Determining the effects of Consumer Sentiments on E-commerce sector using Sentiment Analysis : A Deep Learning Approach

MSc Research Project
Data Analytics

Ayusha Eknath Kashilkar
Student ID: x20239343

School of Computing
National College of Ireland

Supervisor: Dr.Abid Yaqoob

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Ayusha Eknath Kashilkar
Student ID:	x20239343
Programme:	Data Analytics
Year:	2022
Module:	Msc Research Project
Supervisor:	Dr Abid Yaqoob
Submission Due Date:	15/12/2022
Project Title:	Determining the effects of Consumer Sentiments on E-commerce sector using Sentiment Analysis : A Deep Learning Approach
Word Count:	7128
Page Count:	24

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Ayusha Eknath Kashilkar
Date:	29th January 2023

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Determining the effects of Consumer Sentiments on E-commerce sector using Sentiment Analysis : A Deep Learning Approach

Ayusha Eknath Kashilkar
x20239343

Abstract

In this modern world data stands the as one of the most valuable assets to companies. As humans are attached to their digital gadgets which generate immense amount of data per second. This generated data is used by companies to make analytical and financial decision. Ecommerce market stand as 4th on biggest industries in world with market value of 9.09 trillion dollar. As Amazon being on top of the ecommerce market, multiple purchases, transactions, and n number of users are active on amazon site at same moment of time. Companies have grown to keep customer first as policy. Users review about products are critical to company as from these review multiple insights can be found which can affect the company's value, marketing, and financial decision. In this research, Natural Language processing techniques are used to find out the sentiments from customer review on different sectors of amazon using multiple machine learning and deep learning techniques which can be used by manufactures for enhancing their products.

Keywords - Bi-directional Long Term Short Memory, Random forest, Support Vector Machine, NRCLex, GloVe, Ecommerce, text mining.

1 Introduction

The world is growing tremendously in digital era where people are attached to technology. We spend immense amount of time online searching for things, googling, learning, web shopping, etc. While doing all these task data is generated by us for the companies. In current year 2.5 quintillion bytes of data is generated every day[1]. This immense amount of data generation has made data a valuable assets to companies, as data can provide alot of informational insights. Analysing this data through multiple techniques can give informational insights which can benefit companies to take decision regarding marketing, product based and business based.

1.1 Motivation

There are multiple industries that have grown in this digital era, but E-commerce was an industries generated by the digital era stands on 4th largest industry in world. Everything

[1]https://techjury.net/blog/how-much-data-is-created-every-day/gref

is available on e-commerce sites from beauty products to clothes, from electronic tools to computers. People have left the traditional way of shopping and shifted to online shopping. Now since everything in online shopping is virtual the issue comes of trusting, as multiple vendors and manufactures sell the same product for different prices. People used to believe in Word of mouth while shopping traditional. Therefore, to bring in this concept the ecommerce started with reviews of product left by customers with star rating. Now peoples decision get influence by these reviews and star ratings of products may be it positive or negative. These review does impact companies value, market and financial decisions. By tackling review either positive or negative companies can create a good product for customer as well as a gain good client and market value.
In E-commerce market, Amazon being the top companies selling products across the world.[2] Multiple vendors and manufactures sell their products online on amazon. Amazon contains multiple sectors within the products to segregate the number of services it provide. Many products have been returned due to damage or bad products, some times good condition products are also returned due to review of products. Amazon destroys 130000 items a week including the products that vendor choose to house in the warehouse and the products returned. The in-house products in warehouse are not been sold due to bad reviews or overpriced value by vendor. This creates a loss in money as the returned item value needs to be paid back and item is not of any use and destroying so many items is a wastage too.
As reviews play important part in purchase and return of good, these reviews can be analysed using sentiment analysis to find out if positive or negative. Natural language processing is a technique which helps us to analyse the textual data and find out emotions behind it. Text classification and Sentiment analysis is a type of techniques used in natural language processing. The reviews of products can be used for further storing and purchasing of the product. In this research with the use of sentiment analysis using machine learning and deep learning on amazon reviews on different sector, informational insights can be found and used for betterment of product and sector of e-commerce. There are two sentiment analysis techniques - VADER (Valence Aware Dictionary and Sentiment Reasoner) and NRCLex used to determine the sentiments for review body. NRCLex can determine 2 sentiments and 8 emotions for text which can help to classify the exact emotion of text. The modeling techniques used in machine learning are Random forest (RF) and Support Vector Machine (SVM) and in deep learning Bi-directional Long Term Short Memory (Bi-LTSM) has been use. In neural network model for embedding weights Global Vectors for word representation (GloVe) has been used which is a pre trained word vector and has 6 billion tokens and 400K vocab.

1.2 Research Question

"Can Sentiment Analysis using machine learning and deep learning provide insights, which can be used to improve the E-commerce sector and products?"
This research work considers the following research objectives:-

[2] https://www.similarweb.com/top-websites/category/e-commerce-and-shopping/

1.3 Research Objective

1.3.1 Objective 1 :- Comparison of star ratings to their respective reviews

1.3.2 Objective 2 :- Evaluate the Sentiment analysis using machine learning

1. Evaluation the Sentiment Analysis techniques using Support Vector Machine
   (a) Evaluating Support vector machine model using NRCLex
   (b) Evaluating Support vector machine model using VADER (Valance Aware Dictionary and Sentiment Reasoner)
2. Evaluation of Sentiment analysis technique using Random Forest
   (a) Evaluating Random forest model using NRCLex
   (b) Evaluating Random forest model using VADER

1.3.3 Objective 3 :- Evaluate the Sentiment analysis using Deep Learning

1. Evaluate the Sentiment analysis techniques using Bi-directional Long Term SHort Memory (Bi-LTSM) neural network.
   (a) Evaluating Bi-LTSM model using NRCLex.
   (b) Evaluating Bi-LTSM model using VADER.

1.3.4 Objective 4 :- Evaluate the best 2 models from Objective 1 and Objective 2 on unknown dataset

1. Evaluating model 1 on unknown dataset.[3]
2. Evaluating model 2 on unknown dataset.

2 Literature Review

2.1 Sentiment Analysis using Machine learning

With the ease of shopping at home using internet, the e-commerce is growing at an exponential rate. Today almost every user, checks the reviews before investing in product or services, businesses also consider the client feedback to improve their products. Therefore, it's important for an organization to understand their stake holder sentiments in order to succeed the competitive market. In the study (Vanaja & Belwal 2018) and (Hassan & Islam 2021), sentiment analysis was done on amazon product reviews, and classification algorithms were applied to segregate the reviews based on polarity scores. The research (Kothalawala & Thelijjagoda 2020) uses SentiWordNet to calculate the positive, negative and neutral scores , and then the machine learning models based on Naïve Bayes(NB) Classifier and Support Vector Machine(SVM) were trained to predict the polarity in the review. Though these models performed really well, but SVM based model have large time complexity for training and testing, therefore with better computer resources deep learning model can be considered to achieve better accuracy with lower time complexity. There are several text messages that are conveyed and published via social media. The way people express themselves through text is unstructured. By using BERT and machine learning methods, the researchers were able to pinpoint what emotions was underlying

[3]The unknown dataset is amazons review dataset of electronics sector

the text from the IMDB movie review data in this research (Jain et al. 2022). Though the model had an impressive accuracy of 75% but this could have been improved by using RoBERTa process because BERT employs static masking, which means that the same section of the phrase is masked in each Epoch. In contrast, RoBERTa employs dynamic masking, in which various parts of the phrases are veiled for different Epochs.

The user reviews can be found every on every site and applications. Analysing the reviews effectively can provide great insights. Many applications are downloaded and reviewed by people. In this study (Fan et al. 2016), the authors have used the application reviews for processing sentiments. The paper only focuses on sentiment analysis and text classification. The three major task performed were sentiment prediction, polarity prediction and sentiment words extraction. These task were analysed using Navies Bayes algorithm. As for classification of sentiments there are multiple algorithms like pre trained models and neural network which could have been used. The polarity could have been found in different ways which might have added more value to tasks.
Multiple standards have been set for hotels, hotels are ranked on basis of hospitality and services they provide. Customer reviews about hotels can affect the hotels reputation, ranking and business. Analysing the hotel reviews can provide get insights for hotel to work on. In sentiment analysis the text mining process of feature vector has high sparseness and dimensionality problems when applied. To overcome this situation, in this research (Zhang & Yu 2017) the authors have used word2vector tool, ISODATA clustering and K mean clustering algorithm. By this the accuracy of the model has been 0.25% higher that other analysis based on text feature vectorization and K-mean clustering. The draw back of ISODATA clustering is sensitive to the initial cluster center selection which might change the results. Also the methods used usually lose the order of words which is loss of semantic information.

2.2 Sentiment Analysis using Deep learning

The rapid communication has become an integral part of our lives, whether texting on WhatsApp, commenting on Facebook or tweeting on twitter. But the drawback of text communication is that sometimes it becomes difficult for the receiver to understand the senders' emotions. In the research (Goel & Batra 2020), the deep learning approach to classify Hindi tweets into positive, negative, or neutral sentiments is taken by the researchers to understand the emotions of the tweeters. The Logistic regression (LR) was used to extract the emotions from the sentence and then the neural network model was trained, but the accuracy can be improved but using RoBERTa pretrained model because LR only performs lexical analysis (understanding the meaning of words), whereas RoBERTa extracts the emotion from the sentence based on the context and meaning of the whole text.
For governments before introducing a policy or a law, for an organization to improve their products and services, and for an investor to know potential time to put money in the market or pull it, it becomes important to understand the general emotion of the society or the market. This can be done by analysing what people are thinking and talking about on the social media platforms, this is where sentiment analysis comes into play. The project (Jianqiang et al. 2018), focuses on performing semantic analysis on the tweets using unsupervised learning for determining the positive and negative polarity of the textual tweet. The research showed that the convolution neural network based models are better in predicting the sentiment polarity than machine learning based classification models.

In China agriculture is a big sector, as they have started to sell the products online. There have been a lot complex emotions on experiences and products. The sentiment analysis using deep learning has always been easy with feature extraction and English text segmentation, which is not the case with Chinese. In this research (Zikang et al. 2020), the authors have addressed the issue of Chinese sentiment analysis polarity score by creating neural network models like Bi-LTSM Text-CNN and BERT. The Text-CNN has well performed in all three models.

The languages become a barrier in sentiment analysis as the English language can be segmented faster and effectively. The study (Nath Saha & Senapati 2020) has performed Deep learning approach for sentiment analysis on multilingual dataset to overcome the challenges. They have used regularization method and cross validation method to choose the optimal parameters for the deep learning method. Long term short memory basic model was applied with 4 layers which have given moderate results on datasets. The draw backs in this paper were to have a better deep learning algorithm design for better prediction.

The enormous generation of web data in today world created by social media platforms, communities, and online reviews. People have been open about their attitude, thoughts, opinions and feelings about various product manufactures, brands, etc. The Speech recognition domain has lacked information for sentiment classification. In this research (Vimali & Murugan 2021) , Recurrent Neural network has been used to overcome different challenges like text classification on multilingual dataset and verbal assessment. The model has achieved an 86% which was due to use of Word to Vector. The methodology used is very basics of model creations where in text mining section could have been enhanced using new methods available in market like Bi-grams, Global vectors, BERT models.

Social media has been generating ample amount of data per hour. This data can be viewed as a chance to understand the customer's experience, needs and views. In the study (Cheng & Tsai 2019), a novel approach for extracting sentiments from social media using deep learning for short messages. The data has been web scrapped and for sentiment analysis Microsoft text analytics API and NLTK text blob have been used with deep learning models but to improve the model, pre trained models like ROBERTA , VADER could have been a good choice to extract emotions. As per results have highest accuracy of 87.17

2.3 Sentiment Analysis using Hybrid and pre trained models

Due to pandemic, the tourism industry has been affected worse. It is now slowly returning on the track, but still restrictions in some places and covid-19 virus still prevailing in the environment, the tourists are still worried about their travels, therefore it becomes crucial for a travel business to understand it's the emotions of the traveller in order to provide better services and experience while reducing the losses. The study (Wang & Yu 2022), uses BERT model on the tourist feedbacks collected from Trip.com. With the accuracy of 87%, the model was able to classify and generate the deeper insights from the feedback. As discussed in above studies, the RoBERTa based methodology combined with NRCLex can be used to derive more emotions from the text and thus models can be able to predict more accurately.

The technology has taken over as everyone has mobile phones. The mobile phone companies have been increasing features and user experience by launching new phones monthly. In this paper (Srivats Athindran et al. 2018) sentiment analysis has been done on two leading mobile phone brands Vivo and Oppo. They have used Lexicon based sentiment analysis with Naïve Bayes algorithm. As a draw back only one model has been applied, there is no comparison. With lexicon based sentiment analysis other models like deep learning and Bert modules could work better.

Online shopping has been on of the benefits of internet for customers. The customers do represent their views and opinions in review sections. These reviews affect the product sales and online shopping site credibility. In these study (Gaikar n.d.) and (Yang & Yang 2020) , the authors have used SPYDER for sentiment classification and using spacy for aspect based sentiment analysis techniques. There are two model applied pre trained models like BERT, XLNET and RoBERTa. In this application of neural networks with SYPDER might give optimal solution with better accuracy.

The news has been available digitally on various application. People do express their comments on digital platform though comments. Since the traditional method of gathering the data through forms and reviews have been extinct now, data is available online. As Sentiment analysis have been performed on different sectors but not majorly have been explored on new comments. In this research (Mukwazvure & Supreethi 2015) the authors have web scrapped the news comments from online and applied sentiment lexicon for polarity score that is classified into positive, neutral and negative. They have used Support vector machine and K-Nearest Neighbour models. The KNN has failed to identify the neutral sentiment. The results could have been better as classification on small size of data did not predict well also neutral sentiments were not detected majorly.

As online market has been increasing the products and services provided by the sites have been increased too. Amazon being a good option for products purchases online. Mobile phones have been used by elder people as well as students in school, which has made the mobile phone market huge. In this study (Wedjdane et al. 2021), the effects of customer reviews on mobile phones sales online were considered. The accuracy of model achieved is 76.80%. They also confirms the strong correlation of review votes and price and rating and sentiments data does validate each other. In build VADER model has been used for analysing sentiment analysis. The paper lacks comparison of models but does have an excellent exploratory data analysis done to find insights.

3 Methodology

The KDD[4] model is an iterative process which help us to get more appropriate and different results by adding and transforming new data information, refining the data mining, changing the evaluation method accordingly for next iterations. The steps of KDD are mentioned in the Figure 1.

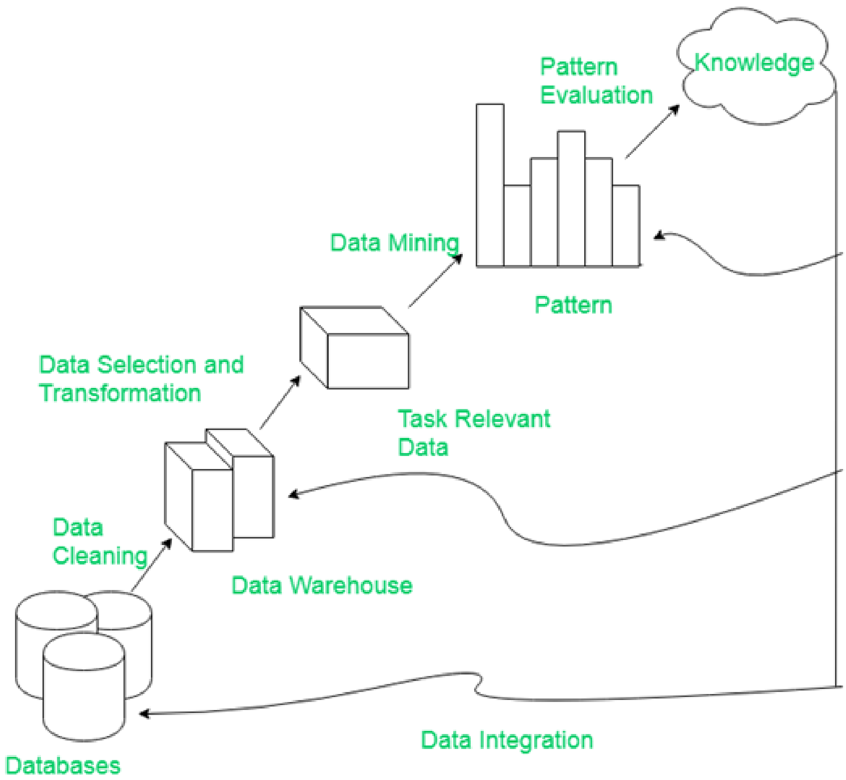

Figure 1: KDD process

[4]https://www.geeksforgeeks.org/kdd-process-in-data-mining/

3.1 Data Gathering

The data was gathered from Amazon web services official site for reviews dataset[5]. The data was downloaded into ZIP folders and then unzipped on the system. The files were in tab separated format, where they were changed to CSV format for preference for further processing. There are total 5 files containing the following sectors – Electronics, Apparels, Personal Care Appliances and Mobile Electronics.

All of the file contains the following columns : Marketplace, customer _ id, Review_id ,Product_id, product_parent, product title, product_category, star_rating, helpful_vote, total_vote, vine, verified_purchase, preview_headline, Review_Body, Review_date. The Electronics dataset contains 82075 records, Apparels contains 82074 records, Personal Care Appliances contains 85924 records whereas, the Mobile Electronics contains 104855 records.

3.2 Data Preprocessing

The complete Data processing and Analysis are done using Python programming language with Google Colab. The data was loaded into google colab using python. As data cleaning is an important process for further modelling. All 5 datasets were loaded into different files for pre-processing. All data files contained 7 columns additional with no values which were dropped. After that the datasets were checked for null values, the null values were found in review body, review headline, votes and rating.
These null values cannot be filled with any other data, the null values were dropped. Just to confirm and check the data types of columns. There are 10 object columns containing string values and 3 float columns and 1 integer values. To clean the reviews data from modelling, the data words were changed to lowercase using lower function of string library. After this from review the punctuation and numbers were removed using replace function of string library and regular expression

3.3 Design Specification

The Figure 2 shows the flow chart of the architecture of the research, The architecture steps are as follows : -

- The dataset was downloaded from amazon dataset site and loaded into Google Colab using python. The CSV files were uploaded to Google Drive for optimum use of them, as at runtime google colab discard locally saved files.
- The data was then pre-processed by removing unnecessary columns, cleaning data by removing null values and junk text values in string.
- Natural language processing technique of text processing was applied where we removed stopword, applied lemmitization and then NRClex and vader_lexicon were used to determine sentiments of text.
- The data had 2 parts 1. NRCLEX sentiment dataset and Vader_Lexicon dataset, which was divided into 80% train and 20% test data. Models like Support vector mechanism, Random forest and Bidirectional LTSM were applied on each part.

[5]https://s3.amazonaws.com/amazon-reviews-pds/tsv/index.txt

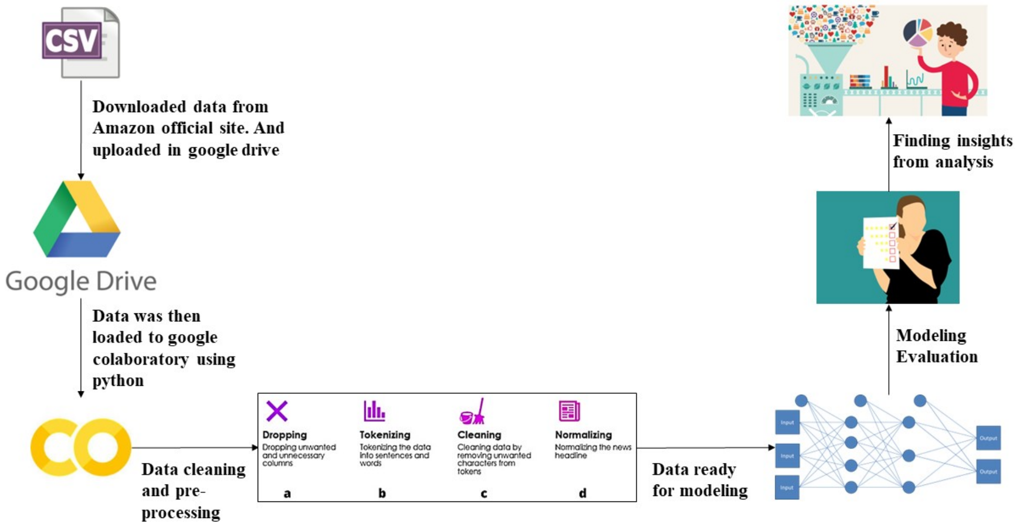

Figure 2: Design Specification Flow chart

- The each model was applied on single dataset – SVM on Apparels dataset, RF on Mobile electronics and Bi LTSM on Personal Care Appliance. Each model was saved

- The saved models were tested on Electronics dataset for evaluation.

Tools used are as follows:-

- Google Drive :- CSV data was uploaded into Google drive for easier access of files and saving. Google drive is a free tool and can hold up to 15 GB of data. The saved models were also stored in google drive

- Google colab :- Google colab is an open source website for python code execution online. As google colab offers RAM , disk space and a amount of GPU for free users. The files are autosaved and is an interactive tool.

3.4 Evaluation Methods

The evaluation methods(Tsai 2010) used in this research are as follows:-

- *Accuracy* :- The accuracy is calculated by dividing the correct predictions of model by the total number of predictions.

- *Precision* :- Precision is a ratio of true positive over addition of true positive and false positive

- *Recall*:- Recall is the ration of true positive over addition of true positive and fase negative

- *Confusion Matrix*:- Confusion matrix provides summary of prediction results containing True positive, true negative, false positive and false negative

- *Accuracy and loss graph*:- the loss graph shows how good or poorly the model was trained . the accuracy graph measures the models performance.

4 Implementation

The implementation is divided into 2 parts. Part 1 consists of training models and saving them. part 2 consists of loading saved model and testing.

4.1 Implementation : Part 1

The Implementation part 1 consists of machine learning and deep learning model training and testing. There are 2 Sentiment analysis techniques used therefore the flow is divided into two parts further as shown in Figure 3

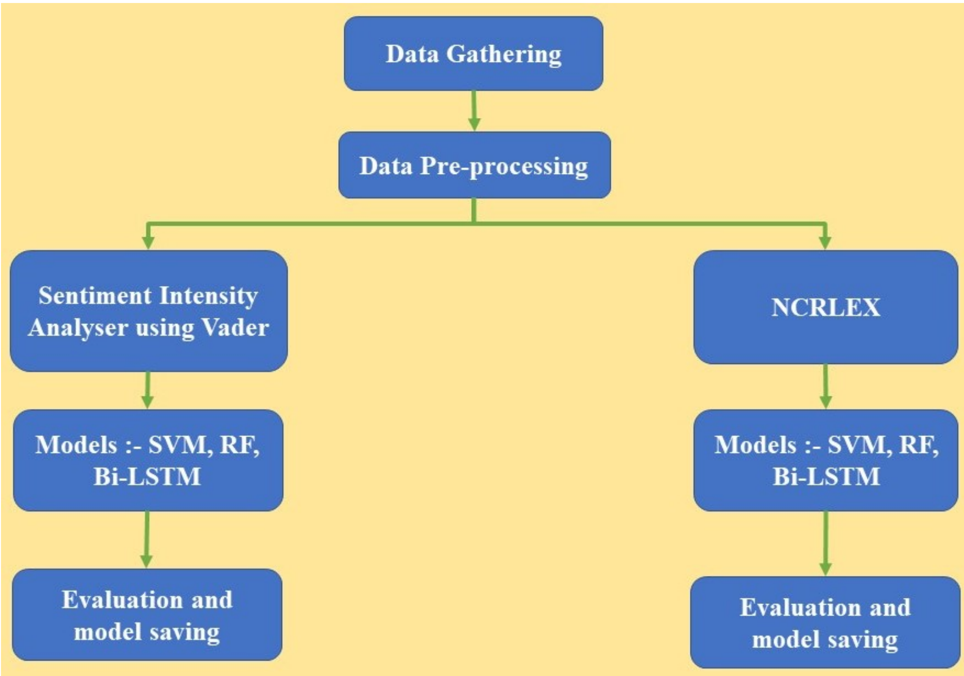

Figure 3: Implementation Part 1 Flow chart

4.1.1 Text Mining

Text mining is conducted on data to find relevant insights and transform data for further analysis. Its uses Natural Language processing for getting insights from data. The Natural Language processing techniques like lemmatization, stop removal, parsing and tokenization can be used. In this research to transform our data with text mining for further analysis we have used the following techniques: -

- *Stop word removal* :-
  It is one of commonly used technique in NLP for data pre-processing before analysis. Stopword removal is basically removal of frequently occurring words in the sentence or corpus. In sentences we use “I” , “in” , “am” , “the” for grammatically correctness but originally these words serve no other purpose or give any other meaning, Such kind of words are omitted by stopword removal. In this thesis, from the reviews body the stop words are removed using stopwords from NLTK corpus library.

- *Tokenization* -
  In review body there are long sentences which might be difficult to interpret the meaning of it, therefore tokenization is applied in Natural Language processing. Tokenization divides sentences and paragraphs into smaller sentences from which meaning can be easily assigned. Word tokenization is a type of tokenization which

splits the sentences where delimiters are used (“,” , “;”) and pauses in speech. In this research with word_tokenize and sent_tokenize from nltk tokenize library is used. New columns were created “tokenized_word” and “tokenized_sent” where the results were saved These tokenized words are saved in an array of words for further process of lemmatization.

- *Lemmatization* -
  Lemmatization is a process of text normalization by changing words to its root base word for example leafs and leaves changed to leaf. These words are further used to detect emotion in analysis. In the research WordNetLemmatizer library is imported from nltk.stem which searches for lemmas of words from Wordnet Database

4.1.2 Sentiment analysis

Sentiment analysis is a Natural Language processing technique through which we can determine the sentiments of text to be positive or negative. In this research, the sentiment analysis was applied on the review body. There are 2 types of sentiment analysis applied in this research to find the emotion of text, they are as follows:-

- *Sentiment Intensity Analyzer using Vader*
  Valence Aware Dictionary for Sentiment Reasoning (VADER) is used to detect the polarity score and intensity of emotion, with polarity score we can determine the positive and negative emotion. The sentiment intensity analyzer function was used from the NLTK.sentiment.vader library. In this research polarity score was obtained from Sentiment intensity analyser on compound value. The compound value is computed by normalizing the negative, positive, and neutral value of the text. These scores were stored in a new column polarity score. With lamba function and looping the polarity score was segregated on basis if polarity score is greater than 0 then its positive else its negative.
- *National Reaersch Council Canada(NRC) affect lexicon (NRCLex)*
  National Reaersch Council Canada(NRC) affect lexicon (NRCLex) is a python library approved by MIT which is able to determine the sentiments from given body of text. NRCLEX library can detect 2 sentiments and 8 emotions – anger, fear, anticipation, trust, surprise, positive, negative, sadness, disgust and joy. It contains 27,000 words and NLTK library wordnet synonym sets for understanding. In this research NRCLex was used, as it can predict multiple emotions and these emotions can be used to know the product value in customers mindset. As for further modelling the target variable had to have 2 value in binary the emotions were separated into positive and negative. The trust, surprise, positive, joy, aniticipation were changed to positive and sadness, negative, disgust, fear and anger were changed to negative.

4.1.3 Models Application

- Feature Extractor
  - *The Term Frequency Inverse Document Frequency (TF-IDF)* is an algorithm to convert text into numerical representation which is used to fit models. It is measured by originality of word comparing with the number of times the word has appeared in a document.
- Employed Model
  1. *Support vector machine*
     Support Vector Machine is a supervised algorithm. It can be used for both regression and classification. Each item in data is represented as point in the n dimensional space. The point is the value of each feature specific coordinate. Then SVM classify the data by selecting the best hyperplane/line that

differentiates two classes. In this research I have applied SVM algorithm as seen in multiple papers it has given effective results in classifying the sentiments of customer. The svm was imported from sklearn library. There were 2 parameters applied to SVM model are as follows:-

(a) Kernal = rbf
As kernel was not specified in the model the default value of kernel that is rbf was taken.

(b) Gamma = 0.001
The Gamma value originally controls the influence of distance of a single training point. The low gamma value results in more points getting group together as it suggests the large similarity radius.

(c) C = 100.
The C value is set to 1 by default. If the dataset contains nosiy observations than C can be decreased, which will correspond to more regulation.

The SVM model was then fit to the two train models and then tested on same dataset test dataset. The xxx data set was split into 70:30 ration of train and test, as there are 2 target values the data has been split twice – once with y value of vader sentiment and other with Y value of NRCLex sentiment.

2. *Random Forest*
Random Forest operates as an ensemble. The algorithm contains multiple decision tree that gives a class prediction. The result is based on the most votes predicted from a class of decision trees. The fundament concept is the wisdom of crowds. Since there are multiple decision tree, this prevent from individual errors unless every decision tree make the same error. The Random forest classifier was imported from sklearn.ensemble library. As no parameters were specified in the random forest model the by default parameters were taken , those are as follows:-

(a) N_estimators: int, default = 100
It defines the number of tree in forest. By default it is 100 as version is 0.22. In older version it was 10

(b) Criterion (default = gini)
The quality of spilt is measured in this function. The by default value is gini

(c) Max_depth
It is the maximum depth of tree. By default it is None

(d) Min_Samples_split
Its is minimum value of sample required to spilt internal node. By default is 2

(e) Min_samples_leaf
It is at leaf node the minimum samples that are to be . By default the value is set to 1

(f) Max_feature
It is when looking for best split to consider this number of features. By default the value is sqrt. Therefore the equation is max features = sqrt(n_featues)

(g) Bootstrap
By default the value is True, therefore the samples are used to build trees. The whole data set is used when the value is False

(h) Random_state
The value of random_state controls the randomization of the bootstrap samples used while building tress, as well as the sampling of features to take into account when looking for the optimal spilt at each node. By default value is None, when bootstrap is True.

As random forest has been used in multiple papers and given optimal results, this has been applied in this research. The apparel data set was split into 70:30 ration of train and test, as there are 2 target values the data has been split twice – once with y value of vader sentiment and other with Y value of NRCLex sentiment.

3. *Bi directional Long Term Short Memory*
   LTSM is a type of recurrent neural network. It is used to recognize the relation between beginning and end values. Bi-directional LTSM is primarily used on natural language processing. In BI directional LTSM the input is from both direction which makes it capable of utilizing information from both sides. The output is combined from both LTSM layers in multiple ways like average, multiplication, sum or concatenation.
   In this research considering the benefits of Bi-LTSM we have applied. The Mobile electronics data set was split into 70:30 ration of train and test, as there are 2 target values the data has been split twice – once with y value of Vader sentiment and other with Y value of NRCLex sentiment. The X_train and X_test were tokenized using the Tokenizer.texts_to_sequence() method and then data was padded at the end using the sequence.pad_sequences() function. To create the weights for embedding layer , an embedding matrix was created using the Global vectors for word representation GloVe. Glove is an unsupervised algorithm which is used to produce vector representation for words from corpus. The multiple layers of Neural network model are as follows:-
   (a) Embedding layer
   For applying Machine learning model, the textual data needs to be converted to numbers. Normally this is done by one hot encoding to convert the categorical features into 0's and 1's. But if this is applied on textual data it will create 1000 features for 1000 words which is not feasible. Therefore embedding layer enables us to create numerical value of fixed length vector for each word which also helps in reducing dimension. As show in fig the model summary The parameters used in embedding layer are as follows:-
      i. Input_dim
      It is the size of vocabulary. The value is previously fetched from the length of tokenize word index into max_vocabulary attribute =1
      ii. Output_dim
      It is the length of vector of each word. The value is 300 Weights
      It is maximum length of sequence. The weights are the embedding matrix created from glove.
      iii. Input_lenght
      A review attribute was passed with embedding layer containing input shaped as (1500,)
   (b) Bi-directional LTSM Layer
   It has LTSM layer as argument in layers containing input value of 300
   (c) Dropout (0.5)
   The drop out layer prevents from overfitting. It describes the frequency rate of each step in training. 0.5 is our rate value.
   (d) Flatten()
   The flatten layer is used to flatten the input. It changes the multiple dimensional input into single dimension.
   (e) 5. Dense(64, activation = "relu", kernel_initializer "he_normal", kernel_regularizer=l2(0.001))
   In Dense layer the neurons of every layer is connected to its preceding layer. It is deeply connected with its previous layer. The parameters of dense layer are as follows:-
      i. Units – 64
      It defines the size of output from the dense layer, In here the value is

64

ii. Activation – relu
It is used for transforming the input value of neurons. In here the value mentioned in relu which rectifies the linear unit of activation function

iii. Kernel_initializer –
This is used to initialize the kernel weights matrix. The value mentioned is Henormal , it a predefined initializers in keras, it takes samples from truncated normal distribution on 0 with standard deviation measured as square root of 2 divided by number of input units in weight

iv. Kernel_regularizer
It is used to regularize the kernel weight matrix. The value mentioned is 12

(f) Dense(8, activation = “relu”, kernel_initializer = ”he_normal”, kernel_regularizer = l2(0.001))
In this dense layer , the unit value is 8

(g) 7. Output layer = Dense(1, activation = sigmoid)
The last output layer contains a dense layer with following parameter Units =1 It is the size of output layer and Activation is sigmoid

The model was compiled on by following parameters

(a) Optimizer = adam
The optimizer is adam optimization which is a stochastic gradient descent method.

(b) Loss = binary_crossentropy
It is to compute the quantityof model to seek minimize during the training.

(c) Metrics = accuracy
To evaluate on metrics of accuracy.

The model was fit on X_train and y_train with batch size of 128, with 15 epochs and validation dataset of X_test and y_test. Model was saved for further testing.

4.2 Implementation : Part 2

As shown in Figure 4, The best model out of 3 were taken in here for testing on Electronics and Home Entertainment. The SVM and Random forest model had performed well based on accuracy. The models were save in ṡav format. The models were than loaded by using the joblib library. There are total 4 model : Random forest with nrclex , Random forest with SIA Vader, SVM with NRCLex and SVM with SIA Vader. The electronics dataset was cleans and processedṪhe data was then spilt into 70:30 using the train_test_split from sklearn.model_selection library. Then these four models were applied on electronics dataset for testing on unknown dataset.

5 Results and Evaluation

5.1 Evaluation

5.1.1 Comparing star ratings to their respective reviews

As shown in below Figure 5, From the dataset there are 5233 positive sentiment reviews marked as star rating as low as 1. Also with rating 5 are there negative sentiment reviews in total 3090. Which contradicts the star ratings and reviews sentiments

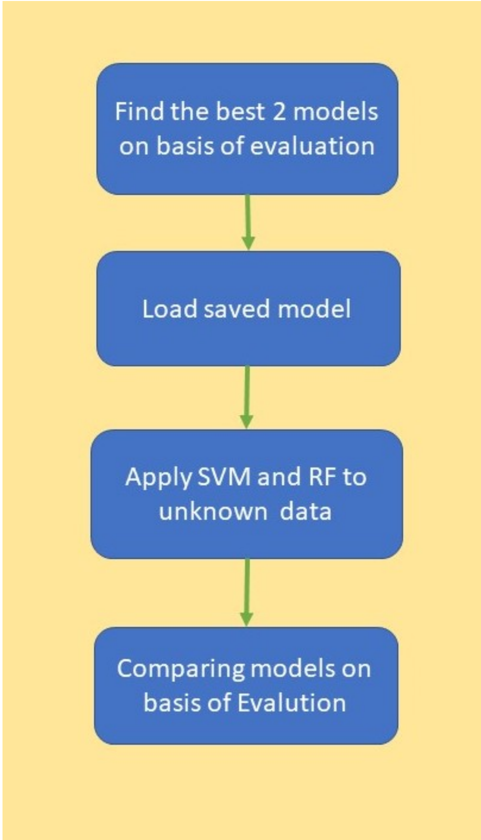

Figure 4: Implementation Part 2 Flow chart

```
#checking how many positive sentiments have star rating low as 1
a = Positive_sent.query('star_rating == 1').sort_values('sentiment_label',ascending=False)['tokenized_sent']
print(len(a))

5233

Positive_sent.query('star_rating == 1').sort_values('sentiment_label',ascending=False)['tokenized_sent'].values[5]

['I bought based good reviews I read Unfortunately case getting pay even It burns batteries like crazy It lasted we
amazon looking replace already']

#checking how many negative sentiments have star rating low as 5

a = Negative_sent.query('star_rating == 5').sort_values('sentiment_label',ascending=False)['tokenized_sent']
print(len(a))

3090
```

Figure 5: Comparison of ratings and reviews

5.1.2 Experimenting Sentiment Analysis using Support vector machine model

The Support vector machine did not take much time to train models on 85926 rows of data. There are two models of support vector machine – NRCLEX sentiment analysis and SIA using Vader sentiment analysis.

1. Support Vector Machine using NRCLEX sentiment analysis
   The data model was split in 80:20 ratio of train and test. The model was trained on training dataset. The model accuracy is 86.42% on the test dataset as shown in Figure 6.

- Precision:- The number of positive reviews predicted by the model is 86% whereas the The number pf reviews predicted that are negative by model is 89%
- Recall :- The number of actual positive reviews that model predicted correctly are 98% whereas the number of actual negative reviews that model predicted correctly is 55%.
- F1 score mentioned for negative reviews in 68% whereas for positive review is 91%.

	precision	recall	f1-score	support
neg	0.89	0.55	0.68	3920
pos	0.86	0.98	0.91	11067
accuracy			0.86	14987
macro avg	0.87	0.76	0.80	14987
weighted avg	0.87	0.86	0.85	14987

Figure 6: SVM using NRCLex : Classification report

2. Support Vector Machine using SIA using Vader
   The data model was split in 80:20 ratio of train and test. The model was trained on the training dataset. The model has accuracy of 87.40% on test dataset. As shown in Figure 7
   - Precision:- The number of positive reviews predicted by the model is 87% whereas the The number pf reviews predicted that are negative by model is 86%
   - Recall :- The number of actual positive reviews that model predicted correctly are 99% whereas the number of actual negative reviews that model predicted correctly is 27%.
   - F1 score mentioned for negative reviews in 41% whereas for positive review is 93%.

	precision	recall	f1-score	support
neg	0.86	0.27	0.41	2442
pos	0.87	0.99	0.93	12545
accuracy			0.87	14987
macro avg	0.87	0.63	0.67	14987
weighted avg	0.87	0.87	0.85	14987

Figure 7: SVM using VADER : Classification report

As shown in Figure 8 The confusion matrix , the 622 cases were predicted and are positive, whereas 1780 were predicted positive but were false. The 12437 case were predicted negative and ts true. The 108 cases were predicted as negative and it was false.

5.1.3 Experimenting Sentiment Analysis using Random forest model

The model took a lot of time to train 765155 items. There are two model of Random forest with NRCLEX sentiment analysis and SIA using Vader sentiment analysis

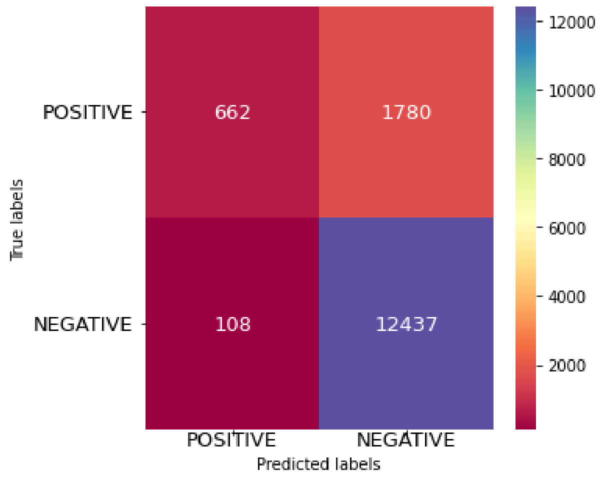

Figure 8: Confusion matrix for SVM

	precision	recall	f1-score	support
neg	0.91	0.56	0.69	41331
pos	0.86	0.98	0.91	111700
accuracy			0.87	153031
macro avg	0.88	0.77	0.80	153031
weighted avg	0.87	0.87	0.85	153031

Figure 9: RF with NRCLex : Classification Report

1. Random forest using NRCLex
   The data model was split in 80:20 ratio of train and test. The model was trained on the training dataset. The model has accuracy of 87% on test dataset. As shown in Figure 9

   - Precision:- The number of positive reviews predicted by the model is 86% whereas the The number pf reviews predicted that are negative by model is 91%
   - Recall :- The number of actual positive reviews that model predicted correctly are 98% whereas the number of actual negative reviews that model predicted correctly is 56%.
   - F1 score mentioned for negative reviews in 69% whereas for positive review is 91%.

   As shown in Figure 10 The confusion matrix , the 23135 cases were predicted and are positive, whereas 18196 were predicted positive but were false. The 109467 case were predicted negative and ts true. The 2233 cases were predicted as negative and it was false.

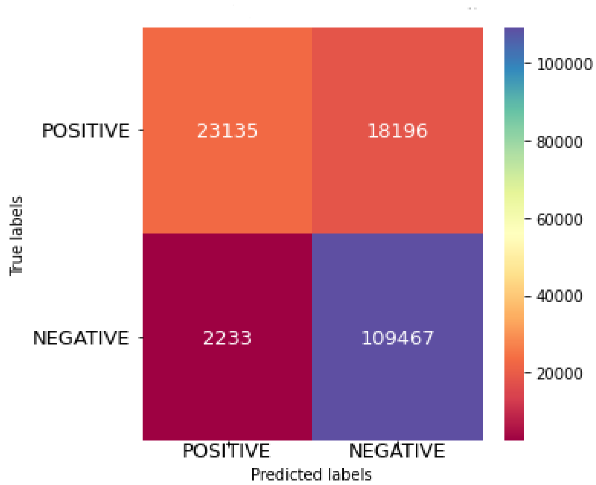

Figure 10: Confusion matrix for Random forest using NRCLex

Random forest using Vader
The data model was split in 80:20 ratio of train and test. The model was trained on the training dataset. The model has accuracy of 94% on test dataset. As shown in Figure 11

	precision	recall	f1-score	support
neg	0.93	0.31	0.47	12633
pos	0.94	1.00	0.97	140398
accuracy			0.94	153031
macro avg	0.94	0.66	0.72	153031
weighted avg	0.94	0.94	0.93	153031

Figure 11: RF using Vader Classification Report

- Precision:- The number of positive reviews predicted by the model is 94% whereas the The number pf reviews predicted that are negative by model is 92%
- Recall :- The number of actual positive reviews that model predicted correctly are 100% whereas the number of actual negative reviews that model predicted correctly is 31%.
- F1 score mentioned for negative reviews in 47% whereas for positive review is 97%.

As shown in Figure 12 The confusion matrix , the 3939 cases were predicted and are positive, whereas 8641 were predicted positive but were false. The 140106 case were predicted negative and ts true. The 345 cases were predicted as negative and it was false.

5.1.4 Experimenting with Sentiment Analysis using Bi-LTSM

The Bi-Directional LTSM did not take much time to train models on 66627 rows of data. There are two models of Bi-directional LTSM – NRCLEX sentiment analysis and SIA using Vader sentiment analysis.

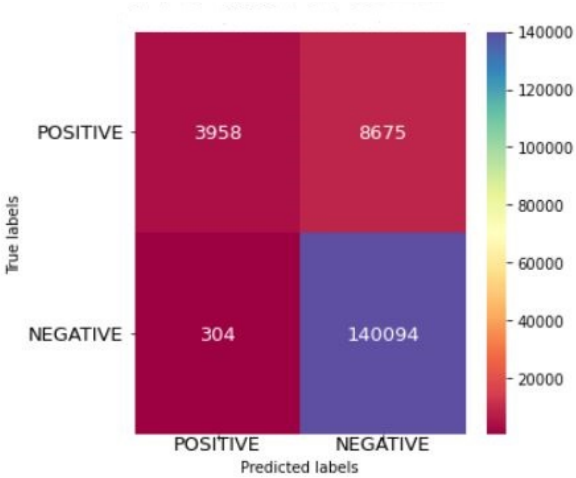

Figure 12: Confusion matrix for Random forest using VADER

1. Bi-LTSM using SIA using Vader
   The data model was split in 70:30 ratio of train and test. The model was trained on the training dataset. The model has accuracy of 80.29% on test dataset. Below figure 13 shows the model building with 15 epochs. As shown in Figure 14, the train

Figure 13: Model execution for Bi-LTSM with SIA VADER

   loss has a dropped till 5th epoch to a point of stability and its same for validation loss where its stable. The stable loss gap in between train and test loss is called generalization gap which is minimum in fractions in our model loss.The training dataset loss should be lower than validation dataset but in our case its vice versa Stating that model is a good fit but a little over-fitted. The accuracy of the model for train and test is quite stable at 79.98 and 80.29 respectively.

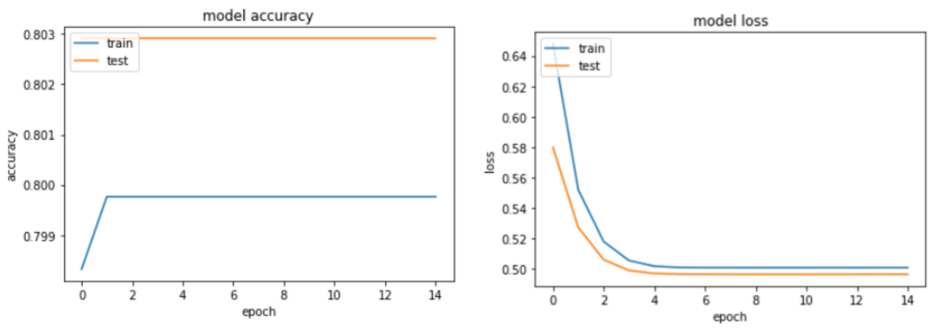

Figure 14: Accuracy and loss graph for Bi-LTSM with SIA VADER

2. Bi-LTSM using NRCLex
   The data model was split in 70:30 ratio of train and test. The model was trained

on the training dataset. The model has accuracy of 71.79% on test dataset. Below Figure 15 shows the model building with 15 epochs.

```
Epoch 9/15
521/521 [==============================] - 128s 246ms/step - loss: 0.5953 - accuracy: 0.7175 - val_loss: 0.5949 - val_accuracy: 0.7179
Epoch 10/15
521/521 [==============================] - 128s 246ms/step - loss: 0.5953 - accuracy: 0.7175 - val_loss: 0.5949 - val_accuracy: 0.7179
Epoch 11/15
521/521 [==============================] - 128s 246ms/step - loss: 0.5953 - accuracy: 0.7175 - val_loss: 0.5949 - val_accuracy: 0.7179
Epoch 12/15
521/521 [==============================] - 128s 246ms/step - loss: 0.5953 - accuracy: 0.7175 - val_loss: 0.5949 - val_accuracy: 0.7179
Epoch 13/15
521/521 [==============================] - 128s 246ms/step - loss: 0.5953 - accuracy: 0.7175 - val_loss: 0.5949 - val_accuracy: 0.7179
Epoch 14/15
521/521 [==============================] - 128s 246ms/step - loss: 0.5953 - accuracy: 0.7175 - val_loss: 0.5949 - val_accuracy: 0.7179
Epoch 15/15
521/521 [==============================] - 128s 245ms/step - loss: 0.5953 - accuracy: 0.7175 - val_loss: 0.5949 - val_accuracy: 0.7179
```

Figure 15: Model execution for Bi-LTSM with NRCLex

As shown in Figure 16 the train loss has a dropped to a point of stability and its same for validation loss where its stable. The stable loss gap in between train and test loss is called generalization gap which is minimum in fractions in our model. Stating that model is a good fit. If the model is continued to train on this it will lead to over fitting. As shown in Figure 16 for accuracy the train model has been trained well to have stable accuracy at end with 71.75% accuracy in training and constant 71.79% accuracy in validation.

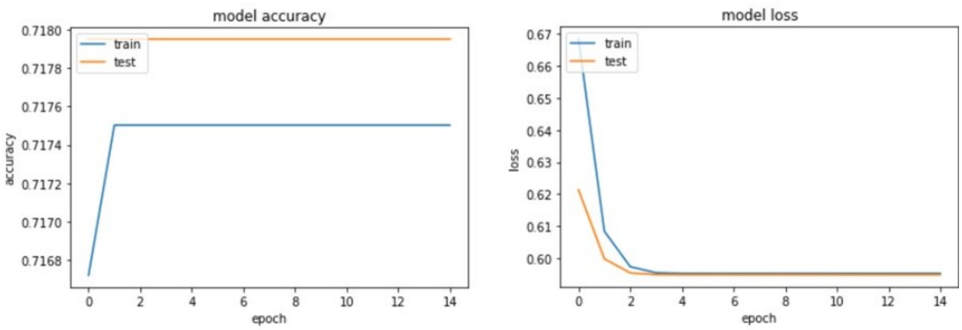

Figure 16: Accuracy and loss graph for BI-LTSM with NRCLex

5.1.5 Comparison of best models on unknown dataset

As Random forest and SVM had performed well, the have been applied on electronics database. Both types of sentiment analysis has been applied for better results

(a) Evaluating Sentiment analysis techniques with Random Forest

- Both of the sentiment techniques - Vader and NRCLex with Random forest have performed well with 94.17% and 94.10% respectively.
- As shown in Figure 17 the Precision of Vader with RF model is 97.84% and recall is 94.17%

```
Accuracy :  0.9417261297995243
Precision :  0.9784201977888956
Recall :  0.9417261297995243
              precision    recall  f1-score   support

           0       0.93      0.32      0.48     12660
           1       0.94      1.00      0.97    140376

    accuracy                           0.94    153036
   macro avg       0.94      0.66      0.72    153036
weighted avg       0.94      0.94      0.93    153036
```

Figure 17: Classification report for Random forest using Vader

- As shown in Figure.18 The confusion matrix , the 3636 cases were predicted and are positive, whereas 8724 were predicted positive but were false. The 140083 case were predicted negative and ts true. The 293 cases were predicted as negative and it was false.

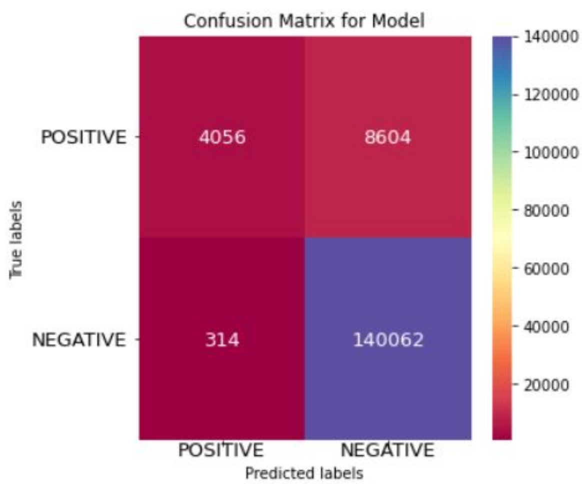

Figure 18: Confusion matrix for Random forest using VADER

(b) Evaluating Sentiment analysis techniques with SVM
- Both models of SVM with VADER and NRCLex have same results and accuracy of 95% as shown in figure 19
- The SVM models precision is 96.54% and recall is 95.20

```
Accuracy :  0.9520439635118534
Precision :  0.965491334920794
Recall :  0.9520439635118534
              precision    recall  f1-score   support

           0       0.83      0.53      0.65     12660
           1       0.96      0.99      0.97    140376

    accuracy                           0.95    153036
   macro avg       0.89      0.76      0.81    153036
weighted avg       0.95      0.95      0.95    153036
```

Figure 19: Classification report for SVM : Vader and NRCLex

- As shown in Figure 20 The confusion matrix , the 6741 cases were predicted and are positive, whereas 5919 were predicted positive but were false. The 138956 case were predicted negative and ts true. The 1420 cases were predicted as negative and it was false.

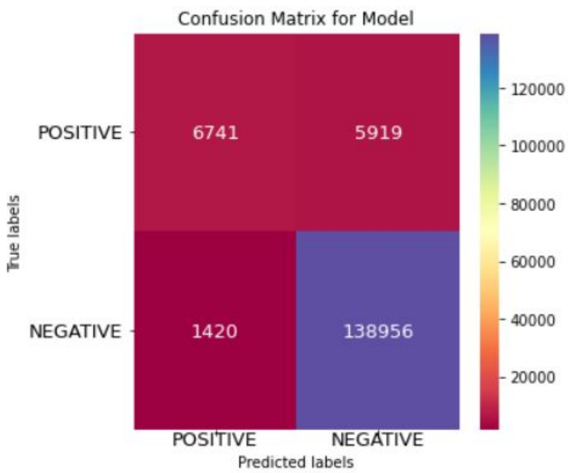

Figure 20: Confusion matrix for SVM : Vader and NRCLex

5.2 Discussion

To answer the research question :- “Can Sentiment Analysis using machine learning and deep learning provide insights, which can be used to improve the E-commerce sector and products?” It was found from evaluations on ratings and reviews sentiments that ratings of a products cannot be confirm if product is good also the sentiment analysis on reviews can provide a clear picture of product. The Sentiment analysis techniques - VADER using Sentiment Intensity Analyzer and NRCLex with machine learning and Deep Learning has provided impressive results. The NRCLex does provides 10 sentimental values for text body where as the VADER does only provide positive and negative. As comparing the models, Random forest model with VADER sentiment analysis have performed well with an accuracy of 94$. The Deep learning model the model has performed moderately due to slight over-fitting and is a good model on basis of graphs. The SVM model has performed well in all cases yet was the 2nd best of them with accuracy of 87%.

6 Conclusion and Future Work

The study is done with a goal to determine the emotions from a reviews of products sold online, and to uncover the hidden insights from the data. The research can help the seller and brands to better understand how their products and services are affecting their clients. To develop a model that can predict sentiments from a textual data, Vader Sentiment Scoring to know how much the sentence is positive, negative or neutral and NRCLex to generate number of emotions from the text were used. The Machine learning algorithm combined with deep learning that is the combination of Random Forest with Vader Sentiment Scoring showed prominent results with the model having high accuracy for predicting the sentiments. The GloVE has been a good parameter in Embedding layer as weights, which has caused model to have constant accuracy and loss.

For future work, merging of Amazons review sentiments with other sectors of business for comparison can be considered. Also other machine learning and deep learning model can be applied in comparison to these model. The product reviews of sectors can be merged with brands reviews of products. Using other sentiment analysis technique or pre trained models would be an interesting area of research.

Acknowledgment

My sincere thanks to the research supervisor Dr.Abid Yaqoob, all the faculty of Research In Computing (National College of Ireland) and to my family and friends who supported me to understand the project and its documentation in the best ways possible.

References

Cheng, L.-C. & Tsai, S.-L. (2019), Deep learning for automated sentiment analysis of social media, *in* '2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)', pp. 1001–1004.

Fan, X., Li, X., Du, F., Li, X. & Wei, M. (2016), Apply word vectors for sentiment analysis of app reviews, *in* '2016 3rd International Conference on Systems and Informatics (ICSAI)', pp. 1062–1066.

Gaikar, P. (n.d.), 'Improving the performance of aspect based sentiment analysis using transformer based techniques msc research project data analytics'.

Goel, A. K. & Batra, K. (2020), A deep learning classification approach for short messages sentiment analysis, *in* '2020 International Conference on System, Computation, Automation and Networking (ICSCAN)', pp. 1–3.

Hassan, R. & Islam, M. R. (2021), Impact of sentiment analysis in fake online review detection, *in* '2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD)', pp. 21–24.

Jain, G., Verma, S., Gupta, H., Jindal, S., Rawat, M. & Kumar, K. (2022), Machine learning algorithm based emotion detection system, *in* '2022 Fifth International Conference on Computational Intelligence and Communication Technologies (CCICT)', pp. 270–274.

Jianqiang, Z., Xiaolin, G. & Xuejun, Z. (2018), 'Deep convolution neural networks for twitter sentiment analysis', *IEEE Access* **6**, 23253–23260.

Kothalawala, M. & Thelijjagoda, S. (2020), 'Aspect-based sentiment analysis on hair care product reviews', *Proceedings - International Research Conference on Smart Computing and Systems Engineering, SCSE 2020* pp. 228–233.

Mukwazvure, A. & Supreethi, K. (2015), A hybrid approach to sentiment analysis of news comments, *in* '2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions)', pp. 1–6.

Nath Saha, B. & Senapati, A. (2020), Long short term memory (lstm) based deep learning for sentiment analysis of english and spanish data, *in* '2020 International Conference on Computational Performance Evaluation (ComPE)', pp. 442–446.

Srivats Athindran, N., Manikandaraj, S. & Kamaleshwar, R. (2018), Comparative analysis of customer sentiments on competing brands using hybrid model approach, *in* '2018 3rd International Conference on Inventive Computation Technologies (ICICT)', pp. 348–353.

Tsai, C.-F. (2010), An evaluation methodology for binary pattern classification systems, *in* '2010 IEEE International Conference on Industrial Engineering and Engineering Management', pp. 953–956.

Vanaja, S. & Belwal, M. (2018), Aspect-level sentiment analysis on e-commerce data, *in* '2018 International Conference on Inventive Research in Computing Applications (ICIRCA)', pp. 1275–1279.

Vimali, J. & Murugan, S. (2021), A text based sentiment analysis model using bi-directional lstm networks, *in* '2021 6th International Conference on Communication and Electronics Systems (ICCES)', pp. 1652–1658.

Wang, M. & Yu, Y. (2022), Deep sentiment analysis of the feelings expressed by tourists based on bert model, *in* '2022 International Conference on Culture-Oriented Science and Technology (CoST)', pp. 130–133.

Wedjdane, N., Khaled, R. & Okba, K. (2021), Better decision making with sentiment analysis of amazon reviews, *in* '2021 International Conference on Information Systems and Advanced Technologies (ICISAT)', pp. 1–7.

Yang, J. & Yang, J. (2020), 'Aspect based sentiment analysis with self-attention and gated convolutional networks', *Proceedings of the IEEE International Conference on Software Engineering and Service Sciences, ICSESS* **2020-October**, 146–149.

Zhang, X. & Yu, Q. (2017), Hotel reviews sentiment analysis based on word vector clustering, *in* '2017 2nd IEEE International Conference on Computational Intelligence and Applications (ICCIA)', pp. 260–264.

Zikang, H., Yong, Y., Guofeng, Y. & Xinyu, Z. (2020), Sentiment analysis of agricultural product ecommerce review data based on deep learning, *in* '2020 International Conference on Internet of Things and Intelligent Applications (ITIA)', pp. 1–7.