2023 26th International Conference on Computer and Information Technology (ICCIT)
13-15 December, Cox's Bazar, Bangladesh

BERT-Driven Sentiment Analysis: Deciphering Amazon Mobile Reviews in the E-commerce Landscape

Syed Istiaque Ahmed, Tahsin Tanim Ramisha, Tasfia Ayesha Maliha, Tasnia Juha,
Ehsanur Rahman Rhythm, and Md Humaion Kabir Mehedi and Annajiat Alim Rasel
Department of Computer Science and Engineering (CSE)
School of Data and Sciences (SDS)
Brac University
66 Mohakhali, Dhaka - 1212, Bangladesh
{syed.istiaque.ahmed, tahsin.tanim.ramisha, tasfia.ayesha, tasnia.juha, ehsanur.rahman.rhythm,
humaion.kabir.mehedi}@g.bracu.ac.bd, annajiat@gmail.com

***Abstract*—Recently, the e-commerce industry has experienced rapid development, leading to a significant rise in online purchasing. Consequently, there has been an increase in the volume of customer reviews for products available online. Purchasing decisions are highly influenced by customer reviews, as they are influenced by recommendations from other consumers. In this research, a dataset of Amazon reviews is analyzed to investigate sentiment classification using different machine-learning methods. In the first stage by using techniques such as (BOW) bag of words, Glove, and tfidf. the reviews are converted into vector representation. Next, multiple machine learning algorithms including Naïve Bayes, Logistic Regression, Bert models, and LSTM or Bidirectional Long-Short Term Memory are trained. The performance of these models is evaluated using metrics such as accuracy, recall, precision, f1 score, and cross-entropy loss function. Furthermore and more importantly, in order to identify the best performing model for sentiment classification analysis purposes specifically, the top-performing model is thoroughly examined. The experiment encompasses multiclass classifications initially, followed by selecting the most successful model which ultimately gets retrained on binary classification tasks.**

***Index Terms*—Product Sentiment, BERT, Data Analytics**

I. Introduction

In today's digitalized world, e-commerce has gained significant prominence for its convenient access to a wide range of products. Moreover, ecommerce websites also provide platforms for individuals to express their thoughts and emotions. The opinions and experiences shared by fellow customers have a significant impact on our purchasing decisions. We often seek insights from others to benefit from their experiences, highlighting the increasing importance of reviews. However, the abundance of these reviews can overwhelm customers.This is where sentiment analysis becomes crucial where it helps us analyze and comprehend the overall sentiments conveyed in these reviews. The research suggests, utilizing advanced machine learning algorithms for sentiment analysis on customer reviews of Amazon mobile phones dataset. This approach aims to predict the polarity of these reviews and facilitate informed purchasing decisions based on others' experiences. Moreover, it provides valuable feedback for businesses to enhance their sells, production, and products by understanding customer opinions and needs. Customer reviews or ratings reflect the writer's experience with a product, which can be positive, negative, or impartial. Some customers express satisfaction by giving a product a rating of four or five stars, while others show dissatisfaction with one or two stars. Analyzing these sentiments poses no difficulty. However, there are instances where customers give three stars despite being satisfied, which can confuse both other customers and companies seeking genuine opinions. As a result, both customers and companies face difficulties when analyzing reviews and understanding consumer satisfaction. It is important to recognize that a three-star rating does not necessarily imply a neutral opiniont. In reality, people who give a three star rating may have biases towards either positive or negative aspects.

This research proposes the utilization of sentiment analysis to determine the polarity of Amazon mobile phone dataset reviews. While keeping the three star rating intact and considering it representative of neutrality, this approach aims to raise the complexity of the study while assessing the effectiveness of advanced Natural Language Processing (NLP) models like BERT in handling intricate classification problems. In addition, three diverse machine learning models - Logistic Regression, BiLSTM and Naïve Bayes will be employed utilizing various feature extraction techniques. Subsequently, an examination will be conducted to determine which model performs best for sentiment classification. Towards the end of this study, with regards to refining model performance, we will take the highest performing model and retrain it without considering the neutral class within our dataset. This adjustment effectively transforms the problem into a binary classification task. By doing so, we aim to evaluate how this recasting impacts overall model performance. The objective is to measure any potential effects resulting from reframing the problem at hand while still maintaining focus on narrative-driven content that presents information logically.

979-8-3503-5901-5/23/$31.00 ©2023 IEEE

II. Literature Review

Opinion mining through document analysis serves a crucial function in understanding consumer characteristics and preferences in relation to particular products. In recent years, there has been an intriguing change in research purpose, with a significant rise of interest directed towards opinion mining. This research includes the datasets and pre-processing, methodology, and assessment measures that we employed in our investigations.

The study by S.A.Aljuhani focuses on extracting data from Amazon customer reviews focusing on three products (Redmi Note 3, Apple iPhone 5S, and Samsung J7), while other researchers have utilized Amazon customer reviews for a variety of reasons [1]. In the research of A. S. Rathor, they utilized the Amazon API in English to retrieve 21,500 reviews, from which they selected 3,000 at random to employ in an experiment [2]. In these two researches of S. A. Aljuhani and B. Bansal based on sentiment analysis and sentiment classification of customer reviews, a dataset from Amazon's mobile phone category containing over 400,000 consumer evaluations was examined [1], [3]. Cernian et al. constructed a dataset of 300 Amazon product reviews [4]. Amazon groups of products including books, cameras, and GPS were analyzed in a separate study by Moraes et al. [5], with each dataset containing approximately two thousand reviews (one thousand positives and one thousand negatives). The researchers divided the reviews into two categories: positive and negative reviews. Using Amazon's API, [6] gathered more than a hundred thousand reviews written in Chinese about various clothing items. The researcher examined more than 1,000 customer comments on product reviews on Amazon [7].

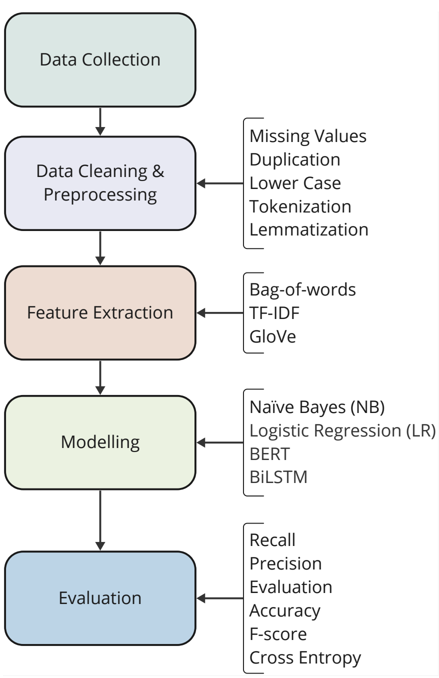

Fig. 1. Sentiment analysis methodology

The interpretation of textual data is carried out in multiple phases using an extensive range of methodologies, although pre-processing of the text is frequently acknowledged as one of the most fundamental NLP steps in boosting data quality. In addition to the various steps that can be taken, the most common ones are tokenization,stop-word removal, POS (Parts of Speech) tagging, stemming, and lemmatization. In the pre-processing phase, stop words are removed because they are not beneficial to explain the text or help in the analysis. Tokenization is a technique of removing punctuation marks from a sentence in order to separate the statement into meaningful tokens, which can be symbols, phrases, or even individual words. The process of merging a word's many diverse forms into one standard form, also known as the canonical form, is known as lemmatization, while stemming is the reverse of this. Stemming is the transformation of a word into its root form. In natural language processing, identifying the various parts of speech contained inside the text is an extremely important step that is accomplished through the use of POS tagging.

In the literature on sentiment analysis, it has been pointed out that Artificial Neural Networks (ANN) are infrequently contrasted to other methodologies. The intended effort of this research [8] was to give a concrete comparison between ANN (Artificial Neural Networks) and SVM (Support Vector Machines) regarding document-level sentiment analysis. In 2015, the authors of the promising research [4] proposed a semantic method for the construction of an application for sentiment analysis, with the use of the SentiWordNet lexical resource serving as the method's primary basis. In another follow-up work [6], the authors adopted a semantic strategy, proposing a word2vec and SVMperf-based sentiment classification. The research has been divided into two distinguishable branches. By using word2vec it better captures the semantic features of the targeted domain by clustering comparable features. In second step of the process, they classified the comment texts with SVMperf. In their study [7], the authors investigated the use of SVM (support vector machine) for sentiment classification. The overall performance of this proposed method is superior to that of separate methods.

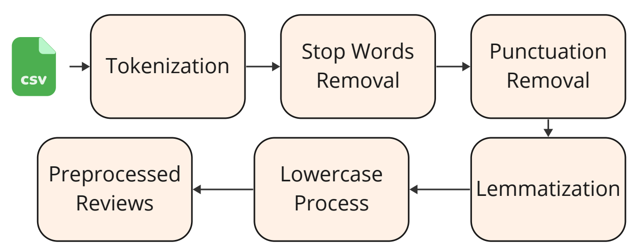

Fig. 2. Preprocessing Steps

III. Methodology

This section summarizes how sentiment analysis is used to examine Amazon's mobile phone reviews. The study progresses through various stages, from data gathering to evaluating each predictive model, as depicted in Figure 1.

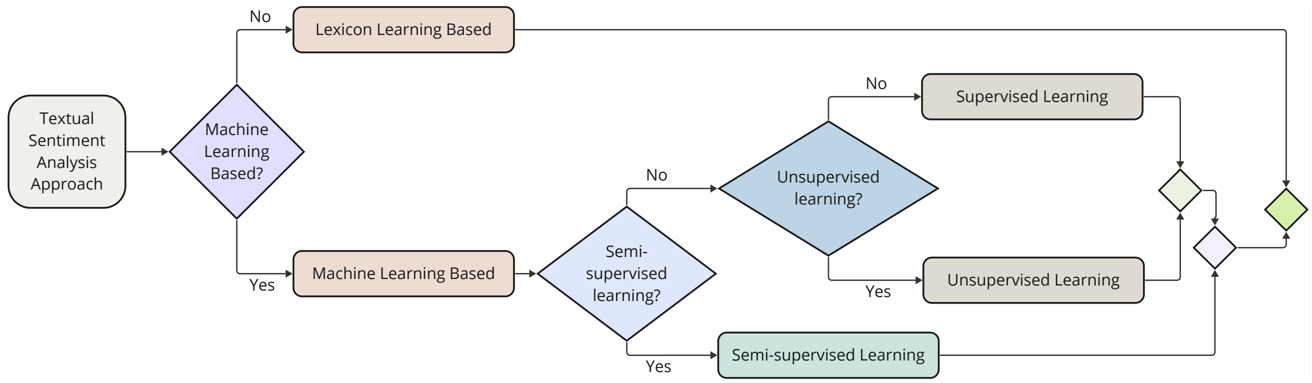

Fig. 3. Sentiment analysis approaches

In Natural Language Processing (NLP), effective text pre-processing plays a crucial role in enhancing the quality of textual data. Specifically, Figure 2. provides an illustrative representation of all the pre-processing steps undertaken for this research on the Amazon mobile phone dataset. To ensure consistency and uniformity, all reviews were subjected to lowercasing, thereby eliminating any mixed capitalization. For instance, words like "Bad" and "ExCeLLent" were transformed into their lowercase equivalents: "bad" and "excellent". Moreover, punctuation marks and frequently occurring insignificant words (known as stop words) such as ("-", "/", ":", "?", "the", and "a") were removed as they do not significantly impact meaning. Another important step involved tokenizing the reviews, which involves breaking down each sentence into individual units referred to as tokens or words. Typically, a space character acts as a separator between tokens. Consequently, word separation relies on this space character during the tokenizing process [9] [10]. Finally, all tokens underwent lemmatization to restore them to their base or dictionary form.

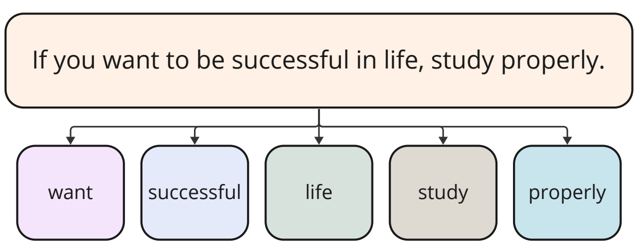

Fig. 4. Review after pre-processing steps

The dataset comprised reviews that were categorized as positive, neutral, or negative based on their star ratings, as outlined in. The data was then divided into different segments for training (60%), testing (20%), and validation (20%) purposes to build standard models. However, when employing deep learning approaches, the distribution changed to (90%) for training, (5%) for testing, and (5%) for validation.

In the realm of data analysis, classification stands as a pivotal method that systematically categorizes data into distinct classes [11]. In Sentiment Analysis, this approach is employed to categorize data into two or three distinct categories namely "negative", "positive", and optionally "neutral". For binary classification, we use positive, negative and for ternary we use positive, negative and neutral. The effectiveness of sentiment analysis heavily relies on this classification process [12]. Predominantly, two methodologies are adopted for sentiment categorization of consumer feedback: the machine learning (ML) approach and the lexicon-based approach, [13] as depicted in Figure 3.

The lexicon-based strategy ascertains the sentiment of textual reviews by relying on words annotated with specific polarity or associated polarity scores [14]. Conversely, machine learning techniques bifurcate into supervised and unsupervised learning categories. Our research predominantly harnesses supervised machine learning, a prevalent method for constructing sentiment classification frameworks in sentiment analysis. Initially, these frameworks curate a training dataset and subsequently annotate this data with sentiment labels. Following this, the suite of featurs is channeled to a classification model by extracting from the training data. Examples of such models include Naïve Bayes (NB) and Logistic Regression (LR) among others [15]. Post the training phase, utilizing sentiment labels, the classifier is adept at predicting the sentiment inclination of new, unseen data samples.

IV. Data Collection and Analysis

There has been an upsurge in the number of different items and services in recent years. As a result, consumer reviews are easily accessible. Typically, there are two aspects that make up a customer review: the star rating and the text review [16]. Our research will employ ratings and reviews from actual customers to establish if a given review falls under the "Positive," "Negative," or "Neutral" heading.

Four hundred thousand customer reviews of unlocked mobile phones that are sold on Amazon.com were gathered by PromptCloud which was carried out in December of 2016 with the assistance of specialist web crawlers that were designed for our data extraction services. The task's goal was to get a better understanding of reviews, ratings, and prices, as well as the relationships between these three factors.

More than 40,000 reviews provided by real users about mobile phones make up the bulk of the collection. In particular, it comprises of 6 features including 413,840 reviews which can be categorized as follows:

1) Information regarding mobile phones, containing brand names, product names, prices, and ratings.
2) Information regarding reviews, consisting of review votes and reviews altogether.

Data cleaning is, without a doubt, the procedure of examining of whole information contained in a data frame and removing or updating any missing informations, wrong, or repeated. This procedure can be beneficial in optimizing the precision of the data, which is essential given that the integrity of the data is vital for obtaining the intended result in an approach that is both accurate and efficient. Figure 5 displays the characteristics of the Amazon mobile phone dataset for which values are missing. Some efforts have been made to modify the data.

Features	Description	Data type	Missing Value
Brand Name	*Brand Name (such as Apple or LG) of Manufacturer*	Object	65,171
Product Name	The label that manufacturers put on their various smartphone models.	Object	0
Price	The cost of the mobile phone	Float	5,933
Rating	Number of star rating [1–5]	Integer	0
Reviews	Feedback about the mobile phone from consumers.	Object	62
Review Votes	The number of consumers who participated in the reviews' voting.	Float	12,296

Fig. 5. Data description of amazon mobile phone

To begin, entries in the “Reviews” column that contained a null value were removed.The second modification was to replace the 12,296 blanks in the “Review Votes” column with the value 0. Due to the fact that these are legitimate reviews that garnered no votes, the value of zero is the one that is considered to be the most suitable in this case. Thirdly, all null values in the column labeled ”Price” were replaced with the number 144.71 because it is the median value of the price and there were 12,296 null values. Furthermore, we dealt with the issue of redundancy in the data frame by removing all of the duplicate data entries (a total of 64,308 records). After That, mobile accessory brand names were removed from the “Brand Name” field. Since many brand names have typos and are not standardized, we used normalization technique for the brand name by switching out the misspelled words with the proper ones and adding uniformity where needed. As an illustration of this stage, the phrase “Lenovo Manufacturer” has been changed to “Lenovo,” and the word “Samsung” has been changed to “Samsung.” Following the completion of this process, the overall quantity of brand names experienced a decline from 382 to 282.

EDA or exploratory Data Analysis, for short, is a process designed to render data that is presented in columns and rows simpler to understand through the use of visualization and interpretation. In addition to this, we managed to derive the maximum amount of clarity possible from the dataset through the use of a variety of visualizations. We investigated how reviews of mobile phones purchased by Amazon.com were dispersed according to both star rating and review count. According to the data presented, we can visualize that the count of five stars is the most common, whilst the count of two stars is the least prevalent. Contrarily, the data presented in figure 6 reveals that 44.2% of customers rated the product five stars, 17.7% rated it four stars, 9.3% rated it three stars, 7.3% rated it two stars, and 21.5% rated it one star.

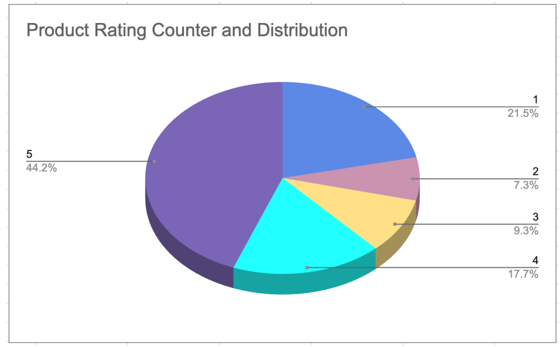

Fig. 6. Rating counter over the reviews

Afterward, an investigation is conducted on the polarity of the dataset pertaining to Amazon mobile phones, which is then visually depicted using the format of a pie chart. The newly incorporated feature in consideration is the review sentiment column, which encompasses three unique indicators. Ratings of 4 or 5 will be classified as ”Positive,” while ratings of 1 or 2 will be classified as ”Negative,” and a rating of 3 will be classified as ”Neutral.” According to Figure 7, the data reveals that 67.5% of the reviews are categorized as “Positive,” while 24.6% fall under the classification of “Negative,” and 7.97% are labeled as “Neutral.” Apart from that, an examination was conducted on the number of reviews and the occurrence of names of brands. According to the statistics, Samsung had the highest number of evaluations, which amounted to 58,394, while Apple and BLU received 51,070 and 51,781 reviews respectively.

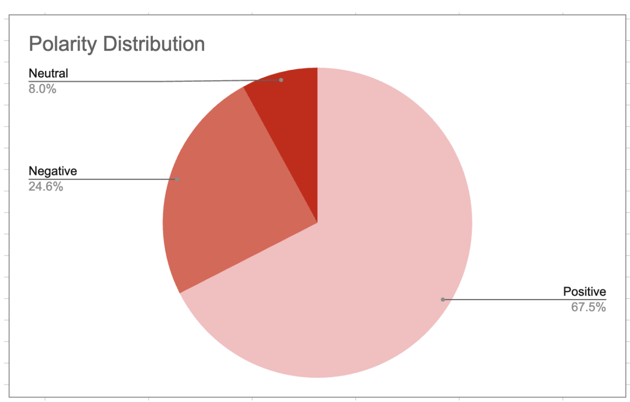

Fig. 7. Amazon mobile phone dataset’s polarity distribution of

V. Results and Analysis

The experiments conducted throughout this project involved two phases. The first phase consisted of a multiclass classification task employing all proposed models. The second phase focused on binary classification using only the model that demonstrated the highest accuracy during validation.

Based on the star ratings, we divided dataset of Amazon mobile phone reviews into three categories: positive (2), negative (1), and neutral (0) attitudes. In particular, ratings of one and two stars were viewed negatively, but those of five and four stars were positive. Three-star ratings fell under the neutral category. All suggested models were used to extract features from these reviews. [15]

By using the Bag-of-Words (BOW), GloVe and TF-IDF feature extraction method, the logistic regression method (LR) was used. The outcomes are shown in Table I. With an accuracy of 85.5% and a cross-entropy loss of 0.59, LR with BOW (Bigram) astonishingly showed exceptional performance. Additionally, it was shown that LR with BOW (Bigram) had higher recall compared to other strategies when evaluating various feature extraction techniques.

TABLE I
RESULT OF (LR) LOGISTIC REGRESSION MODEL

	LR-BOW	LR-BOW-B	LR-BOW-TRI	LR-TF-IDF	LR-Glove
Accuracy	81%	85%	83%	83%	78%
Recall	81%	86%	83.2%	83%	78%
Precision	84%	8%6	84%	88%	84%
F1 Score	82%	86%	83%	85%	80%
Cross Loss	0.72	0.59	0.64	0.5	0.64

Results of the NB classifier is demonstrated in Table II. Here, the results are shown by eliminating the GloVe, the word vector illustration by grouping related words together. However, feature independence is assumed by NB theory. The NB classifier using BOW (Trigram) obtains higher accuracy than among all the feature extraction methods. However, it has also shown a greater cross-entropy loss than any other methods with the value of 78% and 1.176 respectively.

TABLE II
RESULT OF NAÏVE BAYES (NB) CLASSIFIER

	NB-BOW	NB-BOW-Bi	NB-BOW-TRI	NB-TF-IDF
Accuracy	70%	76%	78%	71%
Recall	70%	76%	78%	71%
Precision	74%	78%	79%	74%
F1 Score	71%	76%	78%	72%
Cross Loss	3.04	2.33	1.18	2.99

Table III displays the results of BERT, showcasing an impressive accuracy rate of 94% and a cross-entropy loss of only 0.189. These numbers indicate that the model's performance is deemed quite favorable. Furthermore, this model has good recall and accuracy.

Table IV displays the results of Bi-LSTM by using two different embeddings: jointly trained embeddings and fine-tuned GloVe embeddings. The model reaches a level of accuracy up to 93%, with a cross-entropy loss value of 0.189. This

TABLE III
RESULT OF BERT

	BERT
Accuracy	94.7%
Recall	94.7%
Precision	94.6%
F1 Score	94.6%
Cross Loss	0.189

achievement is noteworthy because the structure of Bi-RNN allows for simultaneous training in both backward and forward time directions, thereby incorporating information from both ends of a sequence. Consequently, this design significantly enhances overall model performance.

TABLE IV
RESULT OF BI-LSTM

	Accuracy	Recall	Precision	F1 score	Cross Loss
Bi-LSTM (Joint trained embedding)	92%	92.9%	92.5%	92.5%	0.24
Bi-LSTM (GloVe embedding)	92.5%	93.3%	92.9%	92.9%	0.234

Table V summarizes all classification model results. Among them, we can see 94.7% of accuracy is from BERT, which outperforms other models used. Moreover, 92.5% accuracy has been seen form Bi-LTSM (GloVE embedding) lowest accuracy whereas LR is second last respectively. Notably, Figure 8 demonstrates that the BERT model outperforms other models in terms of accuracy.

TABLE V
FINAL RESULT FOR ALL CLASSIFICATION MODELS

	LR-BOW-Bi	NB-BOW-TRI	BERT	Bi-LSTM
Accuracy	85.3%	78.4%	94.7%	93.3%
Recall	85.3%	78.4%	94.7%	93.3%
Precision	86.3%	78.4%	96.4%	92.9%
F1 Score	85.7%	77.9%	94.6%	92.9%
Cross Loss	0.596	1.159	0.189	0.234

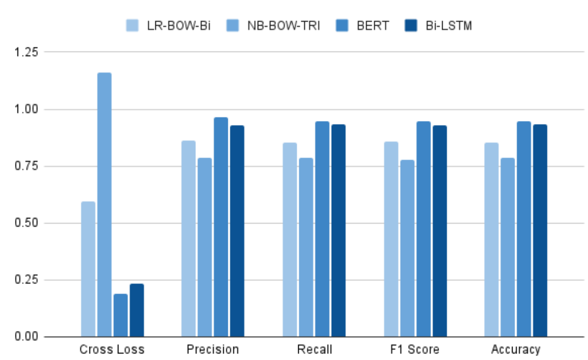

Fig. 8. Final results for multiclass classification

Table VI illustrates the binary classification results. Among these models, BERT stands out with an outstanding accuracy of 98%. Additionally, Bi-LSTM delivers commendable results with an accuracy rate of 97%. As opposed to that, Notably, Figure 8 depicts by comparing with outer models the BERT model excels in binary classification. By recasting the problem as a binary classification scenario we have observed an improvement in the performance of all models. This reinforces that the overall performance is impacted by the neutral class significantly.

TABLE VI
FINAL RESULTS OF VALIDATION DATASET USING BINARY CLASSIFICATION

	LR-BOW-Bi	NB-BOW-TRI	BERT	Bi-LSTM
Accuracy	93.1%	88%	98.3%	97.7%
Recall	93.1%	94.2%	98.6%	97.7%
Precision	93.1%	94.3%	98.6%	97.7%
F1 Score	93.1%	94.2%	98.6%	97.7%
Cross Loss	0.27	0.209	0.064	0.088

TABLE VII
FINAL RESULTS OF TEST DATASET USING BINARY CLASSIFICATION

	LR-BOW-Bi	NB-BOW-TRI	BERT	Bi-LSTM
Accuracy	92.7%	87.8%	98.4%	97.4%
Recall	92.7%	87.8%	98.4%	97.4%
Precision	92.8%	87.6%	984.%	97.4%
F1 Score	92.8%	87.2%	98.4%	97.4%
Cross Loss	0.275	0.429	0.071	0.092

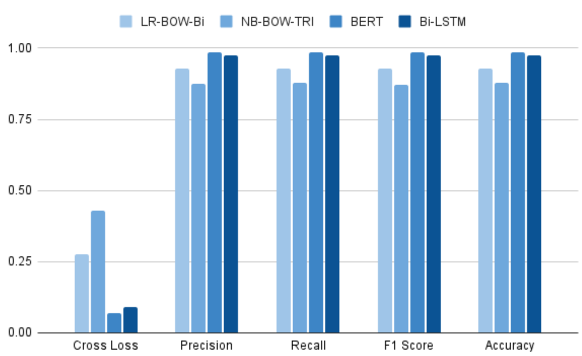

Fig. 9. Final results for Binary classification

In the research “Sentiment classification of online consumer reviews using word vector representations”, researchers employed LR and NB for binary classification [3]. Furthermore, “Random forest and support vector machine based hybrid approach to sentiment analysis” utilized CNN, Naïve Bayes, Stochastic Gradient Descent, and Logistic Regression for multiclass classification [7]. In multiclass classification, our findings utilizing the BERT model reached 94% accuracy. At the same time, the word2vec is used in the CNN model to extract features performed the best, with a 92.7% accuracy. We obtain useful insights into the performance of various models in sentiment analysis activities on Amazon mobile phone reviews by examining these results with current studies.

VI. CONCLUTION

Sentiment analysis plays a crucial role in extracting valuable insights from textual data on e-commerce websites. These platforms generate an immense amount of text daily, consisting of comments, criticism, tweets, and ideas conveying user opinions through emoticons, ratings, and reviews. By analyzing these reviews, customers can make well-informed decisions about products. This study explores the application of binary classification and multiclass techniques for Amazon phones using various machine learning approaches like LR (logistic regression) and NB (Naïve bayes). Various feature extraction methods are employed alongside joint-learned embedding and GloVe embedding with bidirectional Long-Short Memory (Bi-LSTM). Additionally, the Bidirectional Encoder Representations from Transformers (Bart) model is utilized. The Bart model demonstrates promising results with a high accuracy rate of 94% for multi class classification and 98% for binary classification. Equally impressive is the performance achieved by Bi-LSTM with joint-learned embedding, boasting a 93% accuracy in multiclass classification and 97% accuracy in binary classification.

REFERENCES

[1] S. A. Aljuhani and N. S. Alghamdi, “A comparison of sentiment analysis methods on amazon reviews of mobile phones,” *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 6, 2019. [Online]. Available: http://dx.doi.org/10.14569/IJACSA.2019.0100678

[2] A. S. Rathor, A. Agarwal, and P. Dimri, “Comparative study of machine learning approaches for amazon reviews,” *Procedia Computer Science*, vol. 132, pp. 1552–1561, 2018, international Conference on Computational Intelligence and Data Science. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050918308512

[3] B. Bansal and S. Srivastava, “Sentiment classification of online consumer reviews using word vector representations,” *Procedia Computer Science*, vol. 132, pp. 1147–1153, 01 2018.

[4] A. Cernian, V. Sgarciu, and B. Martin, “Sentiment analysis from product reviews using sentiwordnet as lexical resource,” in *2015 7th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, 2015, pp. WE–15–WE–18.

[5] R. Moraes, J. F. Valiati, and W. P. Gavião Neto, “Document-level sentiment classification: An empirical comparison between svm and ann,” *Expert Systems with Applications*, vol. 40, no. 2, pp. 621–633, 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417412009153

[6] D. Zhang, H. Xu, Z. Su, and Y. Xu, “Chinese comments sentiment classification based on word2vec and svmperf,” *Expert Systems with Applications*, vol. 42, no. 4, pp. 1857–1863, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417414005508

[7] Y. Al Amrani, M. Lazaar, and K. E. El Kadiri, “Random forest and support vector machine based hybrid approach to sentiment analysis,” *Procedia Computer Science*, vol. 127, pp. 511–520, 2018, pROCEEDINGS OF THE FIRST INTERNATIONAL CONFERENCE ON INTELLIGENT COMPUTING IN DATA SCIENCES, ICDS2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050918301625

[8] R. Moraes, J. F. Valiati, and W. P. G. Neto, “Document-level sentiment classification: An empirical comparison between svm and ann,” *Expert Syst. Appl.*, vol. 40, pp. 621–633, 2013. [Online]. Available: https://api.semanticscholar.org/CorpusID:5868583

[9] B. Bahrawi, “Sentiment analysis using random forest algorithm-online social media based,” *Journal of Information Technology and Its Utilization*, vol. 2, no. 2, pp. 29–33, 2019.

[10] M. Fikri and R. Sarno, “A comparative study of sentiment analysis using svm and senti word net,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 13, no. 3, pp. 902–909, 2019.

[11] H. Sinha and A. Kaur, “A detailed survey and comparative study of sentiment analysis algorithms,” in *2016 2nd International Conference on Communication Control and Intelligent Systems (CCIS)*, 2016, pp. 94–98.

[12] M. Bouazizi and T. Ohtsuki, “A pattern-based approach for multi-class sentiment analysis in twitter,” *IEEE Access*, vol. 5, pp. 20 617–20 639, 2017.

[13] H. Kaur, V. Mangat, and Nidhi, “A survey of sentiment analysis techniques,” in *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, 2017, pp. 921–925.

[14] Z. Li, Y. Fan, B. Jiang, T. Lei, and W. Liu, “A survey on sentiment analysis and opinion mining for social multimedia,” *Multimedia Tools and Applications*, vol. 78, pp. 6939–6967, 2019.

[15] A. S. M. AlQahtani, “Product sentiment analysis for amazon reviews,” Aug 2021. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3886135

[16] S. M. Mudambi, D. Schuff, and Z. Zhang, “Why aren’t the stars aligned? an analysis of online review content and star ratings,” in *2014 47th Hawaii International Conference on System Sciences*, 2014, pp. 3139–3147.