowadays. Around 2 billion people in the world use e-commerce sites to buy daily necessities. Almost 63 percent of shopping is done on e-commerce sites. Day by day the demand for online shopping is rapidly increasing and the market is growing by 25% to 30% every year. People feel comfortable shopping staying at home. Moreover, while it is a must to stay at home in the corona pandemic, people are fulfilling product needs using online shops. This article [6] informs that online sales growth is 76 percent in June 2020 compared with June 2019 due to the corona-virus pandemic. We have taken a Natural Language Processing approach to solve some e-commerce related issues.

1.1 Motivation

In online shopping, after getting delivery, customers are used to writing reviews or comments to express their sentiment regarding the quality of the products as well as services. However, in online shopping, e-commerce service providers suffer a lot to help customers to select reliable products. Usually, customers look at product reviews to select good products. But sometimes it is very arduous to select a product by looking at a huge amount of reviews manually. Again, online shops are usually third party organizations, who sell merchants' products to customers, composed of different teams to take care of various issues. For example, one team collects complaints from customers to improve their services, in contrast, other team collects customers' recommendations to promote their products. But, for a particular team it is very much time-consuming to read all kinds of reviews and manually separate by category and also needs lots of manpower. That is why they can not provide a fully efficient service to the customers.

If these problems can be solved by an autonomous system many customers, businessmen, and e-commerce service providers will be able to save a huge amount of time. They won't have to read a thousand reviews anymore. It will be beneficial for both consumers, merchants, and service providers. Without reading any review, customers will be able to know if the quality of a product is good or bad and order products with confidence. On the other hand, e-commerce service providers can easily analyze problems with their products as well as customers' needs. Moreover, it will also allow e-commerce sites to promote themselves and solve delivery errors. The improvement team of e-commerce sites will be able to work relatively more efficiently.

In this study, we have intended to work with Bangla text which is spoken by 228 million people all around the world. But in the online platform, Bengali people seem to be more comfortable with Phonetic Bangla and sometimes with English. That is why, we have created our dataset by incorporating the Bangla, Phonetic Bangla, and English texts which makes it a unique dataset as well as more complex.

1.2 Literature Review

There has been an increasing amount of literature on NLP and some Bangladeshi researchers have been emphasized on solving the problems related to Bangla text. But a few work has been conducted those are concerned with the e-commerce business. Some of the state-of-the-art works related to Bangla NLP have been described below.

In [7], the prominent authors classified two types of Bangla language: 'shadhu' and 'cholito'. They applied Multinomial Naive Bayes classification and obtained very promis-Predicteding results with 99% accuracy and 97% f1-score. Contribution in [8], authors took several approaches to classify fake and real news. They applied three machine learning algorithms and built a DNN model which provided the best accuracy score. The contribution on [9] highlighted an approach to classify Bangla and phonetic Bangla reviews using vectorizers and traditional machine learning algorithms. This proposed methodology achieved the best performance on SVM with a 75.5% accuracy score. In another research [10], author highlighted a methodology to establish an autonomous system. They performed their model on restaurant reviews. The suggested model is the Multinomial Naive Bays algorithm that provided a much more improved accuracy score. The author in [11] represents a Deep Learning approach to classify sentiment for Bangla sentences. The author performed a Binary Classification and Multiclass Classification. They implemented Convolution Neural Network(CNN) and Long-Short Term Memory(LSTM) on their dataset which is collected from online news portal articles. The result showed 49% and 75% accuracy respectively on multiclass and binary classification. The paper [12] encompasses some of the contributions in this field. This work proposed a CNN model that is applied to Bangla comments collected from different sources and showed tremendous improvement as a result.

There are relatively few historical studies in the area of detecting abusive comments in social media. One of these works [13] experimented with a model on the toxic comment dataset from Kaggle. Their LSTM and CNN model performed in a way that ensures a 97 percent accuracy score. There is another work in this field[14]. This work highlighted some approaches to classify multiple emotions. They applied CNN and LSTM to data that are collected from youtube comments. 65% accuracy has been found out from this method. In an analysis of [15], Another sentiment analysis approach was performed in this work. The author proposed a method that classifies sentiments from Google Play Store Bangla Comments. The proposed Method highlights the Support Vector Machine(SVM) algorithm to classify sentiments. This experiment accomplished 76.48% accuracy. In [16] , the author came up with an ideal approach that achieves a 0.865 F1 score.

There are a few numbers of works on Sentiment analysis in the E-commerce site. In [17], the author proposed the SVM algorithm to classify comments from various e-commerce sites. Using this approach, researchers were able to achieve an 82.92% accuracy score. We know about [18], established their model to categorize online marketing reviews from amazon. They categorized their data into five sentiments. In a follow-up study, another research has been done on customer perception. The author proposed a Deep Learning method that shows improvement as a result. In an investigation into [19], the Author suggested a model that classified product reviews. SVM and Naive Bayes algorithms are adopted and SVM shows better performance.

1.3 Objectives

In this study, Our objective to solve the aforementioned issues by using Natural Language Processing (NLP) and developed two DNN and four ML based models: one DNN and ML models for for sentiment analysis and the others are for product review classification. The sentiment analysis model will help the service providers to show product quality to the customers on customers' point of view. On the other hand, the product review classification model is not only able to help the service providers but also can help customers. Targeting the merchants and e-commerce sites, we have classified the reviews into four categories namely *'Complain'*, *'Recommended'*, *'Wrong delivery'*, and *'Appreciation'*. The *'Recommended'* category can help a customer to buy a product without any hesitation. *'Wrong delivery'* category can help the merchants to improve their service. *'Complain'* category may help the service providers to solve errors and the *'Appreciation'* category may ensure the merchants that they are providing a good service.

1.4 Thesis Outline

In chapter 1, The motivation behind the research is discussed. Some related works of Natural Language Processing have been described as well as the objectives of the work are also discussed.

Chapter 2 Shows how ML and DNN works and the theoretical explanation of Natural Language Processing(NLP).

In Chapter 3, the Sentiment Analysis and Product Review Classification model are explained and the performance of the model on a real-word dataset is shown.

Finally, Chapter 4 concludes the study and indicates some future work through this study.

Chapter 2

Background and Preliminaries

To increase the data rate in wireless communication with limited transmit power and bandwidth, we have to use multiple antennas at either the transmitter or the receiver. In practical wireless communication systems, there usually exists fading correlation among antenna elements due to insufficient spacing between these elements. This spatial correlation reduces the diversity gain offered by multiple antennas.

2.1 Supervised Learning

Supervised Learning is a Machine Learning technique where the features are labeled. In another word, the answers to the questions in given and the computer learns from the answers and it can predict the answer of new questions. That means outputs are given of corresponding inputs and the machine programs itself when it gets new input and gives the output.

2.2 Classification

In Supervised Learning, If the labeled data is discrete or categorical then it is a classification problem. It can be any binary numbers, name of fruits, colors, discrete numbers. In a word, where the label is not a continuous number then it's a classification problem. There are two types of classification:

2.2.1 Binary Classification

In this classification, the data are classified into two classes. It can be a binary number, yes/no, on/off, etc. Table 2.1 shows a binary classification problem where the "survived" column is categorized into two classes.

Table 2.1: Dataset Sample for Binary Classification

PassengerId	Pclass	Sex	Age	SibSp	Parch	Fare	Embarked	Survived
1	3	male	22	1	0	7.25	S	0
2	1	female	38	1	0	71.2833	C	1
3	3	female	26	0	0	7.925	S	1
4	1	female	35	1	0	53.1	S	1
5	3	male	35	0	0	8.05	S	0
6	3	male		0	0	8.4583	Q	0
7	1	male	54	0	0	51.8625	S	0

2.2.2 Multiclass Classification

In this classification, data are categorized into more than two categories. This type of classification is called a multiclass classification. In the dataset shown in Table 2.2, the data are categorized into three categories which makes this a multiclass classification problem.

Table 2.2: Dataset Sample for Multiclass Classification

Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
1	5.1	3.5	1.4	0.2	Iris-setosa
2	4.9	3	1.4	0.2	Iris-setosa
3	4.7	3.2	1.3	0.2	Iris-setosa
70	5.6	2.5	3.9	1.1	Iris-virginica
71	5.9	3.2	4.8	1.8	Iris-virginica
72	6.1	2.8	4	1.3	Iris-versicolor
73	6.3	2.5	4.9	1.5	Iris-versicolor

2.3 Sentiment Analysis

Sentiment Analysis is a Natural Language Processing technique where the text data are categorized into certain sentiments or groups. It's a supervised learning technique where binary or multiclass classification is done. It's a way to find the sentiment behind the text using Natural Language Processing which done using machine learning or deep learning models.

2.4 Machine Learning Algorithms for Classification

For classification problem there are many effective ML algorithms. Logistic Regression, Naive Bayes, Support Vector Machine(SVM) are popular algorithms for classification. Some theoretical explanation of these algorithms has been discussed in the following sections.

2.4.1 Logistic Regression

Logistic regression models the probabilities for classification problems with two possible outcomes. It's an extension of the linear regression model for classification problems. Instead of fitting a straight line , the logistic regression model uses the logistic function to define the output of a linear equation between 0 and 1 [20]. The logistic function is defined as:

$$f(\sigma) = \frac{1}{1 + e^{-\sigma}} \tag{2.1}$$

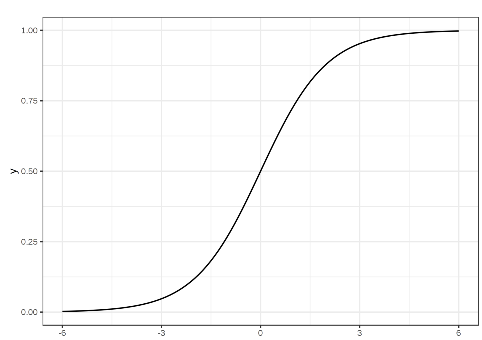

Figure 2.1: Sigmoid Curve [1]

The Logistic Regression curve has been shown in figure 2.1 which is a sigmoid shaped curve.

2.4.2 Support Vector Machine(SVM)

(SVM) is a supervised ML algorithm that can be used for both classification or regression challenges. However, it is mostly used in classification problems. In the SVM algorithm, each data are plotted item as a point in n-dimensional space with the value of each feature being the value of a particular coordinate. where n is the number of features Then, classification is performed by finding the hyper-plane that differentiates the two or more classes[21].

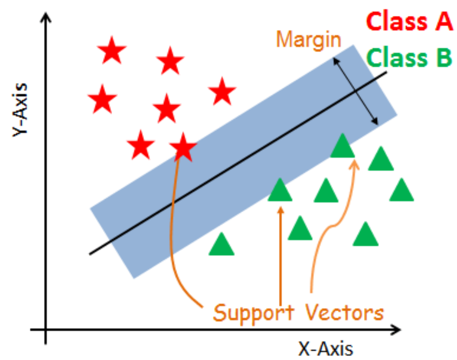

Figure 2.2: Support Vector Machine [2]

Figure 2.2 shows that, *Support vectors* are the data points, which are closest to the hyperplane. These points will define the separating line better by calculating margins. These points are more relevant to the construction of the classifier.

A *hyperplane* is a decision plane that separates between a set of objects having different class memberships.

A *margin* is a gap between the two lines on the closest class points. This is calculated as the perpendicular distance from the line to support vectors or closest points. If the margin is larger in between the classes, then it is considered a good margin, a smaller margin is a bad margin.

2.5 Deep Neural Network

Deep-learning is a machine learning procedure. It instructs a computer to channel inputs through layers to memorize how to predict and classify data. Perceptions can be within the shape of pictures, content, or sound.The larger part of cutting edge deep learning architectures are based on Deep Neural Network (DNN). They utilize numerous layers of nonlinear processing units for include extraction and transformation. Each progressive layer employments the yield of the past layer for its input. What they learn shapes a pecking order of concepts. In this pecking order, each level learns to convert its input data into an increasingly unique and composite representation.

2.5.1 Architecture of DNN

The architecture of DNN: Deep Neral Network contains multiple layers. The first layer is called the input layer and the last layer is called the output layer. The middle layers can be one or more which is called hidden layers. The architecture of DNN is shown in figure 2.3.

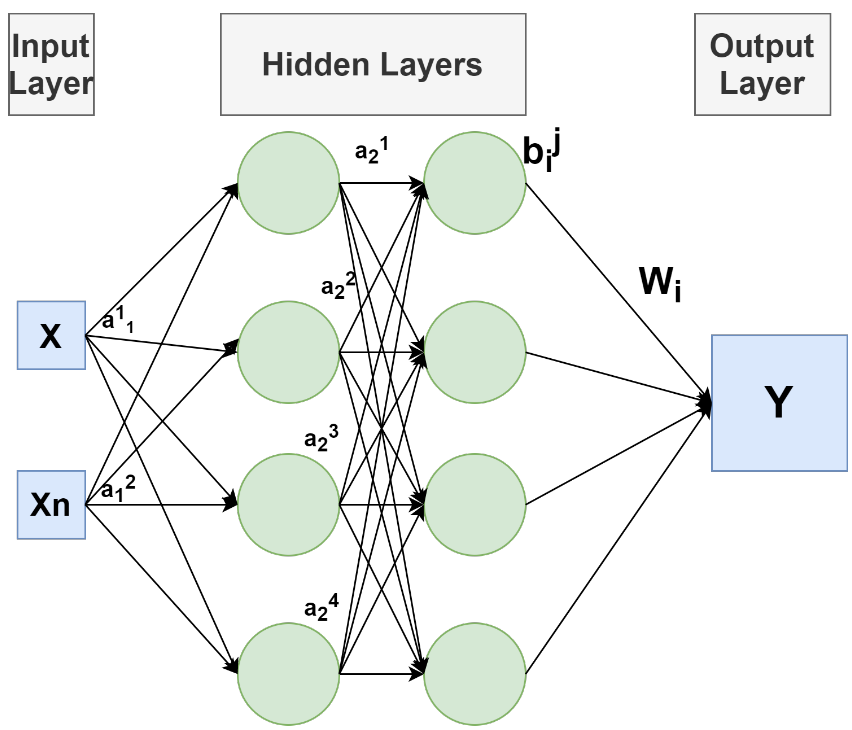

Figure 2.3: Architecture of DNN

2.5.2 Neurons in DNN

The neurons are normally organized in multiple layers. The number of neurons in input layers depends on the input dimension of the data. The number of neurons in output layers depends on the number of the desired output. Neurons in the hidden layers depend on the input layers. There is no hard and fast rule for selecting the number of neurons in the hidden layer. But there are some thumb rules are mentioned in [22]. More the neurons, the calculation, and the model are more complex.

2.5.3 Function Used in DNN

There is a mathematical function that works behind a deep neural network which is given below:

Activation Calculating Function

$$a_j^i = \sigma\left(\sum\left(w_{jk}^i . a_k^{i-1}\right) + b_j^i\right). \tag{2.2}$$

where

a = activation value of the j^{th} neuron in the i^{th} layer

w = weight from the k^{th} neuron in the $i\text{-}1^{th}$ layer to the j^{th} neuron in the i^{th} layer

b = bias of the j^{th} neuron in the i^{th} layer

σ = Activation Function

Cost Function

A cost function is a degree of how great a neural network did with regard to it's given training sample and the anticipated output. It too may depend on factors such as weights and biases. A cost function is a single value, not a vector since it rates how great the neural network did as an entire.

the cost function:

$$C(W, B, Sr, Er) \tag{2.3}$$

where

S = Input of a single training sample

E = Desired output of that training sample

B = B is our neural network's biases

W = Neural network's weights

Activation Function

Activation functions play a very important role in any neural network .The activation function triggers a neuron or node and determines whether the data from this node will be sent to the next node. It normally transfers input signal to output signal. There are some activation functions used for different scenarios. Rectified Linear Unit (RelU) is used when the range of input is between 0 to infinite. When the range of input is between -1 to +1, the tanh function is used. The range of a sigmoid function is between 0 to 1. softmax is another activation function for multiclass classification.

$$g(z) = max\{0, z\} \tag{2.4}$$

$$f(x) = tanh(x) = \frac{2}{1+e^{-2x}} - 1 \tag{2.5}$$

$$f(x) = \frac{1}{1+e^{-x}} \tag{2.6}$$

$$S(y_i) = \frac{e^{y_i}}{\sum_{j=1}^{i} e^{y_j}}. \tag{2.7}$$

The equation no. 2.3, 2.4, 2.5, 2.6 respectively is the equation of ReLu, Tanh, sigmoid and softmax.

Figure 2.4, 2.5 and, 2.6 is the visual of Relu, Tanh and Sigmoid function respectively.

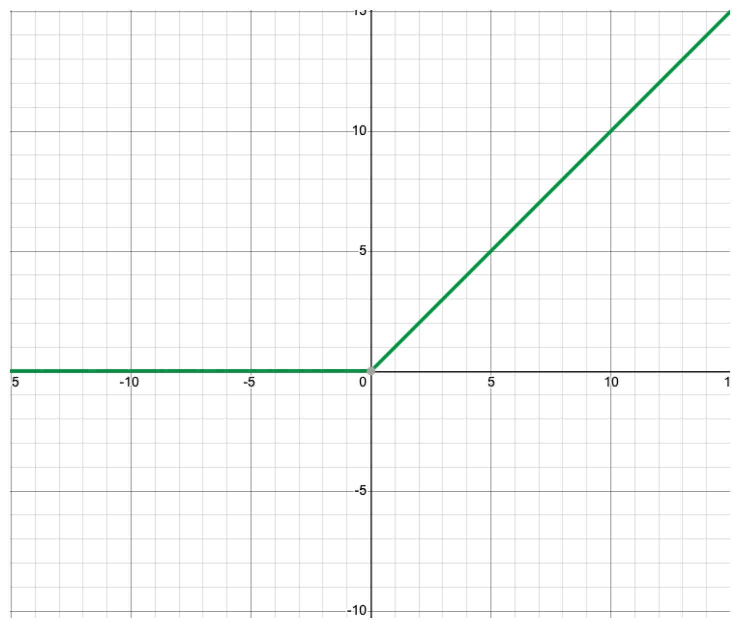

Figure 2.4: Activation Fuction : Relu [3]

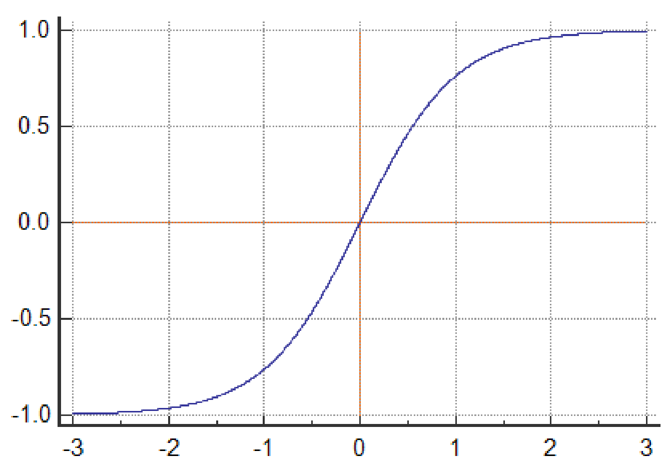

Figure 2.5: Activation Fuction : Tanh [4]

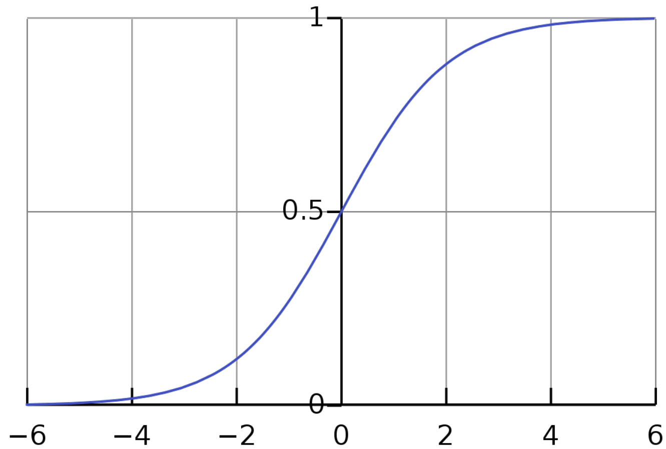

Figure 2.6: Activation Fuction : sigmoid [5]

2.5.4 Optimizer of DNN

The optimization algorithm has a vital role in neural networks that converges the model into the most suitable form by adjusting weights. There are several optimization algorithms such as Adam, RMSprop, SGD and so on. The optimization algorithm basically iterates until the best accuracy of the model is achieved.

2.5.5 Walk-through of Training a Deep Neural Network

The algorithm or workflow of training a dataset into deep neural network is given below step by step:

1. Initiating the weight(w) to small number close to 0.
2. Insterint first observation of dataset into input layer.
3. Forward Propagation
4. Comparing the predicted result to actual result.
5. Back Propagation
6. Updating the weights after a batch of observation
7. Repeat 1-6 untill it reaches the desired result

Chapter 3

Sentiment Analysis and Product Review Classification

In this chapter, The methodology of designing our proposed models are discussed. We started our work by collecting the dataset. Then we have preprocessed the data and extracted the feature. Then we have fed the extracted feature to our DNN and ML models. Finally, We have analyzed the result.

3.1 Data Collection

In this research, we have worked with a real world dataset. We have collected product reviews from various e-commerce sites namely Daraz, BDshop, and Evally which are the topmost online shops in Bangladesh. The uniqueness of this dataset is, it contains Bangla, Phonetic Bangla, and English text. We have put Phonetic Bangla text in the dataset because 80 percent of internet users express their opinion in Phonetic Bangla on the online platform. The purpose of including three types of Bangla text in the dataset is to make the model more generalized.

For two different applications, we have manually annotated each review in two-stages. For our first experiment, we have labeled the reviews *'good'* and *'bad'* based on the quality of products on customers' points of view. In contrast, for the second experiment, the reviews have been categorized based on the expression of customers' comments. In this case, the reviews have been labeled with *'Complain'*, *'Recommended'*, *'Wrong delivery'*, and *'Appreciation'*. The distributions of the different target classes in our prepared dataset have been shown in Table 3.1 and 3.2.

Table 3.1: Class distribution for Sentiment Analysis

Class Labels	**Counts**
good	2830
bad	2279

Table 3.2: Class distribution for Product Review Classification

Class Labels	**Counts**
Complain	1908
Recommended	1562
Appreciation	1185
Wrong delivery	454

From Table 3.1, we can observed that our dataset is properly balanced for the Sentiment Analysis. But Table 3.2 shows, for Product Review Classification, the dataset is slightly imbalanced that has made the model more complex to gain the best performance.

3.2 Data Prepossessing

Usually, raw data contains a big number of unwanted elements that directly affect the performance of the machine learning or deep learning models. That is why, we have removed those elements such as stop-words, punctuations, and unwanted characters from our dataset.

3.2.1 Removing Stop-words

Stop-words are the words that do not contains any significant information, even sometimes decrease the performance of the model. So, we have suppressed all of the stop-words from our dataset.

3.2.2 Removing Punctuation

Punctuations do not carry any informative meaning for the classification tasks. That is why, to reduce the complexity, we have removed the puntuations from the data as well.

3.2.3 Removing Unnecessary Characters

As we have scraped the product reviews from online, the dataset contains many unnecessary characters like '@', '=', '&', and so on. These characters have zero impact and likely to mislead the learning of our models. To improve the model performance, we have normalized our dataset by removing these characters.

3.2.4 Data Partitioning

Data partitioning is an important task since the ratio of the partitions has an impact on the evaluation of the performance of a deep learning model. We have split our dataset into train, validation, and test dataset. In the case of sentiment analysis for DNN, our training, validation, and test dataset contain 4085 (=80%), 715(=14%), and 307 (=6%) instances, respectively. On the other hand, the training, validation, and test dataset for Product Review Classification contain 3574 (=70%), 1073 (=21%), and 460 (=9%) instances, respectively. We have taken 20% of the main data in our test dataset for the ML models.

3.2.5 Feature Extraction

Using FastText

Feature extraction is an indispensable task to train up any DNN model. In order to extract the feature vectors, in this study, we have applied the fastText pre-trained model for Bangla language. FastText, developed by Facebook's AI Research (FAIR) lab, provides word embeddings of shape (1, 300) against each of the word. Thus, for each of the sample text, we have obtained a feature vector of shape (1, 300) by aggregating the individual vectors of the words contained by the text.

3.2.6 Using TF-IDF

TF-IDF is one of the best method for extracting feature for ML model. We have extracted numeric feature from our text data using TF-IDF method. TF-IDF stands for Term Frequency — Inverse Document Frequency and is a statistic that aims to better define how important a word is for a document, while also taking into account the relation to other documents from the same corpus. The multiplication of TF and IDF is the final extracted feature. The mathematical expression of TF-IDF is as follows:

$$tf(w,d) = log(1 + f(w,d)), \tag{3.1}$$

$$tf(w,D) = log\frac{N}{f(w,D)}. \tag{3.2}$$

final step to calculate TF-IDF:

$$tfidf(w,d,D) = tf(w,d) * idf(w,d). \tag{3.3}$$

where

N = number of documents in dataset,

d = given document from dataset,

D = the collection of all documents,

w = a given word in a document.

3.3 Machine Learning Models

3.3.1 Logistic Regression Model

We have two classification problems to perform. One is binary classification and another one is multiclass classification. Logistic Regression is one of the efficient classifiers, especially for binary classification. We have chosen this algorithm because we have a binary classification problem. the logistic regression model uses the logistic function to define the output of a linear equation between 0 and 1. The logistic function is as follows:

$$f(z) = \frac{1}{1+e^{-z}} \tag{3.4}$$

3.3.2 SVM Model

SVM is a kernel-based classifier. Support vector machine classifier can perform classifying and regression studies. In a supervised machine, learning input is fed and the desired output is labeled in the training set. SVM classifier can classify the train data efficiently and predict the valid data into its belonging class. We have used SVM for our two classification problems. It can be more effective on our multiclass classifier.

3.4 DNN Model Architecture

Each of the DNN model contains some common artifacts such as the input and output layers, hiddern layers, activation functions, and optimizers. Proper selection of the number of hidden layers and the number of neurons in the hidden layers have the utmost impact on the performance of the desinged DNN model. In [23], the author extensively describes the way to select parameters for an efficient DNN model and its significance. However, the architectures of our proposed Sentiment Analysis and Product Review Classification models have been shown in Figure 3.1 and 3.2, respectively.

3.4.1 Neurons in the Input and Output Layers

The input layers of a neural network depends on the shape of the training data. As the number of the input features for both of our models is identical, which is 300, we have specified 300 neurons in the input layer. On the other hand, for classification, the number of the neurons in the output layer depends on the number of unique classes. For our Sentiment analysis model, basically, we have performed a binary classification and thus we have specified 1 neuron in the output layer. While we have specified 4 neurons in the output layer of

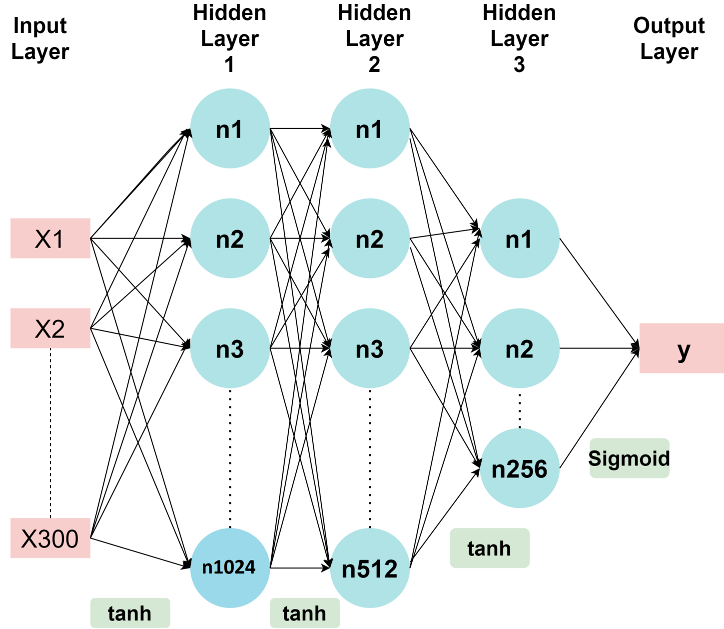

Figure 3.1: DNN Model for Sentiment Analysis

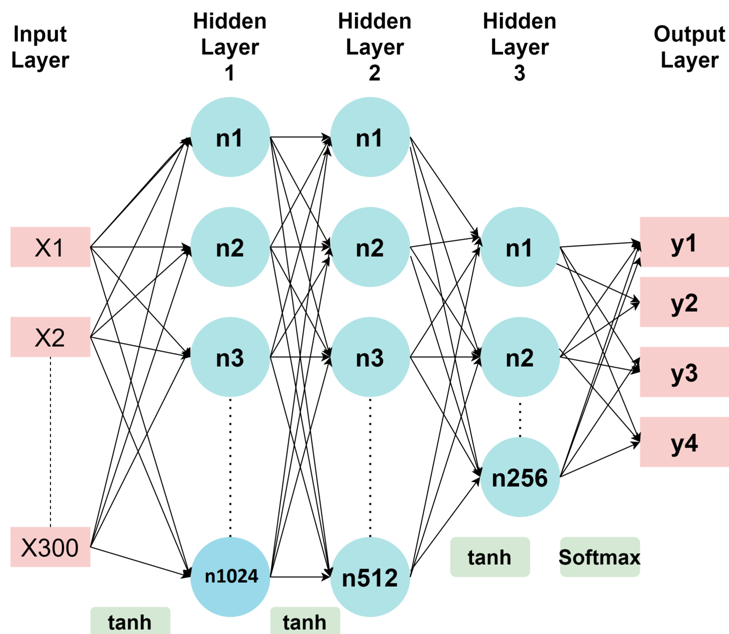

Figure 3.2: DNN Model for Product Review Classification

our Product Review Classification model architecture based on the unique classes.

3.4.2 Hidden Layers and Neurons

It's an important factor determining the number of hidden layers to design a neural network structure. It defines the complexity and efficiency of DNN architecture. A study in [22], Jeff Heaton explained, what should be the number of hidden layers for a DNN architecture. We set three hidden layers for our model according to Jeff's idea. As we don't have huge data we don't need more than three hidden layers. More than 3 hidden layers will make the model more complex, inefficient and it also may cause overfitting. In [24], Moolayi explained a thumb rule for determining the number of neurons in hidden layers. According to Moolayi, if x is the number of input dimensions, the number of neurons in the first hidden layers should be the closest number of 2x in the power of 2. However, in both of the sentiment analysis and product review classification models, we have specified the number of neurons 1024, 512, and 256 for the first, second, and third hidden layer, respectively.

3.4.3 Activation Functions

Activation functions play a very important role in any neural network [25]. The activation function triggers a neuron or node and determines whether the data from this node will be sent to the next node. We have used these three activation functions briefly explained as follow:

tanh

Tanh, stands for tangent hytperbolic function, is a sigmoid shaped function ranges from -1 to 1. As we have negative values and positive values in the vectorized form of our training dataset in the range of -1 to 1, that is why, we have used tanh activation function in the input and hidden layers. The mathematical form of this function can be written as follows:

$$f(x) = tanh(x) = \frac{2}{1+e^{-2x}} - 1 \tag{3.5}$$

Sigmoid

Sigmoid is a popular non-linear activation function and the output range is [0,1]. Since the target columns of out preprocessed sentiment analysis dataset contains either 0 or 1, we have applied the sigmoid activation function in the output layer of the model. The sigmoid function can be expressed mathematically by:

$$f(x) = \frac{1}{1+e^{-x}} \tag{3.6}$$

Softmax

The Softmax is a generalization of the sigmoid function that is used to manage multiple classes (multi-class taxonomy). We have applied this function at the output layer of the Product Review Classification model to obtain the probability of occurrence for each of the four classes. The Softmax function can take any type of score (an array or vector) and return the appropriate probabilities, the larger the scores the larger the return value. The mathematical formula for the Softmax function is:

$$S(y_i) = \frac{e^{y_i}}{\sum_{j=1}^{i} e^{y_j}}. \tag{3.7}$$

3.4.4 Optimization Algorithm

The optimization algorithm has a vital role in neural networks that converges the model into the most suitable form by adjusting weights. There are several optimization algorithms such as Adam, RMSprop, SGD and so on. In our experiment, we have applied the Adam optimization algorithm for both of the models with a learning rate 0.001. The significance of this function is that it has relatively low memory requirements and performs well with a minimal hyper-parameter tuning. The mathematical expressions for Adam are as follows:

$$v_t = \beta_1 \text{Ö} v_{t-1} - (1-\beta_1)\text{Ö} g_t, \tag{3.8}$$

$$s_t = \beta_2 \text{Ö} s_{t-1} - (1-\beta_2)\text{Ö} g_t, \tag{3.9}$$

$$\Delta w_t = -\eta \frac{v_t}{\sqrt{s_t + \epsilon}} \ddot{\text{O}} g_t, \tag{3.10}$$

$$w_{t+1} = w_t + \Delta w_t. \tag{3.11}$$

where

η = Initial learning rate

g_t = Gradient at time t along w_j

v_t = Exponential average of gradients along w_j

s_t = Exponential average of squares of gradients along w_j

$\beta_1; \beta_2$= Hyperparameters

3.5 Evaluation Metrics

In order to validate the classification performance of our models, we have applied several evaluation matrices namely Accuracy, Precision, Recall, and F1 Score. We have generated the confusion matrix for both of the models and calculates the value of the matrices by using the following equations.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{3.12}$$

$$\text{Precision} = \frac{TP}{TP + FP}, \tag{3.13}$$

$$\text{Recall} = \frac{TP}{TP + FN}, \tag{3.14}$$

$$\text{F1-Score} = 2 * \frac{Recall * Precision}{Recall + Precision}. \tag{3.15}$$

Where

TP= True Positive

TN= True Negative

FP= False Positive

FN= False Negative

3.6 Result and Analysis

In this phase, we have analyzed the results to figure out the performance of our proposed models. We have recorded the numeric value of our result in 2 decimal places. Analyzation has been performed in three steps. We have generated the learning curves to analyze training and validation performance which has helped us to determine if there is any underfitting or overfitting issue. And then, We have performed a prediction on test dataset to analyze the performance of the model in reality. And finally, to understand the individual class performance, we have generated a classification report and confusion matrix.

3.6.1 Sentiment Analysis

The performance analysis of logistic regression, SVM and DNN model for Sentiment Analysis is in following sections.

Performance of Logistic Regression Model

We have obtained accuracy scores for both the training and test data set. From table 3.3, we can observe that the training accuracy is much more than the test accuracy. As a result, this model can be called overfitted.

Table 3.3: Accuracy Scores of Logistic Regression for Sentiment Analysis

Dataset	**Accuracy**
Train	0.89
Test	0.83

In table 3.4, we can observe that the model can predict 850 sentiments correctly out of 1022 predictions which can be considered a promising result.

We have obtained a classification report for our logistic regression model which is shown in Table 3.5.

F1-score for the 'good' category is 0.85 which is the maximum f1 score for this model.

Table 3.4: Confusion Matrix of Logistic Regression for Sentiment Analysis

Actual	**Predicted**		
		Bad	***Good***
	Bad	361	87
	Good	85	489

Table 3.5: Classification Report of Logistic Regression for Sentiment Analysis

	precision	**recall**	**f1-score**
bad	0.81	0.81	0.81
good	0.85	0.85	0.85

Performance of SVM Model

We have obtained accuracy scores for both the training and test data set. From table 3.6, we can observe that the training accuracy is much more than the test accuracy as like as Logistic regression model. As a result, this model also can be called overfitted.

Table 3.6: Accuracy Scores of SVM for Sentiment Analysis

Dataset	**Accuracy**
Train	0.98
Test	0.88

In table 3.7, we can observe that the model can predict 903 sentiments correctly out of 1022 predictions which is better than previous model.

We have obtained a classification report for our SVM model which is shown in Table 3.8.

F1-score for the 'good' category is 0.89 which is the maximum f1 score for this model.

Table 3.7: Confusion Matrix of SVM for Sentiment Analysis

Actual		**Predicted**	
		Bad	***Good***
	Bad	398	71
	Good	48	505

Table 3.8: Classification Report of Logistic Regression for Sentiment Analysis

	precision	**recall**	**f1-score**
bad	0.85	0.89	0.87
good	0.91	0.88	0.89

Performance of DNN Model

To obtain the learning graph, We have plotted the loss function with respect to the epochs. Figure 3.3 represents the loss vs epochs for the sentiment analysis model. From Figure 3.3, we can observe that the training and validation loss have been converged consistently and has been intercepted at 13^{th} iteration. It is a obvious sign that our model has been well-learned.

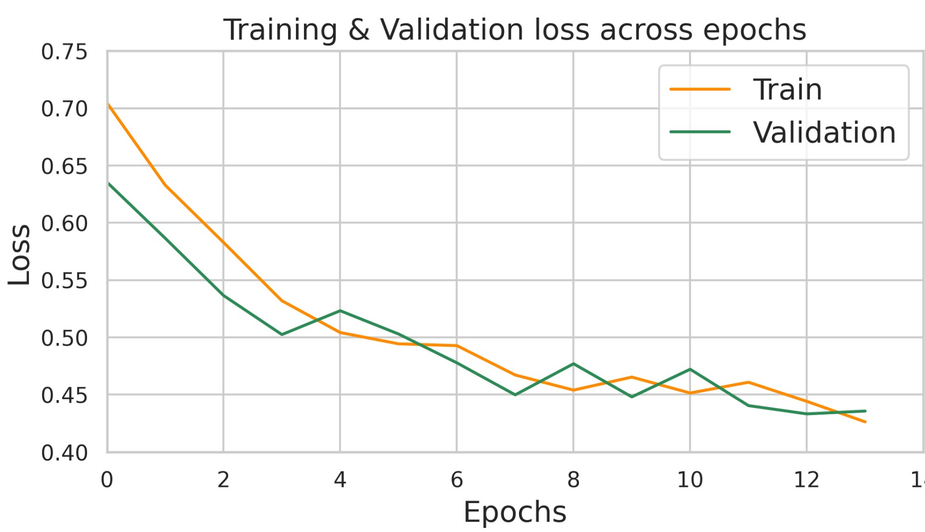

Figure 3.3: Loss vs Epoch Curve for Sentiment Analysis

The training, validation, and testing accuracy for the sentiment analysis model have been shown in Table 3.9. From this table, it is appear that the validation accuracy (=0.83) is slightly lower than the training accuracy that indicates the our sentiment analysis model has not been over-fitted or under-fitted during the training phase. Again, the precision, recall, and f1-score have also been calculated and shown in Table 3.10. From this we can see that out model have resulted in a good f1-score for both of the classes: good (=0.85) and bad (=0.83).

Table 3.9: Performance of DNN for the Sentiment Analysis

Dataset	**Accuracy**
Training	0.84
Validation	0.83
Test	0.84

Table 3.10: Classification Report of DNN for the Sentiment Analysis

	Precision	**Recall**	**F1-Score**
Bad	0.81	0.85	0.83
Good	0.87	0.84	0.85

For understanding the true and false classification behavior of our sentiment analysis model, we have generated the confusion matrix as well which is presented in Table 3.11. From the confusion matrix we can observe that, out of 307 test samples, the number of misclassification for the 'good' and 'bad classes are 28 and 21, respectively.

Table 3.11: Confusion Matrix of DNN for Sentiment Analysis

		Predicted	
		Bad	*Good*
Actual	*Bad*	116	21
	Good	28	142

Based on the above analysis, we can say that the performance of our sentiment analysis model is really good.

3.6.2 Product Review Classification

The performances of Logistic Regression, SVM, and DNN model for product review classification are shown in the following sections.

Performance of Logistic Regression Model

Firstly we have obtained the accuracy score which is shown in table 3.12. The train and test accuracy are respectively are 0.91 and 0.78 which indicates that the model overfitted.

Table 3.12: Accuracy Scores of Logistic Regression for Product Review classification

Dataset	Accuracy
Train	0.91
Test	0.78

We have generated a confusion matrix for our logistic regression model which is shown in Table 3.13. We can observe that it predicted the 'complain' category 275 times correctly. The 'recommended' category predicted incorrectly most of the time. This model has predicted 692 times correctly out of 1022 predictions.

Table 3.13: Confusion Matrix of Logistic Regression for Product Review classification

		predicted			
		appreciation	***complain***	***recommended***	***wrong delivery***
actual	***appreciation***	150	1	12	0
	complain	35	351	55	37
	recommended	44	25	261	5
	wrong deleviry	2	4	4	36

We have finally generated a classification report of or Logistic Regression model. From Table 3.14, we can observe that the f1-score of the 'complain' category is 0.74 which is best. But the worst f1-score 0.60 has been achieved by the 'appreciation' category.

We can conclude that this model performed well except for the overfitting issue.

Table 3.14: Classification Report of Logistic Regression for Product Review classification

	precision	recall	f1-score
appreciation	0.63	0.57	0.6
complain	0.77	0.72	0.74
recommended	0.62	0.72	0.67
wrong delivery	0.65	0.59	0.68

Performance of SVM Model

At the beginning, we have calculated the accuracy score which is shown in table 3.15. The train and test accuracy are respectively are 0.93 and 0.65. The very high accuracy for the training dataset while it falling on test accuracy. It's a clear sign of overfitting.

Table 3.15: Accuracy Scores of SVM for Product Review classification

Dataset	Accuracy
Train	0.93
Test	0.65

We have generated a confusion matrix for our SVM model which is shown in Table 3.16. We can observe that it predicted the 'complain' category 351 times correctly. Also, The 'complain' category predicted incorrectly most of the time. This model has predicted 798 times correctly out of 1022 predictions.

Table 3.16: Confusion Matrix of SVM for Product Review Classification

		predicted			
		appreciation	***complain***	***recommended***	***wrong delivery***
actual	***appreciation***	150	1	12	0
	complain	35	351	55	37
	recommended	44	25	261	5
	wrong delevery	2	4	4	36

We have finally generated a classification report of or SVM model. From Table 3.17, we can observe that the f1-score of the 'complain' category is 0.82 which is best. F1-score for ‘wrong delivery’ category 0.58 which is very poor. It indicates that the model performed well for the ‘complain’ category but fails it predicts incorrectly for the ‘wrong delivery’ category.

Table 3.17: Classification Report of SVM for Product Review classification

	precision	**recall**	**f1-score**
appreciation	0.92	0.65	0.76
complain	0.73	0.92	0.82
recommended	0.78	0.79	0.78
wrong delevery	0.78	0.46	0.58

Performance of DNN Model

To validate the learning behavior in the training phase, like sentiment analysis, we have plotted the loss vs epochs that has been presented in Figure 3.4. Figure 3.4 appears that the losses for both the training and validation has been remaining quite closer to each other with respect to epochs. It is the sign of a well-fitted model.

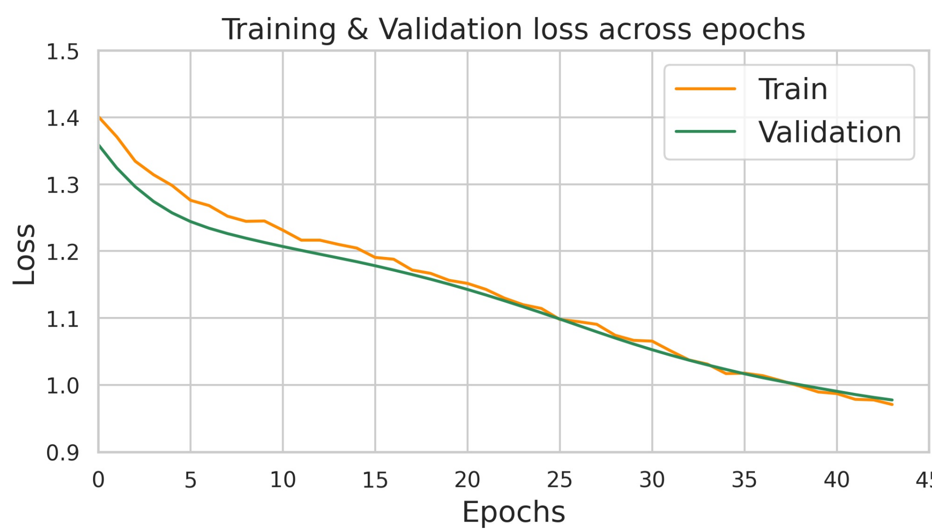

Figure 3.4: Loss vs Epoch Curve for Product Review Classification

The training, validation, and testing accuracy of our Product review Classification model have been shown in Table 3.18. Like sentiment analysis, here the validation accuracy (=0.69) is also quite closer and slightly lower than the training accuracy (=0.70) that indicates this model has been learned from the training data without over-fitting or under-fitting. For imbalanced classes, sometimes the accuracy is not a good measure and thus we have also calculated the precision, recall, and f1-score with the accuracy present in Table 3.19. From Table 3.19, we can see that our model have performed well for each of the 4 different classes with a high f1-score.

Table 3.18: Performance of the Product Review Analysis

Dataset	**Accuracy**
Training	0.70
Validation	0.69
Test	0.69

Table 3.19: Classification Report for Multiclass Classification

	Precision	**Recall**	**F1-Score**
Appreciation	0.72	0.55	0.62
Complain	0.7	0.86	0.77
Recommended	0.66	0.63	0.64
Wrong delivery	0.67	0.54	0.59

During the testing phase, we have also generated a confusion matrix to analyze the class-wise performance represents in Table 3.20. Where 'A', 'C', 'R' and, 'W' is stands for 'Appreciation', 'Complain', 'Recommended' and, 'Wrong Delivery'.

Table 3.20: Confusion Matrix for Multiclass Classification

		Predicted			
		A	*C*	*R*	*W*
Actual	*A*	59	17	30	1
	C	2	148	14	8
	R	21	29	88	2
	W	0	17	2	22

From Table 3.20, it can be observed that, despite of being a multi-classification problem, our model has promisingly performed very well for each of the 4 classes except 'Wrong

Delivery'.

3.6.3 Comparison of the Results of Logistic Regression, SVM and DNN

To compare the model accuracy we generated a bar chart shown in figure 3.5 and 3.6. For the sentiment analysis, we can observe from figure 3.6 that the accuracy score is high SVM classifier. But if we look at the training accuracy it is pretty much overfitted. For the DNN model, both the training and test dataset got an accuracy of 0.84 so it is clearly not an overfitted model. So, we are proposed the DNN model is the best model in the case of Sentiment Analysis. On the other hand, the DNN model for product review classification shows the best performance for the overfitting issue despite the logistic regression is showing the high accuracy.

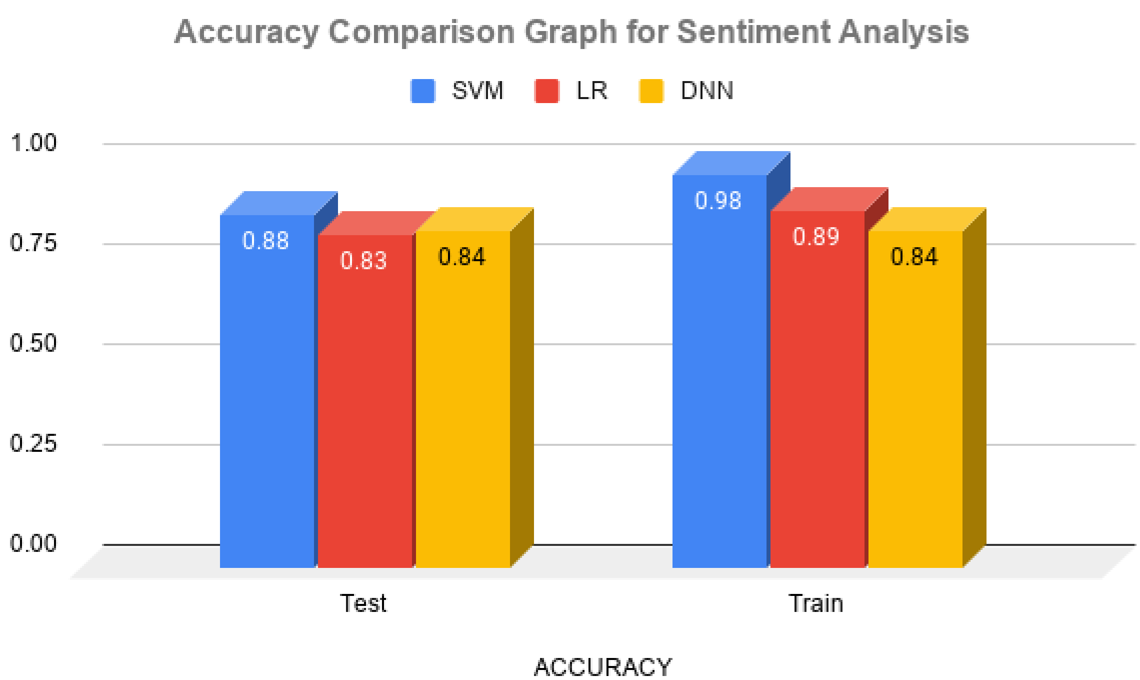

Figure 3.5: Accuracy Comparison for Sentiment Analysis

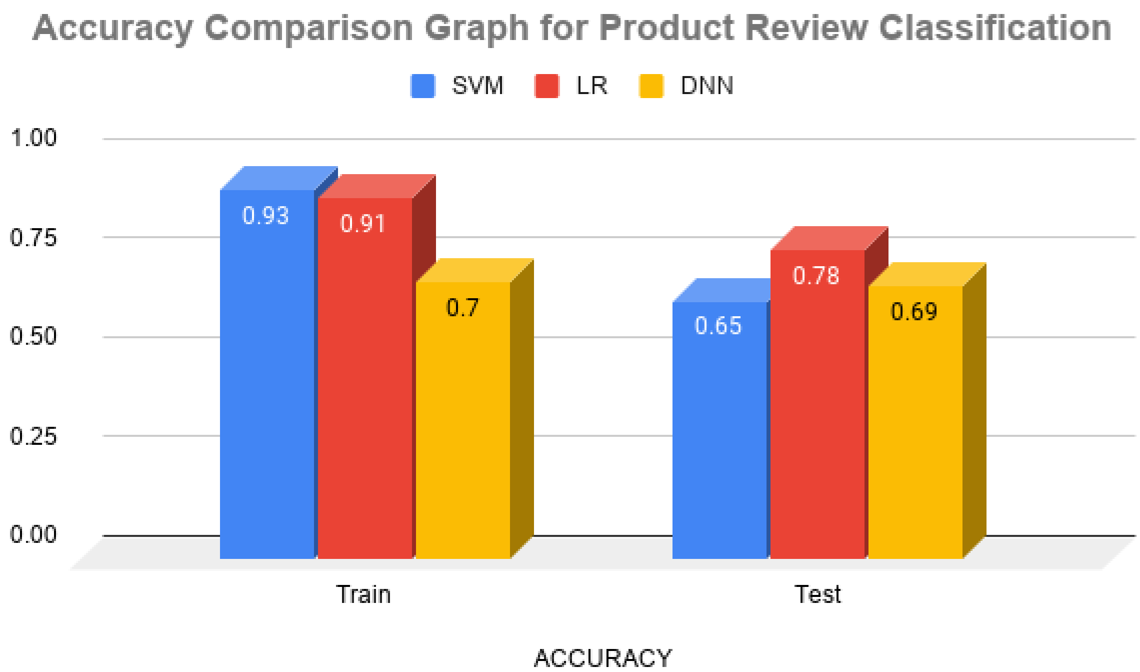

Figure 3.6: Accuracy Comparison for Product Review Classification

Table 3.21: Classification Report Comaprison for Sentiment Analysis

		Bad	Good
precision	SVM	0.85	0.91
	LR	0.81	0.85
	DNN	0.81	0.87
recall	SVM	0.89	0.88
	LR	0.81	0.89
	DNN	0.85	0.84
f1-score	SVM	0.87	0.89
	LR	0.81	0.85
	DNN	0.83	0.85

In Table 3.21 and 3.22 , the comparison of the classification reports are shown for the Sentiment Analysis and Product review Classification Respectively. In the case of Sentiment Analysis, the 'good' category of the SVM model shows the highest f1-score. The 'complain' category shows the maximum f1-score for the SVM model for product review classification. But the recall and precision score for the DNN models are close for both the Sentiment Analysis and Product Review Classification. So, the DNN model can be considered the best model for our application.

Table 3.22: Classification Report Comaprison for Product Review Classification

		appreciation	complain	recommended	wrong delevery
precision	SVM	0.92	0.73	0.78	0.78
	LR	0..63	0.77	0.62	0.65
	DNN	0.72	0.7	0.66	0.67
recall	SVM	0.65	0.92	0.79	0.46
	LR	0.57	0.72	0.72	0.59
	DNN	0.55	0.86	0.66	0.67
f1-score	SVM	0.76	0.82	0.78	0.58
	LR	0.6	0.74	0.67	0.68
	DNN	0.62	0.77	0.64	0.59

Overall, we can conclude that our proposed models have learned efficiently from the data. Indisputably, it is able to predict the quality of the online products and classify the product reviews into the given categories, accurately.

3.7 Comparison with Existing Works

We have compared our proposed work to existing work which is illustrated in Table 3.23. It can be observed that, in our proposed work, the mixed data-set, advanced machine learning techniques, and multistage classification has been done which is missing in the existing work.

Table 3.23: Comparison with Existing Works

Existing Works	Proposed Work
The datasets of existing works are either collected from Kaggle or scrapped from various sites e-commerce site which is only the Bangla text.[12]	The dataset we have collected are real-world datasets which are collected and labeled manually from the topmost e-commerce site of Bangladesh. It not only contains Bangla text. It also contains Romanized Bangla and English text.
Machine Learning algorithms such as Logistic Regression, Support Vector Machine, Gradient Dissent Classifier is applied to classify the sentiments. [15] [16]	In our proposed model, we have used Deep along with ML models Neural Network to classify the sentiments.
Researchers performed single stage classification.[15] [16] [15], [17]	We have performed double stage classification in our study.

Chapter 4

Conclusions

This chapter concludes our whole thesis work and directs some future works for further research.

4.1 Conclusions

In this study, We have aimed to propose a Deep Neural Network(DNN) model for predicting online product quality ('good' and 'bad') and classifying them into four categories based on product reviews of the customers. We have applied ML and DNN model on our won dataset. We have created a dataset that has been collected from the biggest online marketplaces in Bangladesh. We have performed some data preprocessing tasks to clean and prepare the data and performed feature extraction to extract numeric features from text data. To vectorize our text, we have applied TF-IDF and the FastText word embeddings. We have followed some proven methods from a few organized research studies to design our ML and proposed DNN model. In both of the DNN models, we have defined the hidden layers, number of neurons, optimization algorithms, and tuned the hyperparameters for obtaining the best result. We have used the accuracy matrix, precision, recall and f1-score as the evaluation parameters. To validate the traing phase, we have created a loss vs epochs curve that showed a consistent performance for all models. The testing performance that has been found for sentiment analysis is 0.84, and 0.69 for DNN model for product review classification those are very much promising.

4.2 Directions of Future Research

In future, we aim to incorporate more data and compare with different state-of-the-art NLP techniques.We only have applied Recurrent Neural Network but there is a more advanced form of DNN network such as Long Short Term Memory(LSTM), Convolutional Neural Network(CNN). Also, there are many transformation methods like BERT. We are aiming to Use these techniques to improve the results.

The thesis work related codes, dataset and the accepted paper can be found on this link:

https://github.com/Munnamm27/Thesis_work

References

[1] [Online]. Available: https://christophm.github.io/interpretable-ml-book/logistic.html

[2] [Online]. Available: https://www.datacamp.com/community/tutorials/svm-classification-scikit-learn-python

[3] [Online]. Available: https://www.google.com/url?sa=i&url=https%3A%2F%2Fdeeplearninguniversity.com%2Frelu-as-an-activation-function-in-neural-networks%2F&psig=AOvVaw14eZ5mZ7RD3qJZa6BVc59M&ust=1608035931375000&source=images&cd=vfe&ved=0CAIQjRxqFwoTCKCH_ea-ze0CFQAAAAAdAAAAABAD

[4] [Online]. Available: https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.medcalc.org%2Fmanual%2Ftanh_function.php&psig=AOvVaw0SS_a6QDjICpi_X34z1-gO&ust=1608036026201000&source=images&cd=vfe&ved=0CAIQjRxqFwoTCLDml5S_ze0CFQAAAAAdAAAAABAD

[5] [Online]. Available: https://en.wikipedia.org/wiki/Sigmoid_function

[6] J. Clement, "Covid-19 impact retail e-commerce site traffic 2020," Aug 2020. [Online]. Available: https://www.statista.com/statistics/1112595/covid-19-impact-retail-e-commerce-site-traffic-global/

[7] A. B. Parves, A. Al Imran, and M. R. Rahman, "Incorporating supervised learning algorithms with nlp techniques to classify bengali language forms," in *Proceedings of the International Conference on Computing Advancements*, 2020, pp. 1–7.

[8] A. Al Imran, Z. Wahid, and T. Ahmed, “Bnnet: A deep neural network for the identification of satire and fake bangla news.” 2020.

[9] F. Haque, M. M. H. Manik, and M. Hashem, “Opinion mining from bangla and phonetic bangla reviews using vectorization methods,” in *2019 4th International Conference on Electrical Information and Communication Technology (EICT)*. IEEE, 2019, pp. 1–6.

[10] O. Sharif, M. M. Hoque, and E. Hossain, “Sentiment analysis of bengali texts on online restaurant reviews using multinomial naïve bayes,” in *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*. IEEE, 2019, pp. 1–6.

[11] M. Rahman, S. Haque, and Z. R. Saurav, “Identifying and categorizing opinions expressed in bangla sentences using deep learning technique,” *International Journal of Computer Applications*, vol. 975, p. 8887.

[12] M. H. Alam, M.-M. Rahoman, and M. A. K. Azad, “Sentiment analysis for bangla sentences using convolutional neural network,” in *2017 20th International Conference of Computer and Information Technology (ICCIT)*. IEEE, 2017, pp. 1–6.

[13] M. Anand and R. Eswari, “Classification of abusive comments in social media using deep learning,” in *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)*. IEEE, 2019, pp. 974–977.

[14] N. I. Tripto and M. E. Ali, “Detecting multilabel sentiment and emotions from bangla youtube comments,” in *2018 International Conference on Bangla Speech and Language Processing (ICBSLP)*. IEEE, 2018, pp. 1–6.

[15] M. M. J. Soumik, S. S. M. Farhavi, F. Eva, T. Sinha, and M. S. Alam, “Employing machine learning techniques on sentiment analysis of google play store bangla reviews,” in *2019 22nd International Conference on Computer and Information Technology (ICCIT)*. IEEE, 2019, pp. 1–5.

[16] T. Rabeya, N. R. Chakraborty, S. Ferdous, M. Dash, and A. Al Marouf, “Sentiment analysis of bangla song review-a lexicon based backtracking approach,” in *2019 IEEE*

International Conference on Electrical, Computer and Communication Technologies (ICECCT). IEEE, 2019, pp. 1–7.

[17] M. G. Sarowar, M. Rahman, M. N. Y. Ali, and O. F. Rakib, "An automated machine learning approach for sentiment classification of bengali e-commerce sites," in *2019 IEEE 5th International Conference for Convergence in Technology (I2CT)*. IEEE, 2019, pp. 1–5.

[18] A. Z. Adamov and E. Adali, "Opinion mining and sentiment analysis for contextual online-advertisement," in *2016 IEEE 10th International Conference on Application of Information and Communication Technologies (AICT)*. IEEE, 2016, pp. 1–3.

[19] S. Ramaswamy and N. DeClerck, "Customer perception analysis using deep learning and nlp," *Procedia Computer Science*, vol. 140, pp. 170–178, 2018.

[20] R. E. Wright, "Logistic regression." 1995.

[21] S. Vishwanathan and M. N. Murty, "Ssvm: a simple svm algorithm," in *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No. 02CH37290)*, vol. 3. IEEE, 2002, pp. 2393–2398.

[22] J. Heaton, *Introduction to neural networks with Java*. Heaton Research, Inc., 2008.

[23] V. Peddinti, D. Povey, and S. Khudanpur, "A time delay neural network architecture for efficient modeling of long temporal contexts," in *Sixteenth Annual Conference of the International Speech Communication Association*, 2015.

[24] J. Moolayil, "Learn keras for deep neural networks. apress," 2019.

[25] S. Sharma, "Activation functions in neural networks," *Towards Data Science*, vol. 6, 2017.