Genuine Online Reviewer Identification using ML Techniques

MSc Research Project
Data Analytics

Gayathri Chengalakkattu Ajayan
Student ID: X21157103

School of Computing
National College of Ireland

Supervisor: Prof. Jorge Basilio

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Gayathri Chengalakkattu Ajayan
Student ID:	X21157103
Programme:	Data Analytics
Year:	2022
Module:	MSc Research Project
Supervisor:	Prof. Jorge Basilio
Submission Due Date:	01/02/2023
Project Title:	Genuine Online Reviewer Identification using ML Techniques
Word Count:	2417
Page Count:	23

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:	
Date:	31st 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):	

Genuine Online Reviewer Identification using ML Techniques

Gayathri Chengalakkattu Ajayan
X21157103

Abstract

Customers are constantly trusting digital reviews as a form of insight before taking any purchase choice. Authentic item evaluations generally get a big consequence upon future product orders. Under this method, we put forth the effort in analysing costumers' feedback that has been posted in effort to comprehend the aspects of their actions. In contrast side, finding bogus comments and filtering these from the dataset employing multiple Natural Language Processing (NLP) procedures becomes essential in a wide range of contexts. The hybrid Convolution Neural Network-Long Short Term Memory (CNN-LSTM) Machine Learning (ML) system is applied to review database to train it to forecast how a comment is positive or negative utilizing the sentiment analysis (SA) approach. On the basis of the Crowd Sourcing concept's consolidated result, the Genuine Reviewer is determined. Prior to actual purchase, the customer reliance on item ratings in the e-commerce sector and possibly on different forums is growing. Therefore, the situation of false reviews should be handled so those multiple firm like eBay, Amazon etc could resolve it while also getting rid of the fraudsters and bogus critics, preserving customers' faith in websites and other internet portals. To evaluate the chosen model, accuracy, recall, precision and F1 Score are employed. The accuracy for the Baseline model is 97 percent.

Keywords: Natural Language Processing (NLP), Genuine Reviewer, sentimental analysis (SA), Crowd Sourcing, Machine Learning (ML).

1 Introduction

1.1 Background and Motivation

People in this digital age rely on technology for all of their day-to-day needs. Technology is continually developing. Techniques that are more contemporary and sophisticated are continually overtaking older ones. Such contemporary technology enables individuals to do their work swiftly. The public has become increasingly aware of e-commerce websites as a result of their accessibility and time-saving capabilities. Almost all of us examine buyer ratings prior buying any kind of a service or an item. Today, firms may benefit greatly from internet evaluations as a form of acknowledgment. They also contribute significantly to the publicizing and marketing of goods and services. As digital services gain traction, bogus web comments pose an increasing problem. People are capable of writing misleading comments to discredit authentic consumers and advertise their respective products. Often consumers are swayed and post reviews that are contrary to

the essence of the goods or item. Knowing if the critic is authentic or bogus is crucial because several of consumers count upon that. Hence, NLP methods have been applied to detect the comments rather than mistakenly identifying a phoney comment like sentimental analysis does. Organizational judgments depending on sentimental analysis are vital while selecting an item or merchandise.
The Hybrid Deep Learning (DL) network is used to detect both positive and negative terms in order to identify the Genuine Reviewer. The methodology relies on customer feedback and evaluation in order to determine if the phrase and evaluation are either favorable or detrimental. It processes a corpora as its intake, performing activities like phrase division, vectorization, special characters and stopwords elimination, and reducing stem of the words to root word during the preprocessing stage. The convolutional layers extracts top order properties, while the LSTM structure discovers long-term associations among terms. By studying the standing and history of its customers in digital sources, like reputable press sites, forums, even weblogs, several businesses already started to produce income channels. The difficulty is in correctly identifying the orientation. In perspective of sentimental analysis, the integration of ML with crowdsourcing provides a variety of benefits. Employing crowdsourcing and ML, the main goal is to generate objective sentimental analysis of a changing set of comments, emphasizing commonalities and differences among diverse sources and spotting patterns. This technique can establish the authenticity of a customer's feedback via sentimental analysis and the idea of crowdsourcing. Python calculation is employed to compute the public opinion of the item and to determine whether a rating is real. The DL approach is applied to categorize reviews either as pleasant or unfavorable.

1.2 Research Question

In today's internet-driven world, the majority of evaluators conduct item ratings online. Many clients are persuaded to write reviews that are at odds with the core of the product. The majority of consumers rely on this, therefore knowing if the critic is legitimate or not is essential.
How does the suggested system employ Deep Learning techniques to determine whether the review is positive or negative and how does Python computation work to determine the product's crowd sentiment before determining the reviewer's authenticity?

1.3 Research Objective

In addition to offering methods for identifying the Genuine Reviewer, the objectives described below address the research question posed in section 1.2.

1. Identifying good data source for sentiment analysis and fine-tune the dataset using NLP techniques and extract useful features to develop sentiment classification using deep learning.

2. The architecture of the system is created, that will aid in deploying the hybrid CNN-LSTM DL system and categorizing reviews as either positive or negative employing DL approaches.

3. Assessment of the model and design that will aid in determining the genuine reviewer utilising the user's crowdsourcing and product sentimental analysis concepts.

2 Related Work

This section offers a succinct summary of the publications that employed a hybrid DL system and sentiment analysis to decide whether or not evaluations were favorable. The proposed method and the dataset that researchers utilised are used to classify the study papers. The suggested system is primarily divided into three methods for recognising the Genuine Reviewer: sentiment analysis, crowdsourcing, and deep learning and also the NLP methods employed in these research papers, which are described below.

2.1 Survey and comments

Elmogy et al. (2021) recommends a system using Machine Learning techniques to spot fake comments. This work includes a gamut of featured design methods in addition to the extracting of features used on the assessments to record different customers' actions. The study evaluates the efficacy of a variety of studies conducted using a genuine Yelp data made up of reviews of restaurants that include and exclude factors determined by customer choices. The study evaluated the effectiveness of many models in both scenarios, including Naive Bayes (NB), RF, KNN, LR and SVM. Within evaluations, a variety of n-gram linguistic patterns, particularly bi and tri-grams was also taken into account. The outcomes show as KNN works better than various models in terms of f-measure with the maximum 82.4 percentage f-measure.

Fornaciari et al. (2020) contrasts a collection of actual phoney internet comments along a set of crowd-source (phoney) opinions. They assess how close they are to one another as well as its value in developing methods to identify unreliable comments. This study discovered how the bogus comments gathered through community sourcing vary considerably out from fraudulent ones posted on web. It appears as these deliberately generated misleading sentences' subject closeness to the intended audiences has a considerably greater impact on the classifiers' efficacy over its deception. This implies why methods developed for recognizing comment fraud via crowd-sourcing sets of data might not be suitable to the actual job of identifying false ratings. It suggested approaches centered on the statistical tagging of unmarked phrases, depending on the utilization of contextual-info often accessible on e-commerce portals, as a substitute way to produce massive data for the false comment recognition job. Regardless of the content of the comment, these strategies allow for the training of reliable algorithms for the detection of fraudulent comments.

To safeguard consumers, goods, as well as shopping websites interests, Jacob et al. (2020) created a method towards recognising and deleting bogus comments. Therefore, the primary objective of this project is to discover unjust Ratings on amazon utilising sentimental analysis and supervised ML methods in online shopping company setting. On a corpus of customer ratings for smartphones from Amazon, sentimental analysis classifiers are applied. In order to discover unjustified or acceptable favourable and unfavourable ratings, this research utilized 3 distinct methods-logical and linear regression, and also CNN and RNN of the supervised ML methods. These methods entail vetting, cooperative filtration, and expelling with an ideal accuracy yield. Exploring a DL method which provides maximum efficiency in the spotting of false comments is the main objective or focal point.

Punde et al. (2019) Several rating platforms contain excellent auditing created by the goods association personnel themselves in an effort to produce assessments that aren't accurate. This study developed a method which identifies fraudulent item ratings and removes them all employing ML in order to get rid of such kind of false item rating. It eliminated comments which an advertising firm has flooded online in an effort to raise an item's ratings. Lastly, a sentimental analysis is performed, whereby each term is given an emotion tag using the Bag-of-words method. Aspects of info extraction as well as text processing are included in sentimental analysis. A significant obstacle in sentimental analysis is feature weighing, that is essential for accurate categorization. The research may be expanded to new fields including e-learning and analysis using soft computation methods like NNs.

Tufail et al. (2022) The detection of fake or false comments has become a contentious issue due to the significance of item evaluations in guiding purchasers in making judgments. Considering the influence such ratings exert on a company, the perilous actions of writing fraudulent ratings for individual benefit had gotten worsened throughout. In this study, they developed a methodology for detecting bogus reviews utilizing textual categorization and ML approaches. In attempt to recognize phoney comments focusing on the quantity of pronouns, adjectives, and emotions, they employed predictors like SVM, KNN and LR. Upon this Yelp collection as well as the TripAdvisor collection, the suggested approach for spotting fraudulent internet ratings exceeds existing cutting-edge methods by 95 and 89.03 percent correctness, respectively.

In effort to effectively estimate how authentic the comments in a provided records were, two distinct ML methods were applied to train the false comment database in this paper by Anas and Kumari (2021). While relying on customer ratings for the goods discovered using internet on various sites and apps, the frequency of fraudulent evaluations in the e-commerce sector and perhaps other sectors is rising. Prior buying an item, customers trust the firm's offerings. In order for the big e-commerce organizations like eBay, Amazon, etc. to manage and get rid of spamming and fake evaluators, maintaining customers' faith in online shopping platforms, the issue of fake comments should be tackled. This technique may be applied to apps and sites with only a small number of customers, predicting the validity of reviews so that site admins can respond to them appropriately. The approaches of Naive Bayes and Random Forest are used to create this system. One might rapidly determine the quantity of bogus ratings on an app or a site by employing these methods. A robust algorithm which was constructed using comments in millions was needed in combating such scammers. The "amazon yelp dataset," a relatively small amount of information that may be built up to attain exceptional efficiency and personalization, is used to train the algorithms in this study.

This study employ sentimental analysis, where textual mining is performed primarily on the feelings of the perspective, to discover reviews that are phoney affirmative or phoney adverse. In an effort to distinguish between phoney favourable as well as false negative film ratings utilising two distinct sets of data, Elmurngi and Gherbi (2017) contrasts five distinct Supervised ML classifiers. They come to the conclusion as to SVM algorithm performs quite accurately than the other 4 methods such as Decision Tree, K-Star, KNN, and NB.

2.2 NLP and other ML techniques for Pre-Processing

The sentimental analysis method which can clearly differentiate among good and bad ratings is presented by Hassan and Islam (2021). Using a collection of online ratings, the suggested sentiment's method is used to assess the influence of a potential emotion's scoring on the recognition of falsely web ratings. The probability emotion scoring offered by this approach helps the fraudulent comment recognition mechanism operate well. With 89 percent accuracy, the method's TF-IDF (Term Frequency and Inverse Document Frequency) and emotion factors can recognize fraudulent ratings. This work develops a method for recognizing bogus ratings employing only textual data.

Alrehili and Albalawi (2019) The sentimental analysis technique is applied on a collection of user ratings that were obtained via Amazon. Employing the ensemble ML approach, categorizing every comment as either a favourable one or a bad one is done. NB, SVMs, RF, Bagging and Boosting were the five predictors that were merged as part of the ensemble ML approach called Voting. This paper has tested six distinct hypotheses in order to compare the suggested framework to the five classifiers. The situations employ unigrams, bigrams and trigrams (with or without) stopwords removal. The outcome demonstrates that when employing unigrams (with) stopwords, the RF approach provides the maximum efficiency that is equivalent to 89.87 percent accuracy.

Chong et al. (2014) outline the initial research on sentiment analysis of tweets in this research. This experiment aims to identify sentiment based on topics seen in tweets. Using Natural Language Processing techniques, it finds the sentiment that pertains to the particular topic. Our experiment uses three key processes to categorize sentiment: subjectivity classification, semantic association, and polarity classification. By outlining the grammatical connection between sentiment lexicons and subject, the experiment makes use of sentiment lexicons. Experimental results show that the suggested method outperforms current text sentiment analysis algorithms because tweets don't have the same structure as traditional text.

2.3 Deep Learning Methods to Predict Positive or Negative Reviews

Shahariar et al. (2019) to recognize the bogus textual feedback, with both marked and unmarked input, DL techniques have been employed in this research. These techniques include CNNs, Multi-Layer-Perceptrons, and LSTM a form of Recurrent NNs (RNN). In effort to recognize fraudulent comments and compare the effectiveness of standard and DL predictors, this study have also deployed certain classic ML predictors. In attempt to get improved term vector renderings as well as boost the precision of predictors learned using conventional ML methods, DL algorithms like Word2Vec are also employed. After a specific amount of training set, conventional ML methods like SVM and NB will no longer enhance efficiency. The limitation of this DL approach is the limited data that causes an overfitting issue.

Tammina and Annareddy (2020) employed the deep convolutional neural network, a deep learning technique, to collect sentiment from IMDB and Amazon product reviews. The authors of the study compare the deep learning model to the traditional machine learning algorithms Random Forest and Naive Bayes. Convolutional neural networks scored 74 percent on the Amazon Product Review dataset and 68 percent on IMDB, proving to have the highest accuracy, according to the results.

Rafay et al. (2020) have suggested a sentimental analysis model to conduct the categorization on industry evaluations with an emphasis on ratings for all types of restaurants. Utilizing a sizable and comprehensive textual review data given by Yelp, basic binary and multilabel classifiers are employed to get reliable outcomes. Comparing the outcomes of the DL methods such as convolution-LSTM and multinomial NB along word2vec and global vector (Glove) is carried out in this research. The optimal system for categorising the evaluation scores is C-LSTM, it has been discovered by examining the efficacy of every system using various measures. Using a word to vector encode strategy, the finest predictor C-LSTM had achieved generally favourable results in both binary and multi-class categorization.

In the study by Yang and Yang (2020), In order to create a new model that handles Aspect Based Sentiment Analysis (ABSA), convolutional neural networks and a self-attention mechanism were merged.Both aspect-category sentiment analysis (ACSA) and aspect-term sentiment analysis (ATSA), two types of aspect-based sentiment analysis, are the focus of investigations. The experiments employed the SemEval 2016 and SemEval 2014 datasets, respectively. The proposed model incorporates a softmax layer for the ACSA tasks, a full connection layer, three CNN filters, a gating mechanism, and a self-attention algorithm. In the convolutional neural network area of the suggested ATSA model, there is an additional CNN filter. The dot-product algorithm is the main source of inspiration for the self-attention algorithm, whose primary goal is to enhance learning of the correlation between the words in the text. The authors presented respectable findings using accuracy as their main evaluation criteria.

Three polarity categories—Positive, Negative, and Neutral—were used in the article by Monika et al. (2019) to categorize tweets regarding six US airlines. Recurrent neural network and long short-term memory models were used to accomplish this. Twitter is the dataset's primary source. The investigations consist of two stages; in the first, GloVe vector word embeddings are used to pre-process the tweets. This is followed by the model's training stage. The model design consists of the input layer, two LSTM layers, dropout layers on each of them, a fully connected Dense Layer, and an output layer with a Softmax activation function. The authors present encouraging findings that enable them to draw the conclusion that the suggested strategy is a trustworthy choice for forecasting the polarity of tweets about US airlines. However, the authors suggest that additional research investigations on the Bidirectional LSTM(Bi-LSTM) Model be conducted in order to increase performance.

Sarkar (2019) looked into using a Long Short-Term Recurrent Neural Network to determine the polarity of sentiment in Bengali tweets. Due to the availability of the Indian language on social media and the fact that most research on sentiment polarity recognition has been conducted in English, the study focuses exclusively on tweets in Bengali.

An integrated CNN and LSTM-RNN based like Deep NN Opinion mining was created by Kaladevi and Thyagarajah (2021). The vast amount of online networking info is subjected to sentimental analysis using the ICNN-LSTM-DNN approach. Regarding sentimental analysis, the comments that were made on Twitter throughout the 2019 Indian Elections are used as the source of information. The suggested technique mechanically differentiates user's factual information from opinions in live twitter posts. In grounds of F-measure, accuracy, recall, and precision, this strategy is better than the currently used strategies.

3 Methodology

The proposed study employs the CRISP-DM approach.
In the modern world, many things are bought from e-commerce websites, and before making a selection, consumers often read online reviews and customer ratings of a given product. As a result, the orders for and sales of the products are greatly influenced by these reviews and ratings. Therefore, it is crucial to analyze these online evaluations and determine whether they are real or fraudulent. As a result, by effectively identifying and removing bogus reviews, the sale of the listed goods can be increased by using e-commerce platforms like Amazon, eBay, and others. Additionally, it contributes to client loyalty.
The suggested approach generates a CNN-LSTM Hybrid DL Sentiment Analysis model using the collected data for identifying the sentiment associated to the reviews of each customer. The elements required to recognize the Genuine Reviewer are listed in figure 1.
The proposed methodology for identifying Genuine Reviewer can be broadly classified into 4 different stages such as:

1. Sentiment Analysis Model
2. Product Crowd Sentiment Analysis
3. User Crowd Sourcing Behavior
4. Genuine Reviewer Rating

1. **Sentiment Analysis Model**
   Identification of the emotion or mood associated with the certain situation is done using the sentimental analysis technique. It is mainly related to investigating into and determining the sentiment or intent behind any sort of communication, whether it be spoken, written, or otherwise.

   Read the Text Dataset:
   Algorithms, a ton of data, and other various things are required for machine learning to work successfully. In order to correctly develop the training algorithm, the information gathered should include the necessary qualities. It is essential to make sure there are enough entries because more information is generally helpful.

   A count plot of all the ratings is plotted as shown in figure 2. The dataset contains five different categories of ratings, and figure 2 illustrates the Rating Countplot as a bar chart. This diagram demonstrates how the distribution of the ratings is negatively skewed. The five Rating classes are displayed on the x-axis, and the count for each Rating class is displayed on the y-axis.

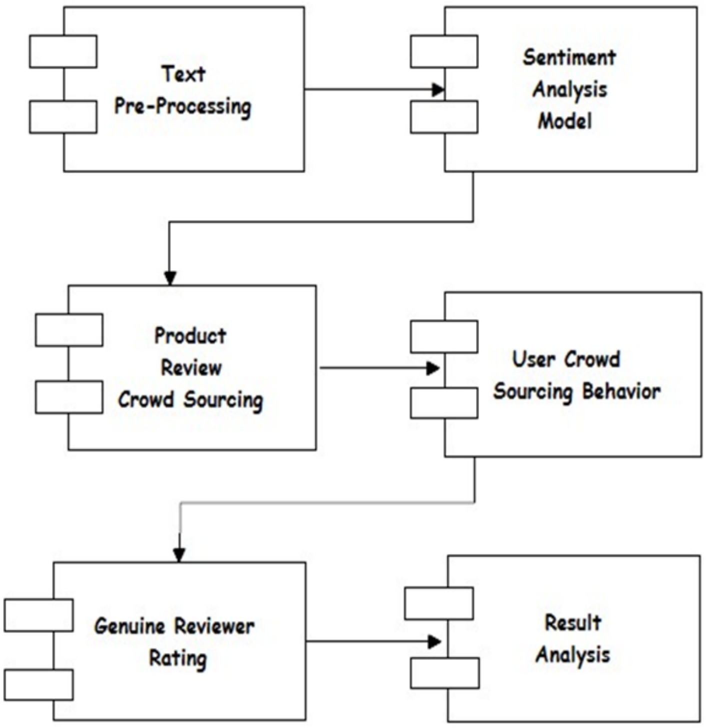


Figure 1: Proposed Methodology for identifying Genuine Reviewer

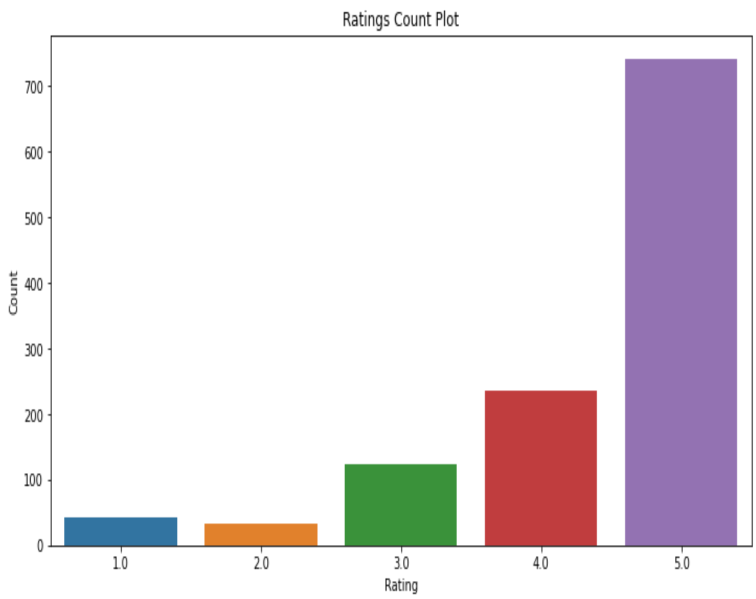


Figure 2: Ratings Count Plot

A pie chart for the different brands is plotted as shown in the figure 3. It displays a pie chart of the brands included in the dataset; according to the legend, Amazon has the most devices, while Moshi has the fewest.

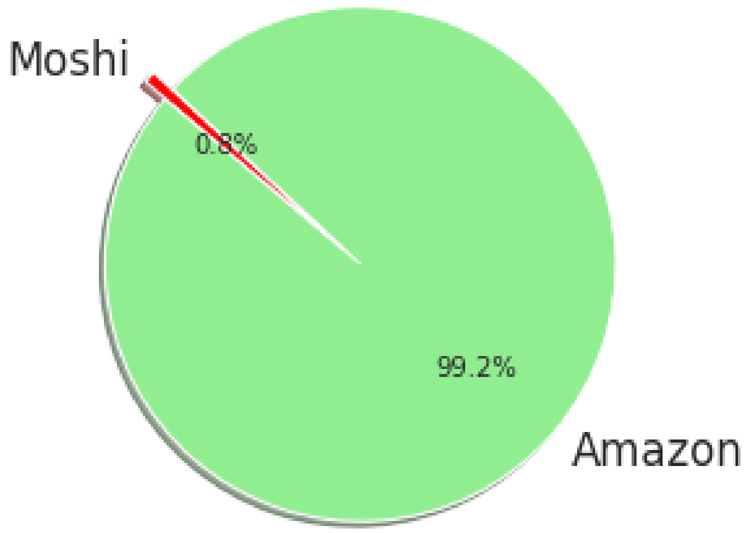


Figure 3: Brands Pie Chart

Preprocess Dataset using NLP Techniques:
Data preprocessing is a method for transforming impure inputs into clean sets of data. In other words, gathering data from many sources results in unstructured data collection, which hinders study. Data preparation is necessary since actual data is sometimes uneditable. Accuracy, consistency, and the presence of noisy data make up the majority of the real data. Reviews that are null are removed since they are useless and cannot be translated into sentiment labels. To ensure that the data doesn't contain any duplicate reviews, checks are done.

Figures 4 and 5 show the target variable before and after addressing the class inbalance, respectively. Figure 4 displays the class inbalance that exists in the target variable during the Exploratory Data Analysis step.

The most frequently used words in the reviews are then compiled into a Word Cloud, as seen in figure 6.
A word cloud is a group or collection of words that are shown in various sizes. The more pronounced and bold a word is, the more frequently an audience member chooses it or votes for it. Word clouds are an effective tool for visualizing the opinions of the audience on a certain subject. They are quick to manufacture, simple to understand, and easy to read.
Words like Amazon, Kindle, Tablet, Fire, Alexa, etc. are among them.

Using Vectorization Technique to create Bag-of-Words:
This is referred to as a "bag" of phrases since no attention is paid to the structure or organization of the phrases inside the text. The algorithm doesn't care where in

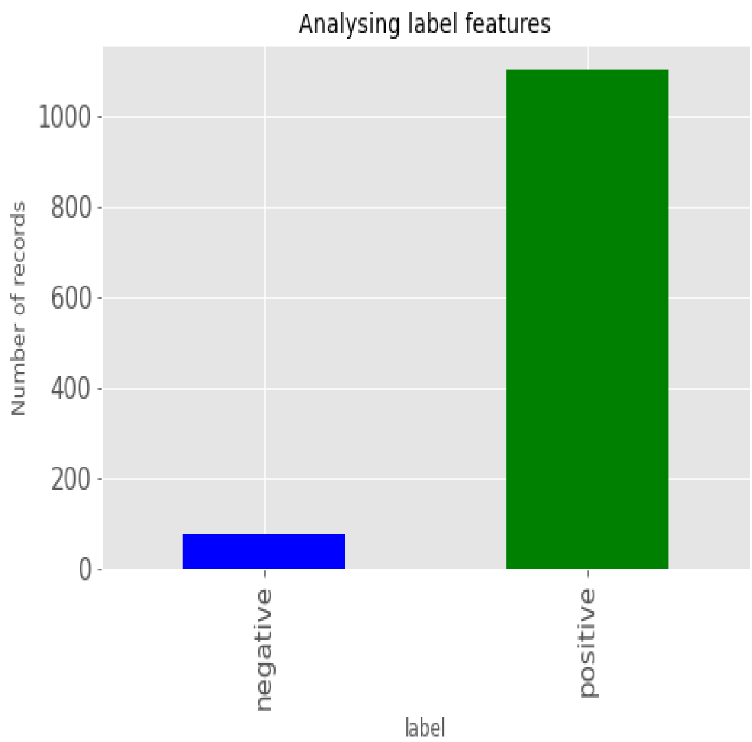


Figure 4: Target Variable: Class Inbalance

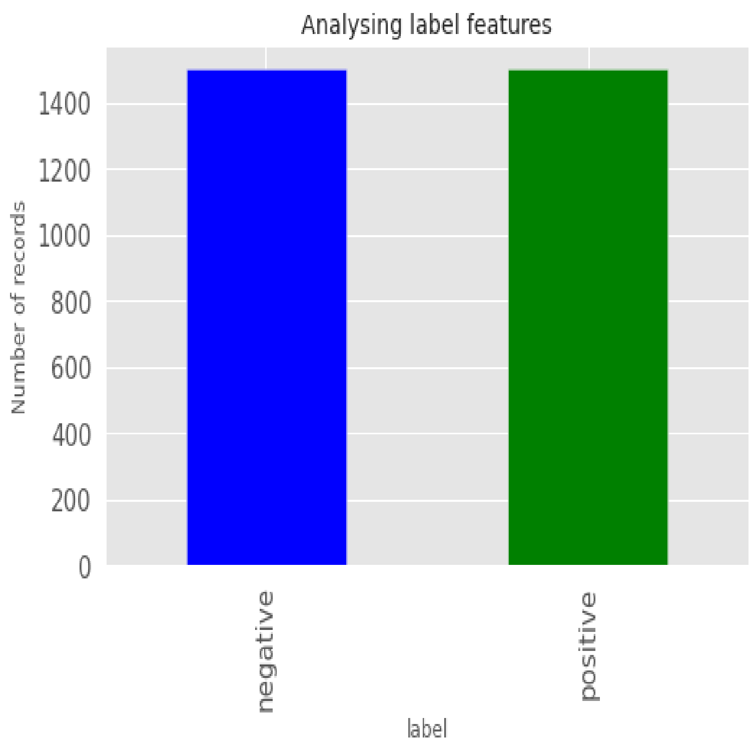


Figure 5: Target Variable: UpSampled

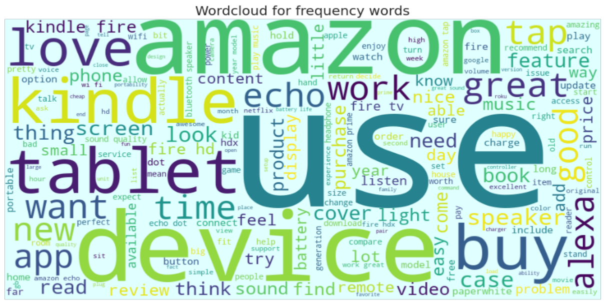


Figure 6: Review Word Cloud

the article a term appears; it is only interested in if it is present.

Split the Bag-of-Words into Training Data and Test Data:
By dividing the training and testing datasets before whatever data preparation, we can keep in mind that such a constraint is taken into account when our models are evaluated. This ensures that throughout the process of dataset processing and model training, we are unable to leverage any information from the test set which would enhance the way we organize the data. The data is divided as 70 percentage training and 30 percentage testing. Whih means that the final 30 percentage of the reviews each of which is positive and negative will be utilized as the test set and remaining 70 percentage reviews will be utilized as the training dataset.

The Sentiment Analysis System is built using the Training Data with the CNN-LSTM and the Sentiment Model is saved.

2. **Product Crowd Sentiment Analysis**
By sending the reviews to the sentiment model, the sentiments of the reviews are discovered. The sentiment model is then used to calculate the total number of reviews for each and every product, both positive and negative thus finding the product's crowd sentiment.

3. **User Crowd Sourcing Behavior**

Step 1: Count the positive and negative reviews for the product to determine the crowd sourcing sentiment.
Step 2: Calculate the proportion of each user's reviews that are in accordance with the general consensus and the proportion that are not.
Step 3: Provide users with marks based on the total count of genuine review counts and non genuine review counts.

4. **Genuine Reviewer Rating**

The Top K Genuine Reviewers are listed according to the scores in descending order of their scores. Each genuine review will earn the user +2 points, and this information will be kept in the variable "GenuineReviewCounts." Each user will then receive a score of -2 for each review that isn't genuine, and that score will be saved in the variable "NotGenuineReviewCounts." The overall genuine score for each user will be determined by the difference between the two.

3.1 Data Understanding

We'll use the publicly accessible Amazon product reviews dataset from the website www.kaggle.com [1] for this project. The collection includes excerpts from Amazon reviews for various commodities and products. The dataset contains 1,597 customer reviews that were chosen at random from a range of customers. Amazon tries to identify the primary topics of these ratings in order to categorize them for simpler searching. A user review of a product on Amazon is included with each insight. The customer reviews combine the overall rating and the text comment. Understanding the dataset, doing the necessary cleaning, and developing a solid subject modeling strategy are the objectives. There are 27 different attributes in the dataset. Figure 7 displays the dataset.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	
id	asins	brand	categories	colors	dateAddec	dateUpdat	dimension	ean	keys	manufactu	manufactu	name	prices	reviews.da	reviews.dc	reviews.nu	reviews.ra	reviews.sc	reviews.te	reviews.tit	reviews.us	reviews.us	re
AVpe7AsM	B00QJDU3	Amazon	Amazon Devices,maz		2016-03-0	2017-07-1	169 mm x 117 mm x 9		kindlepape	Amazon		Kindle Pap	[{"amount	2015-08-08T00:00:00		139	5	https://wv	I initially h	Paperwhite voyage, no regrets!			Cr
AVpe7AsM	B00QJDU3	Amazon	Amazon Devices,maz		2016-03-0	2017-07-1	169 mm x 117 mm x 9		kindlepape	Amazon		Kindle Pap	[{"amount	2015-09-01T00:00:00		126	5	https://wv	Allow me t	One Simply Could Not Ask For M			Ri
AVpe7AsM	B00QJDU3	Amazon	Amazon Devices,maz		2016-03-0	2017-07-1	169 mm x 117 mm x 9		kindlepape	Amazon		Kindle Pap	[{"amount	2015-07-20T00:00:00		69	4	https://wv	I am enjoy	Great for those that just want ar			Te
AVpe7AsM	B00QJDU3	Amazon	Amazon Devices,maz		2016-03-0	2017-07-1	169 mm x 117 mm x 9		kindlepape	Amazon		Kindle Pap	[{"amount	2017-06-16T00:00:00		2	5	https://wv	I bought o	Love / Hate relationship			Do
AVpe7AsM	B00QJDU3	Amazon	Amazon Devices,maz		2016-03-0	2017-07-1	169 mm x 117 mm x 9		kindlepape	Amazon		Kindle Pap	[{"amount	2016-08-11T00:00:00		17	5	https://wv	I have to s	I LOVE IT			Mi
AVpe7AsM	B00QJDU3	Amazon	Amazon Devices,maz		2016-03-0	2017-07-1	169 mm x 117 mm x 9		kindlepape	Amazon		Kindle Pap	[{"amount	2015-07-08T00:00:00Z				http://ww	My previo	Great device for reading. 8 peop			Ke
AVpe7AsM	B00QJDU3	Amazon	Amazon Devices,maz		2016-03-0	2017-07-1	169 mm x 117 mm x 9		kindlepape	Amazon		Kindle Pap	[{"amount	2015-09-01T00:00:00Z				http://ww	Allow me t	One Simply Could Not Ask For M			Ri
AVpe7AsM	B00QJDU3	Amazon	Amazon Devices,maz		2016-03-0	2017-07-1	169 mm x 117 mm x 9		kindlepape	Amazon		Kindle Pap	[{"amount	2015-07-03T00:00:00Z				http://ww	Just got mi	Definitely better than the previo			Ba
AVpe7AsM	B00QJDU3	Amazon	Amazon Devices,maz		2016-03-0	2017-07-1	169 mm x 117 mm x 9		kindlepape	Amazon		Kindle Pap	[{"amount	2015-08-08T00:00:00Z				http://ww	I initially h	Paperwhite voyage, no regrets! 1			Cr
AVpe7AsM	B00QJDU3	Amazon	Amazon Devices,maz		2016-03-0	2017-07-1	169 mm x 117 mm x 9		kindlepape	Amazon		Kindle Pap	[{"amount	2015-07-20T00:00:00Z				http://ww	I am enjoy	Great for those that just want ar			Te
AVpe7AsM	B00QJDU3	Amazon	Amazon Devices,maz		2016-03-0	2017-07-1	169 mm x 117 mm x 9		kindlepape	Amazon		Kindle Pap	[{"amount	2015-07-01T00:00:00Z				http://ww	As reviewe	Give this to a book-loving-new-n			Mi
AVpe7AsM	B00QJDU3	Amazon	Amazon Devices,maz		2016-03-0	2017-07-1	169 mm x 117 mm x 9		kindlepape	Amazon		Kindle Pap	[{"amount	2015-08-05T00:00:00Z				http://ww	My new Ki	Trouble-free international shippi			Ma
AVpe7AsM	B00QJDU3	Amazon	Amazon Devices,maz		2016-03-0	2017-07-1	169 mm x 117 mm x 9		kindlepape	Amazon		Kindle Pap	[{"amount	2015-07-20T00:00:00Z				http://ww	I am enjoy	Great for those that just want ar			Te
AVpe7AsM	B00QJDU3	Amazon	Amazon Devices,maz		2016-03-0	2017-07-1	169 mm x 117 mm x 9		kindlepape	Amazon		Kindle Pap	[{"amountMax":139.99,"amountMin":139.9				4	http://ww	Had older	Liked the smaller size			Ja
AVpe7AsM	B00QJDU3	Amazon	Amazon Devices,maz		2016-03-0	2017-07-1	169 mm x 117 mm x 9		kindlepape	Amazon		Kindle Pap	[{"amountMax":139.99,"amountMin":139.9				5	http://ww	This is a re	Superb reading device - but whic			Jo
AVpe7AsM	B00QJDU3	Amazon	Amazon Devices,maz		2016-03-0	2017-07-1	169 mm x 117 mm x 9		kindlepape	Amazon		Kindle Pap	[{"amountMax":139.99,"amountMin":139.9				5	http://ww	I love my k	I love it!			sa

Figure 7: Amazon Dataset

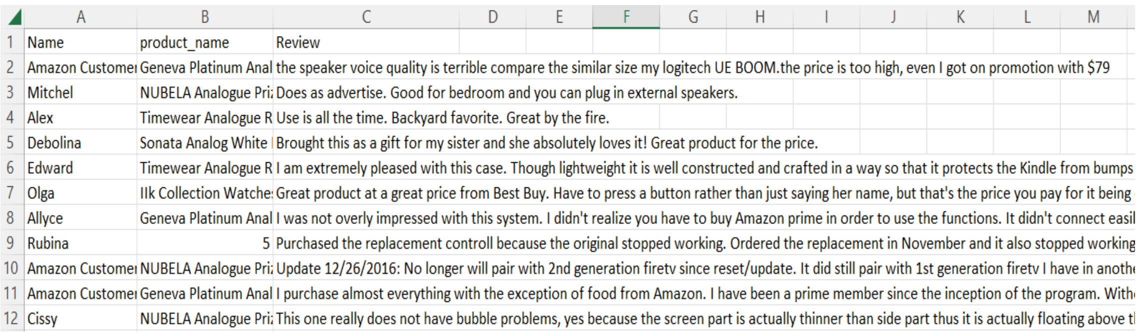
The dataset can be found here:
`https://www.kaggle.com/datasets/yasserh/amazon-product-reviews-dataset`

[1]https://www.kaggle.com/datasets/yasserh/amazon-product-reviews-dataset

Attributes List:

id	ID
asins	Product id asins
brand	Product Brand
categories	Product Category
colors	Product Color
dateAdded	Product Added Date
dateUpdated	Product Update Date
dimension	Dimension of the Product
ean	EAN Code of the Project
name	Name of the Product
keys	A certain product key assigned
prices	Price of the Product
manufacturer	Manufacturer of the Product
manufacturerNumber	Number of the Product Manufacturer
reviews.date	Date of the Product Reviews
reviews.doRecommend	Product Recommendation
reviews.rating	Ratings of the Product
reviews.numHelpful	Number of upvotes that are Helpful for reviews
reviews.sourceURLs	Review sources
reviews.text	Review of Product
reviews.userProvince	Product Review User Province
reviews.userCity	Product Review User City
reviews.title	Title of Review
reviews.username	Product Review User Username
sizes	Product Size
upc	Product UPC
weight	Product Weight

Figure 8 shows the second dataset which is used for the Genuine Reviewer Identification. This dataset has created by researcher herself with the help of Mockaroo: an online data generator tool. It has 3 attributes where Name indicates the customer name, product-name indicates the name of the product and Review indicates the customer reviews of the products.



	A	B	C
1	Name	product_name	Review
2	Amazon Custome	Geneva Platinum Anal	the speaker voice quality is terrible compare the similar size my logitech UE BOOM.the price is too high, even I got on promotion with $79
3	Mitchel	NUBELA Analogue Pri	Does as advertise. Good for bedroom and you can plug in external speakers.
4	Alex	Timewear Analogue R	Use is all the time. Backyard favorite. Great by the fire.
5	Debolina	Sonata Analog White	Brought this as a gift for my sister and she absolutely loves it! Great product for the price.
6	Edward	Timewear Analogue R	I am extremely pleased with this case. Though lightweight it is well constructed and crafted in a way so that it protects the Kindle from bumps
7	Olga	Ilk Collection Watche	Great product at a great price from Best Buy. Have to press a button rather than just saying her name, but that's the price you pay for it being
8	Allyce	Geneva Platinum Anal	I was not overly impressed with this system. I didn't realize you have to buy Amazon prime in order to use the functions. It didn't connect easil
9	Rubina	5	Purchased the replacement controll because the original stopped working. Ordered the replacement in November and it also stopped working
10	Amazon Custome	NUBELA Analogue Pri	Update 12/26/2016: No longer will pair with 2nd generation firetv since reset/update. It did still pair with 1st generation firetv I have in anoth
11	Amazon Custome	Geneva Platinum Anal	I purchase almost everything with the exception of food from Amazon. I have been a prime member since the inception of the program. With
12	Cissy	NUBELA Analogue Pri	This one really does not have bubble problems, yes because the screen part is actually thinner than side part thus it is actually floating above t

Figure 8: Second Dataset

3.2 Data Preparation

The information obtained from the set of data must be preprocessed before being used in the modeling of the review dataset. The data is preprocessed using a variety of NLP approaches, including the removal of stopwords, special characters, and punctuation, altering the case of the words, stemming, and lemmatization.
The simplest way to represent text in vectors is as a bag of words. In a vector, each column denotes a word. The values in a row's cells indicate how many times a word appears in a sentence. The texts are converted into sequences and the total unique tokens size is printed. Pad sequences is then used to ensure that all sequences in a list have the same length with maximum length set as 200 in this study.
The base dataset is then splitted into train, validation and test sets.
Then Glove word Embedding is used to group the terms together that are similar. GloVe is a method for obtaining word vector representations using unsupervised learning. GloVe, then, gives us the ability to naturally translate each word in a corpus of text into a location in a high-dimensional space. It is also known as Global Vectors for word representation.

4 Proposed Design Specification

The System Design Architecture for the Genuine Reviewer Identification is shown in Figure 9. The procedures involved in finding the genuine reviewer are shown in the suggested design. Following preprocessing as explained in Section 3, the proposed implementation designs are then applied as discussed in Section 5. Upon applying the NLP techniques and Product Crowd Sentiment models as discussed, the final step is to carry out the proposed model assessment steps as discussed in Section 6. By analyzing the model, we can get an idea of how well it performs in predicting the outcome as to genuinity of the review; after getting an evaluation result we can easily identify the genuine reviewer based on the product crowd sentiment analysis and user crowd sourcing behavior, this will help the product owners and companies efficiently to perform identification with less amount of time and with correct accuracy. This section identifies and presents the specifications related to the design that supports the implementation.

Firstly, the sentiment dataset whih is the Amazon dataset is loaded and various data processing is done using NLP techniques. Reviews that are null are removed since they are useless and cannot be translated into sentiment labels. To ensure that the data doesn't contain any duplicate reviews, checks are done. Variables are checked for class inbalance and treated to remove the inbalance. Vectorization Technique was used to create Bag-of-Words. Then the Bag-of-Words is splitted into Training Data and Test Data. The Sentiment Analysis System is built using the Training Data with the CNN-LSTM and the Sentiment Model is saved. Then the second dataset which contains user name, product name and reviews are loaded. All the same data processing techniques applied to the base dataset is applied to the second dataset also to avoid misclassification. The saved sentiment model is loaded and then it is used to calculate the total number of reviews for each and every product, both positive and negative thus finding the product's crowd sentiment. In the next step, the proportion of each user's reviews that are in accordance with the general consensus and the proportion that are not are calculated. Then users are provided with marks based on the total count of genuine review counts and non genuine review counts. For each genuine review, +2 marks will be given to the users and for each

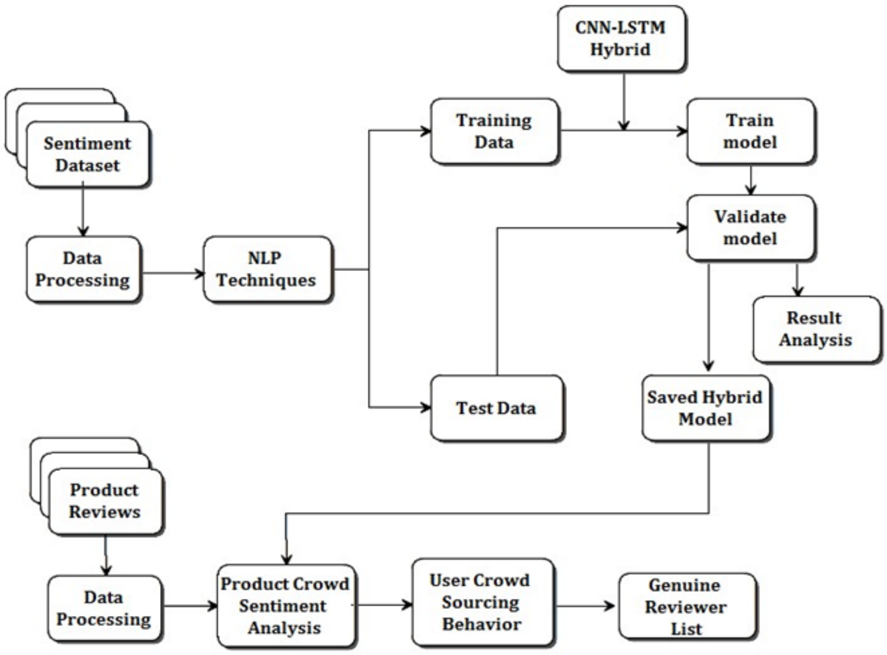


Figure 9: System Design Architecture for identifying the Genuine Reviewer

non genuine review, -2 marks will be given to the users. The difference of these two will give the Genuine Score for each user. And finally, the top K Genuine Reviewers are listed according to the descending order of their scores.

5 Implementation

In addition to providing a hybrid CNN-LSTM model, this study offers a practical and quick method for distinguishing between favorable and unfavorable phrases in the Review dataset. We can determine the reviewers who are being truthful by taking into account the overall point of view regarding a specific item. It is novel because we use the crowdsourced scoring system to decide if the user is in favor of or against this after deciding whether the rating is fair or unfair. A reviewer's reliability is taken for granted and their sincerity is considered when they concur with the bulk of the audience's ratings for the product. Figure 10 depicts the state diagram for the proposed system implementation.

5.1 CNN-LSTM Hybrid DL Model

The most common applications for CNNs are in identifying and intelligent perception tasks. Fully connected (FC), pooling (PL), and convolutional (CL) layers are their three main forms of layering. The topmost part of a CNN is called the CL. Even if there may be further CL or PL tiers, the FCL comes after them. With the aid of responsive connections, LSTM neural networks are RNNs that can take in long-term interactions.
The suggested model makes use of LSTM layers' capacity to identify both long and short term correlations as well as LSTM layers' ability to probe the internal structure of time-series data and extract essential information from it.

While LSTM models are good at predicting both short-term and long-term relationships, CLs are better at extracting important data and understanding the fundamental

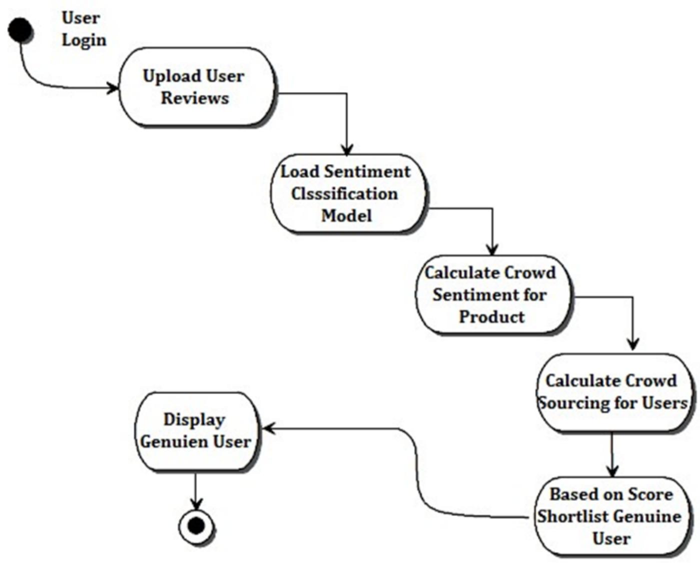


Figure 10: State Diagram for the Proposed System Implementation

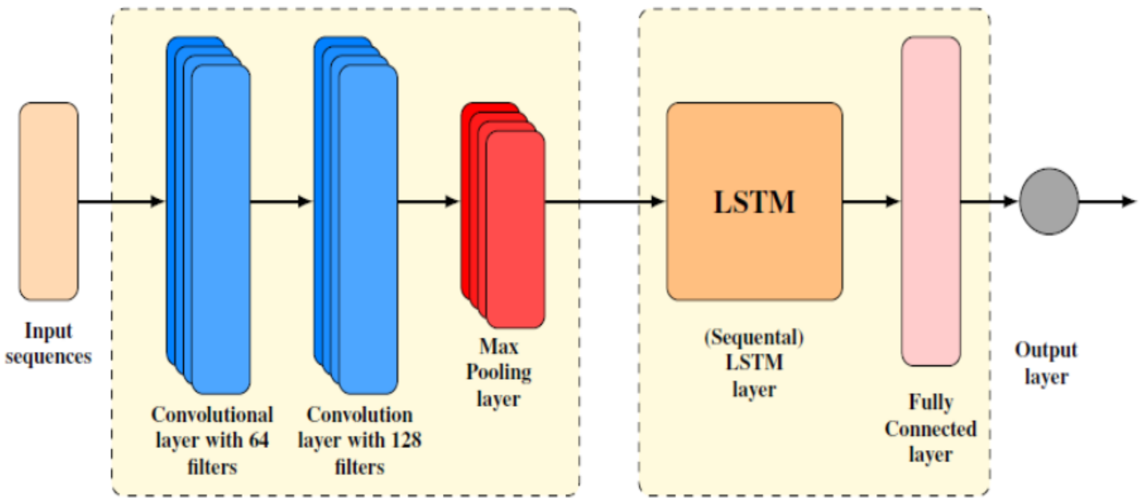


Figure 11: CNN-LSTM Architecture

workings of temporal data. The suggested approach's main goal is to effectively incorporate the benefits of such DL techniques. To achieve this, we describe the CNN-LSTM approach, which consists of 2 crucial elements. The intricate mathematical processes required to determine the properties of the dataset are performed in the first portion using CL and PL, while the features produced by LSTM utilizing packed layering are used in the second half. The CL, PL, and LSTM layering are the three primary parts of the proposed model, and the architecture of this layering is shown in Figure 11. [2]

CNN-LSTM Hybrid DL Model is constructed for text analysis employing a bidirectional long short-term memory (Bi-LSTM). The semantic and syntactic connections between text utterances are captured by the Bi-LSTM model. Thus, the model upholds the long-term dependencies in the text in this way. Two convolutional layers of 64 neurons each, a pooling layer, two Bi-LSTM layers of 400 neurons each, and a uniform kernel initializer are used to construct a sequential model. Then, a dense layer of 512 neurons is added. Function of activation These layers are activated using ReLu, while the output layer with its two output neurons is activated using the sigmoid activation function. The Bi-LSTM layer uses a 0.2 recurrent dropout rate. The figure 12 shows the summary of the implemented Bi-LSTM model.

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 embedding (Embedding)       (None, 200, 50)           173150

 conv1d (Conv1D)             (None, 198, 64)           9664

 conv1d_1 (Conv1D)           (None, 196, 64)           12352

 max_pooling1d (MaxPooling1D  (None, 98, 64)           0
 )

 bidirectional (Bidirectiona  (None, 98, 400)          424000
 l)

 bidirectional_1 (Bidirectio  (None, 400)              961600
 nal)

 flatten (Flatten)           (None, 400)               0

 dropout (Dropout)           (None, 400)               0

 dense (Dense)               (None, 512)               205312

 dense_1 (Dense)             (None, 2)                 1026

=================================================================
Total params: 1,787,104
Trainable params: 1,787,104
Non-trainable params: 0
```

Figure 12: Model Summary

5.2 Crowd Sourcing Concept

The main goal is to create objective sentimental analysis of a dynamic corpus of media posts using crowdsourcing and machine learning, showing trends and discrepancies across various sources.

[2]https://www.researchgate.net/figure/The-architecture-of-the-proposed-model-of-CNN-LSTM-with-two-convolution-layers-a$_{fig}$4$_3$47917347

The Product Review Crowdsourcing Algorithm will evaluate each item rating submitted by the client and determine whether or not it was favorable. Additionally, it will show which demographics make up the product's target market. A list of all the unique products are developed from the second dataset with the column "product name". Then a product dictionary is created with product name as the key and value is a list that contains the sentiment. Then the overall sentiment is calculated for the product using crowd sentiment analysis.
Nextly, User Crowd Sourcing Behavior is evaluated. The first thing that is discovered using the quantity of positive and negative reviews is the crowd-sourced sentiment of the product. The number of reviews for each user that are in line with the general consensus and those that are not are then calculated. Each genuine review will earn the user +2 points and -2 points for each review that isn't genuine. The overall genuine score for each user will be determined by the difference between the two.

6 Evaluation

The project's performance metrics are listed below:

1. **Classification Report:** The Machine Learning efficiency assessment measure is a classification report. It is employed to display the trained classified model's f1-score, support, recall and precision. The accuarcy of the model is 97 percentage.

	precision	recall	f1-score	support
positive	0.96	0.98	0.97	465
negative	0.98	0.96	0.97	435
accuracy			0.97	900
macro avg	0.97	0.97	0.97	900
weighted avg	0.97	0.97	0.97	900

Figure 13: Classification Report

2. **Confusion Matrix:** A table called a confusion matrix is created to describe how well a classifying system performs. The output of a classifier is shown and summarized in this matrix. It is very helpful for assessing recall, precision, accuracy and, perhaps crucially, the AUC-ROC curves.

   • **True Positive (TP):** Both the actual value and the forecasted value are true. Consequently, the value anticipated is positive and the forecast is accurate. In the evaluation results, we got 458 True Positives.

   • **True Negative (TN):** Both the actual value and the forecasted value are false. This indicates that the classification of the expected and actual values as negative is accurate. In the evaluation results, we got 416 True Negatives.

   • **False Positive (FP):** In this case, the model predicted a True value, but the

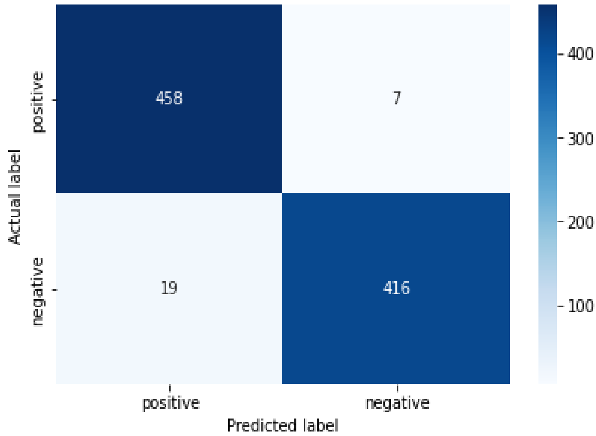


Figure 14: Confusion Matrix

actual value is False. This indicates that while the anticipated value was negative, the actual value was positive. This is defined as the Type I error. In the evaluation results, we got 7 False Positives.

- **False Negative (FN):** In this instance, the model's predicted value and actual value are both False, meaning that while the predicted value was positive, the actual value is classified as negative. This is referred to as a Type II error. In the evaluation results, we got 19 False Negatives.

3. **Recall:** The Recall value can be calculated by considering how many of the positive categories we properly predicted.

$$\text{Recall} = \text{TP}/(\text{TP} + \text{FN})$$

4. **Precision:** The Precision score can be calculated using the formula below by counting how many of the classes that were expected to be affirmative actually are.

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP})$$

5. **F-score:** Evaluating two models with high recall but low accuracy is challenging. As a result, we employ F-Score to analyze both.

$$\text{F} - \text{Measure} = 2 * (\text{Recall} * \text{Precision})/(\text{Recall} + \text{Precision})$$

6. **Accuracy:** The percentage of both positive and negative classifications that were accurately predicted. Maximum precision is required.

7. **Data Visualization:** It is the visualisation of data using standard images, including infographics, charts, plots, word clouds and pie charts. Using the Matplotlib library, such instructive graphical displays offer intricate relationships in dataset and data-driven findings simpler to grasp.

   The accuracy increases significantly in the first two epochs, as seen in figure 15, showing that the network is learning quickly. The curve flattens after the sixth epoch, indicating that fewer epochs are needed to train the model further. In general, overfitting occurs when the accuracy of the training data keeps increasing while the accuracy of the validation data decreases. Therefore, it is evident from figure 15 that there is no overfitting issue in this scenario.

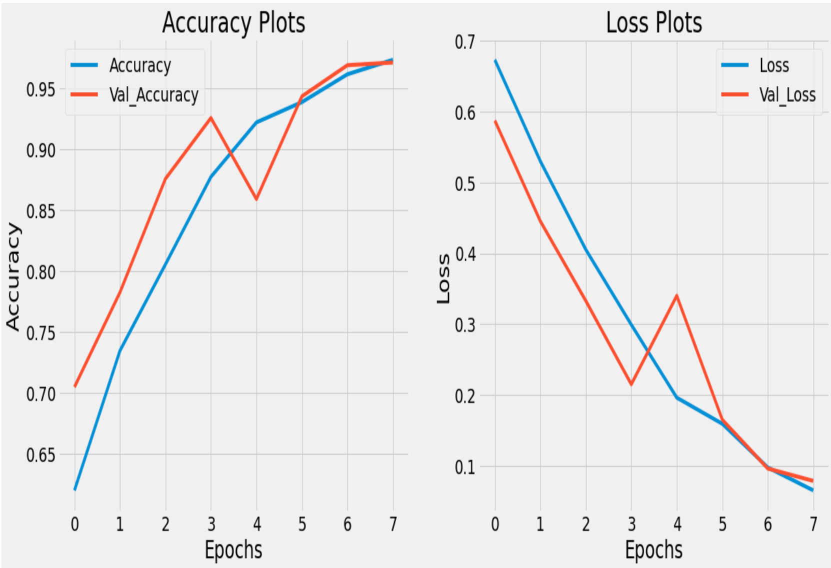


Figure 15: Accuracy graph and loss function graph for training and testing data

6.1 Experiments and Result

Apart from the classification report and confusion matrix which are used to analyse the model's efficiency, few more measures such as MSE(Mean Squared Error) and RMSE(Root Mean Squared Error) between the testing sets are also evaluated in this study.
A typical metric for assessing a model's predictive power is mean squared error.
In this research study:
MSE: 0.021
RMSE: 0.148
As both the values are less, smaller is the error and better the estimator.

Figure 16 shows the original output VS predicted output plot. It can be observed that the predicted output is similar to the original output.

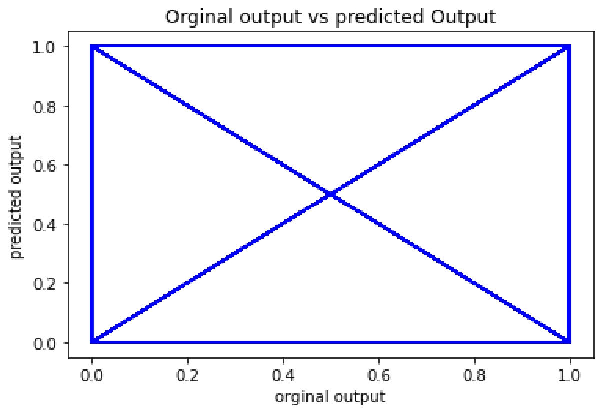


Figure 16: Original Output VS Predicted Output Plot

7 Discussion

The "Amazon Product Review dataset" efficiency results for the CNN-LSTM Hybrid DL model were studied in this work, and their significance for using the model in practical contexts was discussed. The classification report and the confusion matrix of the model used in the study are shown in Figures 13 and 14. The findings of the f1-score, accuracy rate and precision score show that the CNN-LSTM Hybrid DL model has surpassed all other available techniques.

The Amazon dataset utilized for sentiment analysis had a class imbalance in the target variables, which could make the baseline model overfit the training set. As a result, upsampling the target variables is used to address the class imbalance. As a result, the CNN-LSTM Hybrid DL model performs better overall.

The mean squared error between the testing sets were calculated and found to be 0.021. The computation time of the model was less than different alredy existing models. CNN-LSTM Hybrid DL Exhibited the recall core of 96 percentage which is comparitively gretater than that obtained in the study Monika et al. (2019) that signifies correct sentiment prediction associated with each review in the dataset.

Then the overall sentiment is calculated for the product using crowd sentiment analysis. By sending the reviews to the sentiment model, the sentiments of the reviews are discovered.The total number of reviews, both positive and negative for each and every product is then determined using the sentiment model.

As part of the identification of user crowdsourcing behavior, it is computed how many of each user's reviews are in line with the crowd and how many are not. Additionally, the users receive marks depending on user sentiment. The Top K Genuine Reviewers are successfully listed in descending order of scores at the end.

8 Conclusion and Future Work

The CNN-LSTM Hybrid Deep Learning model's efficiency results for the "Amazon Product Review dataset" are examined in this study, along with their importance for implementing the model to real-world applications. As an outcome, the CNN-LSTM model being a sentiment classification model provided results with a high level of accuracy. The primary objective was to employ a method to accomplish the process of identifying the genuine reviewer employing sentiment analysis and the conception of crowdsourcing. The genuine reviewer identification problem is thus successfully resolved and offered a reasonable knowledge of its validity and requirements.

In the future work, with the help of feature selection strategies, such as the Wrapper method, which randomly chooses a set of features, the classification model's accuracy can be increased.
The proposed approach is restricted to a collection of word processing techniques, which limits its applicability to e-commerce platforms. As a result, it might be expanded by being trained on a variety of datasets from different online sources.
Another interesting area to investigate is models that can handle different languages.

References

Alrehili, A. and Albalawi, K. (2019). Sentiment analysis of customer reviews using ensemble method, *2019 International Conference on Computer and Information Sciences (ICCIS)*, IEEE, pp. 1–6.

Anas, S. M. and Kumari, S. (2021). Opinion mining based fake product review monitoring and removal system, *2021 6th International Conference on Inventive Computation Technologies (ICICT)*, IEEE, pp. 985–988.

Chong, W. Y., Selvaretnam, B. and Soon, L.-K. (2014). Natural language processing for sentiment analysis: an exploratory analysis on tweets, *2014 4th international conference on artificial intelligence with applications in engineering and technology*, IEEE, pp. 212–217.

Elmogy, A. M., Tariq, U., Ammar, M. and Ibrahim, A. (2021). Fake reviews detection using supervised machine learning, *International Journal of Advanced Computer Science and Applications* **12**(1).

Elmurngi, E. and Gherbi, A. (2017). Detecting fake reviews through sentiment analysis using machine learning techniques, *IARIA/data analytics* pp. 65–72.

Fornaciari, T., Cagnina, L., Rosso, P. and Poesio, M. (2020). Fake opinion detection: how similar are crowdsourced datasets to real data?, *Language Resources and Evaluation* **54**(4): 1019–1058.

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

Jacob, M. S., Rajendran, S., Michael Mario, V., Sai, K. T. and Logesh, D. (2020). Fake product review detection and removal using opinion mining through machine learning, *International Conference on Artificial Intelligence, Smart Grid and Smart City Applications*, Springer, pp. 587–601.

Kaladevi, P. and Thyagarajah, K. (2021). Retracted article: Integrated cnn-and lstm-dnn-based sentiment analysis over big social data for opinion mining, *Behaviour & Information Technology* **40**(9): XI–XIX.

Monika, R., Deivalakshmi, S. and Janet, B. (2019). Sentiment analysis of us airlines tweets using lstm/rnn, *2019 IEEE 9th International Conference on Advanced Computing (IACC)*, IEEE, pp. 92–95.

Punde, A., Ramteke, S., Shinde, S. and Kolte, S. (2019). Fake product review monitoring & removal and sentiment analysis of genuine reviews, *International Journal of Engineering and Management Research (IJEMR)* **9**(2): 107–110.

Rafay, A., Suleman, M. and Alim, A. (2020). Robust review rating prediction model based on machine and deep learning: Yelp dataset, *2020 International Conference on Emerging Trends in Smart Technologies (ICETST)*, IEEE, pp. 8138–8143.

Sarkar, K. (2019). Sentiment polarity detection in bengali tweets using lstm recurrent neural networks, *2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP)*, IEEE, pp. 1–6.

Shahariar, G., Biswas, S., Omar, F., Shah, F. M. and Hassan, S. B. (2019). Spam review detection using deep learning, *2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, IEEE, pp. 0027–0033.

Tammina, S. and Annareddy, S. (2020). Sentiment analysis on customer reviews using convolutional neural network, *2020 International Conference on Computer Communication and Informatics (ICCCI)*, IEEE, pp. 1–6.

Tufail, H., Ashraf, M. U., Alsubhi, K. and Aljahdali, H. M. (2022). The effect of fake reviews on e-commerce during and after covid-19 pandemic: Skl-based fake reviews detection, *IEEE Access* **10**: 25555–25564.

Yang, J. and Yang, J. (2020). Aspect based sentiment analysis with self-attention and gated convolutional networks, *2020 IEEE 11th International Conference on Software Engineering and Service Science (ICSESS)*, IEEE, pp. 146–149.