**ABSTRACT:**

Online shopping is currently experiencing rapid growth. As a consequence, buying stuff online has increased, which caused an increase in customer feedback on the internet of items. Because the Consumer's View Regarding the Goods or Service is Affected by Other Consumers' Suggestions or Concerns, the Suggested Views in Reviews from Customers Have a Huge Impact on Customer's Buying Behaviour. This study analyses the Amazon Reviews Dataset and investigates sentiment classification using several machine learning approaches. Initially the reviews were converted to vector representations by multiple methods such as Bag-Of-Words, Tf-Idf, and Glove. Afterwards we trained a variety of machine learning algorithms, including Logistic Regression, Random Forest, Naive Bayes, Linear Support Vector Machine, and BERT. The mathematical models were then evaluated using Accuracy, F1-Score, Precision, and Recall. Following that, we investigated the Sentiment Classification of the most successful Algorithm. The study was carried out on Multiclass Classifications, after which we chose the best performing model.

**INTRODUCTION:**

Everything around us is growing more computerised. Online shopping is gaining traction in our digitalized environment due to the accessibility of items within consumers' grasp. Furthermore, the online store allows individuals to express their thoughts and feelings. In reality, users are increasingly depending on other consumer feedback. Our views and purchase decisions are influenced by the experiences of others and their product reviews. We always seek people for their opinions in order to profit from their experience; therefore, the value of feedback has risen. Nevertheless, it is very difficult for consumers to read all of these evaluations; hence, sentiment analysis plays an important part in analysing them. Using supervised machine learning methods, this study presents an analysis of sentiment to forecast the polarity of Amazon mobile phone dataset reviews. Emotion research assists customers in making purchase decisions based on the experiences of others. Furthermore, knowing what consumers think and wants will assist businesses in improving their products. [1].

**Problem Statement:**

Customer reviews or ratings seek to define the writer's attitude towards the product. It could be either positive, negative, or neutral. Some consumers offer a product four or five stars to show their overall contentment, while others assign it one or two stars to show their overall discontent. This poses no difficulties in analysing emotions. Others, on the other hand, give it three stars, indicating their overall contentment. This confuses consumers as well as companies that wish to know their true feelings. As a result, customers and businesses are having problems analysing feedback and understanding consumer happiness. As a result, the three-star rating does not truly indicate a neutral feeling, because in reality, people who give a product or service three stars do not necessarily imply that they are evenly balanced in their positive and negative opinions. This study presents an analysis of sentiment to predict the neutrality of Amazon mobile phone dataset reviews based on this argument. We'll leave the 3-star rating alone because it represents a neutral viewpoint. This is done to increase the study's complexity and difficulty, as well as to assess the effectiveness of cutting-edge NLP models like BERT in tackling challenging categorization issues. This study will also employ four machine-learning models with distinct feature extraction techniques: Logistic Regression, Linear Support Vector, Nave Bayes, Random Forest and BERT. The best performance model is then examined in order to explore sentiment categorization. At the end of the study, we will retrain the highest performing model on the dataset without the neutral class, essentially recasting the problem as a binary-classification problem. We want to see how this shifting of the problem impacts the performance of the model.

**RELATED WORKS:**

**1. Data Collection and Pre-processing:**

In various sentiment analysis studies, data has been collected from diverse platforms, including Twitter tweets, business evaluations, and consumer comments. For instance, in [2], researchers focused on extracting Amazon website reviews for products like the Apple iPhone 5S, Samsung J7, and Redmi Note 3. In [3], 21,500 Amazon reviews were acquired and 3,000 were randomly selected for analysis. [1] and [4] performed experiments using a dataset of over 400,000 user evaluations in the mobile phone industry. [5] utilized a collection of 300 Amazon evaluations for digital products, while [6] analyzed Amazon reviews for various product categories. Moreover, [7] gathered around 100,000 Chinese feedback on apparel goods from Amazon, and [8] evaluated over 1,000 Amazon ratings. In each case, text pre-processing is considered a crucial phase to enhance NLP accuracy. It involves stages like stop-word elimination, tokenization, stemming, lemmatization, and POS labelling.

**2. Text Pre-processing Techniques:**

Text pre-processing is a critical step in natural language processing (NLP) to improve text-based analysis. In [2], all stop words were removed from Amazon reviews, while in [3], letter sequences used for emphasis were deleted and replaced. [4] employed stemming, lowercasing, punctuation removal, and white space deletion. [1] included tokenization, lowercasing, spelling verification, and lemmatization in their pre-processing. [6] combined tokenization and stemming in their pre-processing. Notably, [9] underwent various processing procedures, including character translations and punctuation removal. [5], on the other hand, identified parts of speech following data extraction and then divided statements into vectors of words with meanings derived from SentiWordNet. In [7], Chinese comment strings were split into words and labeled with appropriate POS tags using the ICTCLAS 4 system.

**3. Sentiment Analysis Methods:**

In sentiment analysis, researchers utilized both machine learning approaches and lexicon-based methods. For example, [2] used the Naive Bayes classifier (NB) to categorize reviews as good or poor, outperforming Logistic Regression (LR) and SentiWordNet. Performance was evaluated using Recall, Precision, and F-measure. [9] tested NB at both phrase and review levels, calculating TF-IDF word frequency ranges. [3] employed Support Vector Machine (SVM), NB, and Maximum Entropy (ME) for identifying comments as positive, negative, or neutral, achieving an optimal accuracy of 81% with SVM. [4] used Continuous Bag of Words (CBOW) and skip-gram approaches with various classification algorithms, with Random Forest (RF) and CBOW providing the highest accuracy (91%). [5] explored machine learning algorithms, including LR, SGD, NB, and CNN, using various feature extraction methods, with CNN and word2vec achieving 91% accuracy and using Lime for analytical reasoning. It's worth noting that Artificial Neural Networks (ANN) are rarely employed in sentiment analysis research, as highlighted in [6].

**4. Advanced Sentiment Analysis Approaches:**

In some cases, more advanced sentiment analysis techniques were employed. [6] compared SVM and ANN for document-level sentiment analysis, shedding light on the underutilization of ANN in sentiment analysis research. In [5], a semantic method based on the SentiWordNet lexical resource was used, while [7] presented a sentiment classification based on word2vec and SVMperf, employing semantic aspects from word2vec and classifying feedback texts using SVMperf. In [8], sentiment categorization was investigated using SVM, RF, and a hybrid technique called Random Forest Support Vector Machine (RFSVM), which outperformed specific approaches.

**Methodology:**

This part offers a description of the suggested sentiment analysis approach for Amazon mobile phone reviews. Figure 1 demonstrates the steps of this particular effort, beginning with data collection and ending by the evaluation every categorization algorithms.

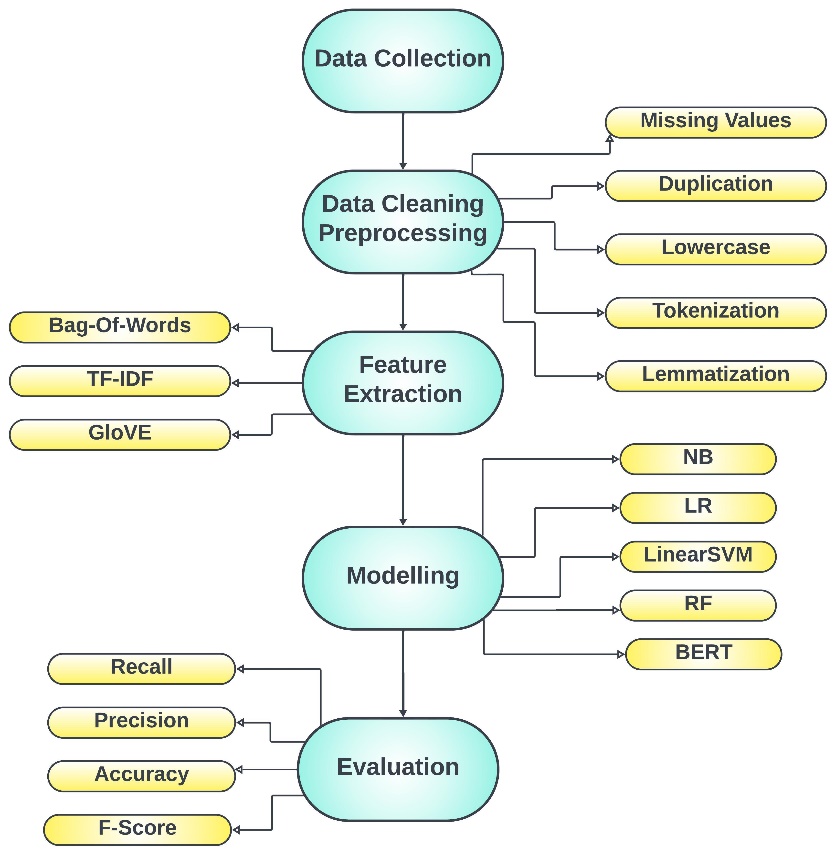


Figure 1 General sentiment analysis approach for Amazon mobile phone reviews

**Data-Preprocessing:**

Text pre-processing is an essential phase in NLP to improve textual data integrity. The picture following Figure 2 displays all pre-processing processes used in this study on the Amazon mobile phone data. The feedbacks where evaluations have been processed by changing all of the characters to lowercase, rather than combining capitals and lowercase; for example, "THANK YOU," and "HapPy" are transformed to "thank you" and "happy." In addition, any punctuation and stop words that occur often but have no effect on meaning, such as "-, /,:,?, the, a, an" were removed. Furthermore, the reviews were tokenized, which is a method of breaking down a text into a series of terms known as "tokens." Typically, a space character distinguishes or separates one token from others; hence, the tokenizing process relies on the space character to execute word divisions [11] [12]. Then, using the lemmatization procedure, all tokens were restored to their base or vocabulary version.

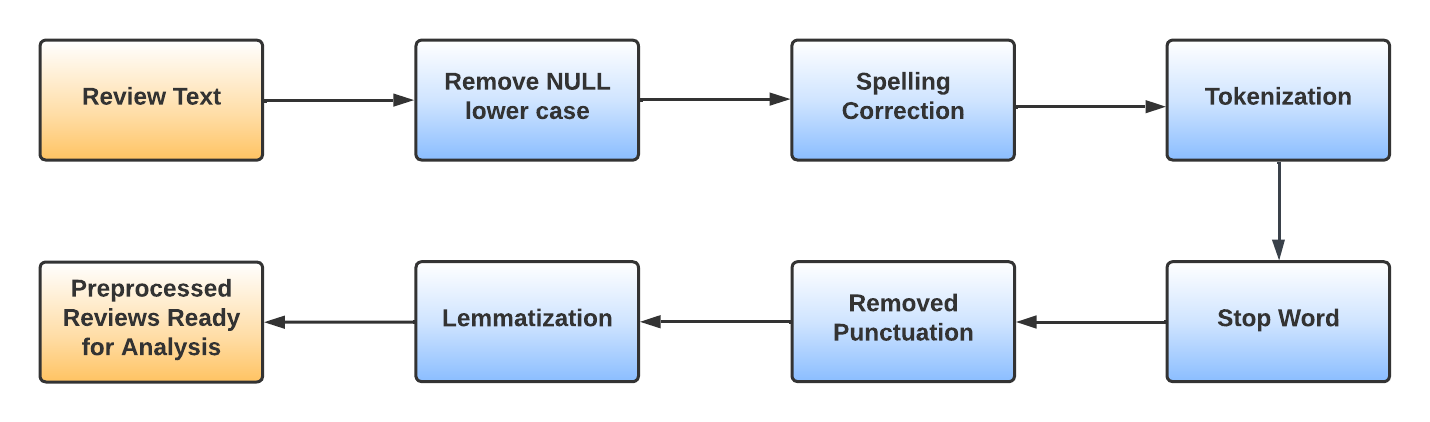


Figure 2 Overview of Pre-processing Procedures

In a similar manner, every review in the dataset was labelled as positive, negative, or neutral based on its star rating. The dataset was then partitioned between 60% training, 20% validation, and 20% testing for initial models.

**Feature Extraction:**

Natural Language Processing (NLP) calls for machines to decode words spoken by humans. Initially, the written input is translated to numbers that can be used by machine learning models. This project makes use of Bag of Words (BOW), Term frequency - Inverse document frequency (TF-IDF) , and GloVe. They are described in the sections that follow.

**Sentiment Analysis Methods:**

Classification of information is a strategy for classifying data [13]. It is used in the field of Sentiment Analysis for categorising data into binaries as "positive", "neutral" and "negative" and ternary such as "positive," "negative," and "neutral" classifications, and then the sentiment analysis procedure is finished [14]. In sentimental categorization of consumer evaluations, two methodologies are commonly used: lexicon-based and machine learning [15], as seen in Figure 3 .

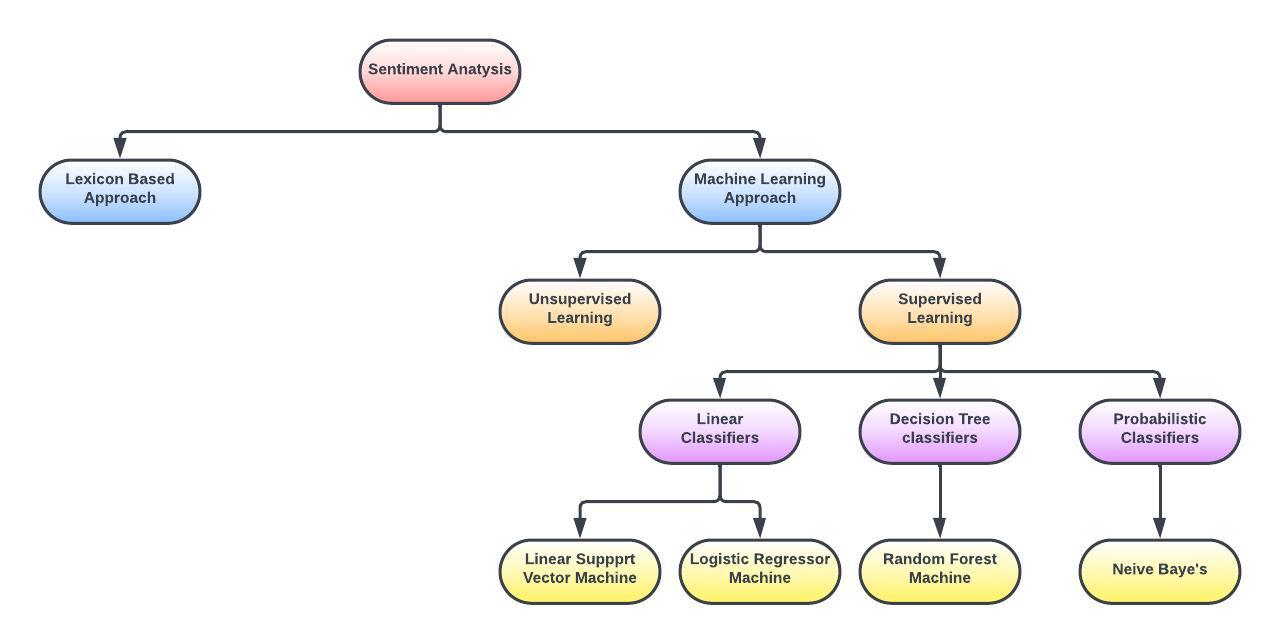


Figure 3 Techniques of sentiment analysis

Utilising terms tagged with polarity or polarity ratings, lexicon-based techniques forecast the polarity of text reviews [16]. Machine learning approaches, on the other hand, are classified as supervised learning and unsupervised learning. In the subject of sentiment evaluation, supervised machine learning is commonly utilised to develop categorization of sentiment methods. The aforementioned models begin by creating a training set and tagging it with feelings. Following that, using the training data, a set of features is extracted and fed into a classifier model, such as Naive Bayes (NB), Logistic Regression (LR), Linear Support Vector Machine(LinearSVM), Random Forest (RF), and so on. Following the sentiment tag training step, the classifier may be used for predicting the sentiments direction of an experiment on fresh data.

**Performance Evaluation Parameters:**

The classification techniques' performance may be determined using Accuracy, F-Score, Recall, and Precision. Based on a component from a matrix known as the confusion matrix or contingency table [17], these metrics are useful for evaluating the effectiveness of supervised machine learning algorithms. A confusion matrix is commonly used to visualise the performance of a method. Classification phrases such as 'True Positive (TP)', 'False Positive (FP)', 'True Negative (TN)', and 'False Negative (FP)' are used to compare class labels in this matrix, as shown in Below Table 1. True Positive reflects favourable feedback that the classifier categorised as positive, whereas False Positive is predicted to be unfavourable but is classified as negative. True negative, on the other hand, reflects unfavourable evaluations that the classifier categorised as negative, whereas False Negative is projected as positive but actually classed as negative. Precision, recall, f-measure, and accuracy are used to evaluate classification performance based on confusion matrix.

Predicted Values

Actual values

|  |  |  |
| --- | --- | --- |
|  | Positive | Negative |
| Positive | True Positive (TP) | False Positive (FP) |
| Negative | False Negative (FN) | True Negative (TN) |

Table 1 Confusion Matrix

1. **Precision**

The following is described as the ratio of the number of positive reviews accurately categorised to the total number of favourably classified reviews.

1. **Recall**

This is described as the proportion of positive feedbacks accurately categorised to the total number of positive reviews.

1. **Accuracy**

This is the proportion of correctly categorised reviews to the total number of reviews.

1. **F1-Score**

This is a composite accuracy and recall metric.

**DATA COLLECTION & ANALYSIS:**

Feedback from consumers for a wide range of products and services have recently been more widely available. Reviews from consumers generally consist of two parts: star ratings and review content. In this study, we will utilise customer reviews and ratings to categorise the review as Positive, Negative, or Neutral.

**Data Gathering and Interpretation:**

For the sentiment analysis, an Amazon dataset extracted using Prompt Cloud is used. The dataset pertains to unlocked mobile phones and was obtained in December 2016. It is a Kaggle.com dataset that is freely available to the public. The Amazon reviews dataset contains over 400,000 consumer evaluations in the mobile phone category. It specifically covers 413,840 reviews and 6 attributes, which are categorised as follows:

1. Mobile phone information (Brand Name, Model Name, Price, Rating).
2. Review info (Reviews and Review Votes).

Without a doubt, data cleansing is the act of going through all of the data within a data frame and either removing or updating any incomplete, erroneous, or redundant information. This stage is important for maximising data accuracy since data quality is critical for achieving the desired outcome with high efficiency and precision. The Amazon mobile phone dataset has missing values in various characteristics, as shown in below Table 2, and some procedures have been made to clean the data.

|  |  |  |  |
| --- | --- | --- | --- |
| Features | Description | Data type | Missing value |
| Product Name | A company's name for each type of mobile phone. | Object | 0 |
| Brand Name | The name of the manufacturer | Object | 65,171 |
| Price | The price of a mobile phone | Float | 5,933 |
| Rating | Rating [1-5] | Integer | 0 |
| Reviews | Consumer’s feedback | Object | 62 |
| Reviews Votes | The number of consumers that participated in the review voting. | Float | 12,296 |

Table 2: Summary of Amazon mobile phone data

Firstly, items in "Reviews" having a null value (62) were removed. Secondly, in "Review Votes" (12,296), all null values were replaced with 0. Because these are real reviews that have gotten no votes, i.e. they have earned 0 votes, zero is the most acceptable rating. Third, all null values in "Price" were replaced with 144.71, which is the cost's median value (12,296 null values). Furthermore, duplication in the data was handled by eliminating all duplicated records (64,079 records). The brand names of mobile devices were removed from the "Brand Name" field. Furthermore, several trademark names with extra quotes were eliminated. Many brand names include spelling errors and are not standardised; in these cases, we normalised the brand name by inserting the right names and unifying them. To demonstrate this process, suppose "New Nextel Rugged Motorola I680 Cell Phone" was changed by "Motorola " as well as "motorola" was replaced by "Motorola". Following this phase, the overall number of brand names decreased from 382 to 282.

**Exploratory Data Analysis:**

Exploratory Data Analysis (EDA) is a technique for visualising and analysing data buried in rows and columns. Furthermore, by utilising a variety of visualisations, it allows us to gain the most insight into the dataset. We looked at the Amazon mobile phone rating distribution by star rating as well as the number of reviews. The statistic shows that the five-star rating is the most common, while the two-star rating is the least common. Below Figure 4 reveals, on the other side, that 52.7% of customers awarded 5 stars, 15.1% provided 4 stars, 7.9% got 3 stars, 6.2% got 2 stars, and 18.1% gave 1 star.

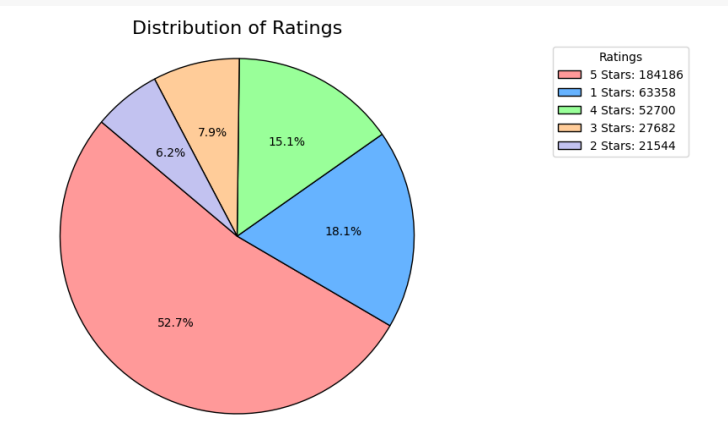


Figure 4 Distribution of Rating

The polarity of the Amazon mobile phone dataset is then investigated and visualised using a bar chart. A new column with three values is Review Sentiment. Reviews rated 4 or 5 are deemed "Positive," whereas 1 or 2 are considered "Negative," and 3 are considered "Neutral". According to Figure 5 , 67.78% of the evaluations are "Positive," 24.29% are "Negative," and 7.92% are "Neutral."

We also looked at the amount of reviews and the brand name. It shows that Samsung received the most reviews (58,439), while BLU, Apple, LG and BlackBerry earned (51,780) , (50,077), (22,131) and (14,973) reviews, respectively.

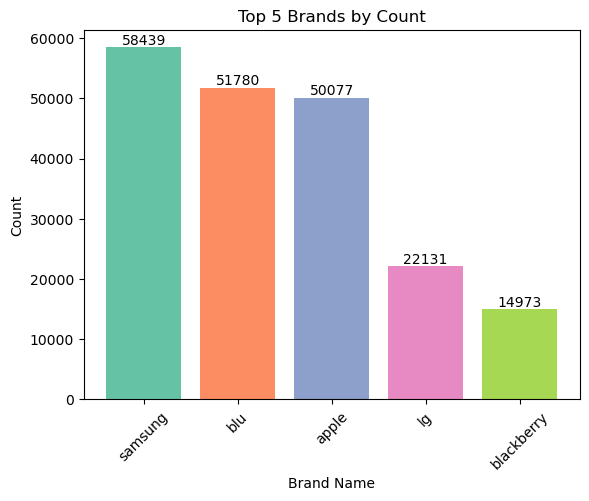
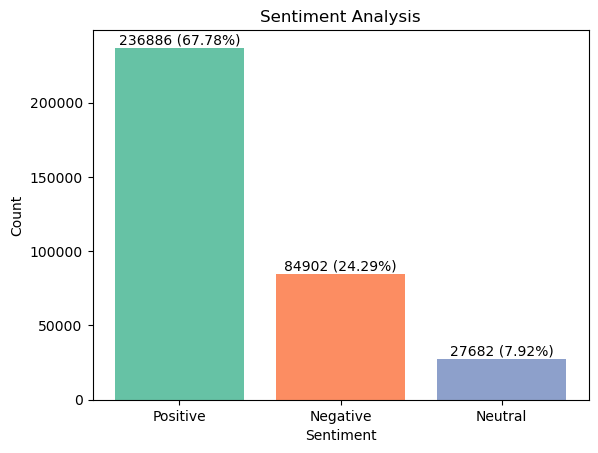


Figure 5 Sentiment Analysis and Top 5 Brands by Count

**RESULTS AND DISCUSSION:**

During this endeavour, Python and its Jupyter-lab, were utilised alongside other supporting libraries to do data purification, visualisation, pre-processing, and machine learning modelling.

Furthermore, for deep learning, the CUDA and cuDNN libraries were employed for quicker implementation. The findings of the proposed models: LR, RF, NB, Linear SVM, and BERT are shown in this section. Various assessment measures were employed to evaluate the models explored here, including Accuracy, Precision, Recall, and F-score, which are all explained in below section. Initially we'll go over the results for all models using the validation dataset and several feature extraction approaches, including BOW with N-gram, TF-IDF, and Glove. Then, using the test dataset, we will depict the ultimate outcome for each classification model. The trials began with a multiclass classification using all of the proposed models. Furthermore, they were carried out for a binary classification using the model having the best validation accuracy.

**Multiclass classification:**

For the present research, Amazon mobile phone reviews were categorised as positive, negative, or neutral based on star rating, with one and two stars represented negative, four and five stars regarded positive, and three stars as neutral. As a result, all recommended models were used in conjunction with every feature extraction methods.

The Logistic Regression model was used in conjunction with BOW, TF-IDF, and GloVe. Table 3 displays the Logistic Regression findings for all of these techniques. The Logistic Regression with BOW (Bigram) obtained great accuracy. Furthermore, the Recall, Precision and F1 score of the Logistic Regression using BOW (Bigram) is higher than that for various feature extraction approaches.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Logistic Regression | Accuracy | Precision | Recall | F1-score |
| BOW UNIGRAM | 0.8572 | 0.8391 | 0.8572 | 0.841 |
| BOW BIGRAM | **0.8938** | 0.8914 | 0.8938 | 0.8867 |
| BOW TRIGRAM | 0.8527 | 0.8687 | 0.8523 | 0.8379 |
| TF-IDF | 0.8703 | 0.8539 | 0.8703 | 0.8526 |
| GloVe | 0.7997 | 0.7581 | 0.7998 | 0.7665 |

Table 3 The Results of Logistic Regression Model

Below Table 4 also shows the assessment outcomes of the Random Forest Regression model using various feature extraction methodologies. When compared to previous techniques, the Random Forest Regression with TF-IDF has 92.31% greater accuracy. The reasons for this strong performance could include a well-chosen model, appropriate hyperparameters, and a clean and informative dataset.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Random Forest Regression | Accuracy | Precision | Recall | F1-score |
| BOW UNIGRAM | 0.918 | 0.9183 | 0.918 | 0.9125 |
| BOW BIGRAM | 0.8946 | 0.8932 | 0.8946 | 0.8893 |
| BOW TRIGRAM | 0.8573 | 0.8641 | 0.8574 | 0.8464 |
| TF-IDF | **0.9231** | 0.9245 | 0.9231 | 0.9175 |
| GloVe | 0.8988 | 0.9016 | 0.8988 | 0.8929 |

Table 4 The Result of Random Forest Regression Model

Under the counter Table 5 provides a comparative view of the Linear SVM Regression model's performance with different text representations. It highlights that the choice of text representation can significantly impact the model's accuracy and other performance metrics, with TF-IDF and BOW unigrams generally leading to the best results, while GloVe embeddings result in slightly lower performance.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Linear SVM Regression | Accuracy | Precision | Recall | F1-score |
| BOW UNIGRAM | 0.8781 | 0.8768 | 0.8781 | 0.8622 |
| BOW BIGRAM | 0.8678 | 0.8741 | 0.8678 | 0.8538 |
| BOW TRIGRAM | 0.8143 | 0.8467 | 0.8143 | 0.7878 |
| TF-IDF | **0.8822** | 0.8751 | 0.8822 | 0.8671 |
| GloVe | 0.8053 | 0.8165 | 0.8053 | 0.7689 |

Table 5 The Results of Linear SVM Regression Model

The following Table 6 shows the findings of the Neive Bayes Regression classifier assessment using every suggested feature extraction technology in this research except GloVe. GloVe is a word vector format that groups with related words, although the Neive Bayes Regression concept maintains that characteristics are independent. When contrasted with each of the other techniques, the Neive Bayes Regression classifier using BOW (Bigram) has greater accuracy with 88.36%.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Neive Bayes Regression | Accuracy | Precision | Recall | F1-score |
| BOW UNIGRAM | 0.8443 | 0.8283 | 0.8443 | 0.8221 |
| BOW BIGRAM | **0.8836** | 0.8878 | 0.8836 | 0.8701 |
| BOW TRIGRAM | 0.8729 | 0.8803 | 0.8729 | 0.8622 |
| TF-IDF | 0.8162 | 0.8162 | 0.8162 | 0.7755 |
| GloVe | **-** | **-** | **-** | **-** |

Table 6 The Result of Neive Bayes Regression Model

BERT results are shown in below Table 7. It obtained 88% accuracy and a Precision, recall as well as F1-score as 0.8587, 0.8799 and 0.8665 respectively, indicating that the model performed well. The techniques also has a good accuracy and recall.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Accuracy | Precision | Recall | F1-score |
| BERT | **0.8799** | 0.8587 | 0.8799 | 0.8665 |

Table 7 The Result of BERT Model

**Final Evaluation:**

Here below Table 8, we present the accuracy results of various regression models applied to different feature representations. The Random Forest Regression model, utilizing the TF-IDF feature representation, achieved the highest accuracy at 92.31%, demonstrating its superior predictive performance. The Bag of Words (BIGRAM) representation with Logistic Regression closely followed with an accuracy of 89.38%, while Linear SVM Regression using TF-IDF achieved an accuracy of 88.22%. The Naive Bayes Regression model with BIGRAM representation performed at 88.36% accuracy. Lastly, we tested the BERT-based model, which achieved an accuracy of 87.99%. These findings provide valuable insights into the comparative performance of these models and feature representations, aiding in informed decision-making for these applications.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Models | Accuracy | Precision | Recall | F1-score |
| Logistic Regression BOW BIGRAM | 0.8938 | 0.8914 | 0.8938 | 0.8867 |
| Random Forest Regression TF-IDF | 0.9231 | 0.9245 | 0.9231 | 0.9175 |
| Linear SVM Regression TF-IDF | 0.8822 | 0.8751 | 0.8822 | 0.8671 |
| Neive Bayes Regression BOW BIGRAM | 0.8836 | 0.8878 | 0.8836 | 0.8701 |
| BERT | 0.8799 | 0.8587 | 0.8799 | 0.8665 |

Table 8 Final Classification Model Results

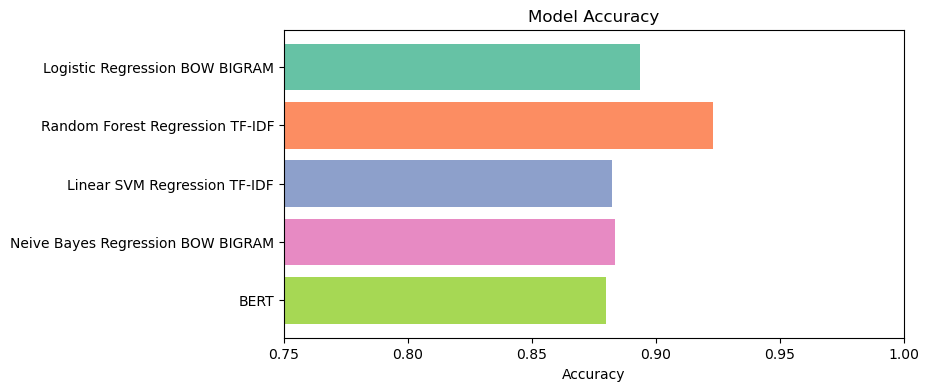


Figure 6 Final Classification Model Results

**Analyse the Performance of Random Forest Regression model:**

The accuracy of the Random Forest Regression model is 94%. As a result, we wish to study its feelings categorization ability. The current emphasis is on understanding models and checking incorrect classifications. We need to understand what causes models to classify incorrectly in certain reviews while accurately recognising in others. As a result, we sought misclassified reviews with the largest loss values. In the exact same manner that we understood them, we used them to clarify the Random Forest Regression model's decision boundaries. We can see from Figure 7 that the Random Forest Regression model categorised the review as satisfactory. While the review's real label was impartial. This is a prime instance of a case where the Random Forest Regression model is right but the tag is incorrect. The consumer awarded the product three stars, clearly indicating his overall happiness with it. That is the reason why as we stated in the outset, three-star ratings do not necessarily convey a neutral mood.

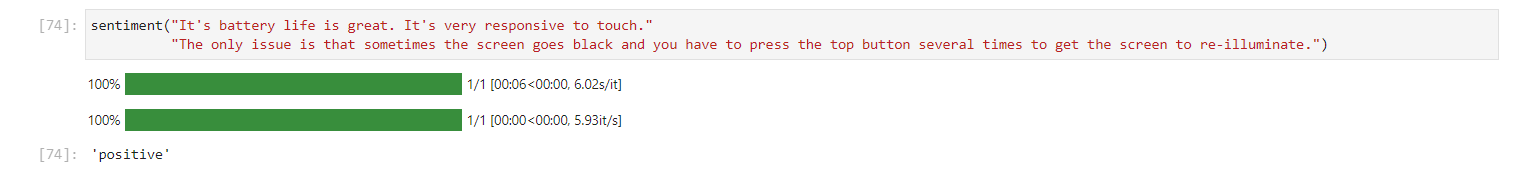


Figure 7 A case of a consumer misclassifying a review.

Furthermore, as seen in below Figure 8, the consumer conveys his unhappiness with the cell phone. In reality, the model accurately labels the review as negative, although the actual class was incorrectly assigned as positive.

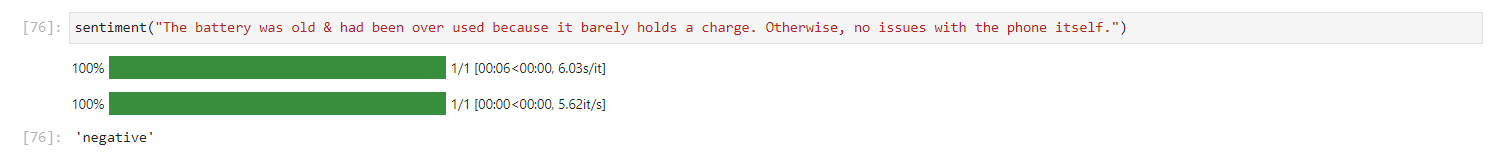


Figure 8 A scenario of an incorrect review label

**Discussion:**

The following part examines the outcomes of our models and compares them to some of the common models used for the same objective, namely sentiment analysis on Amazon mobile phone reviews. Some models utilised in this experiment, such as BERT, Linear SVM, have never been employed with the Amazon mobile reviews dataset; hence, comparing current research will not cover all angles. The investigators of study [4] used Logistic Regression, Neive Bayes, Support Vector Machine , and Random Forest in a binary classification problem. Researchers achieved an excellent accuracy of 90% with Random Forest and word2vec as feature extraction, whereas Random Forest plus GloVe achieved 89.99% in our trial. In addition, the researchers of [5] used four different approaches in multiclass classification: Logistic Regression, Nave Bayes, Stochastic Gradient Descent, and Convolutional Neural Network (CNN). The CNN model using word2vec for feature extraction got the highest accuracy of 92% on multiclass classification, whereas the model we used with BERT came to 87.99%.

**Conclusion:**

Analysis of sentiment is an unavoidable and widely used method for collecting information from text data in online stores. Every day, online shopping sites generate vast amounts of textual information in the manner of ideas, feedback, tweets, and comments. Furthermore, reviews, ratings, and emoticons imply the views of individuals. Gathering details regarding an item from a review will assist a consumer in learning more about the item and making a choice. The study investigates multiclass classification for Amazon mobile phones utilising supervised machine learning techniques like as Logistic Regression, Nave Bayes, Random Forest, and other feature extraction methodologies. Furthermore, GloVe embedding were used in this study. Furthermore, the Bidirectional Encoder Representations from Transformers (BERT) model was used. The BERT model performed well in multiclass classification, with accuracy of 87.99%, respectively. Random Forest with word embedding (TF-IDF) beats two baseline models, Logistic Regression and Neive Bayes, with multiclass classification accuracy of 92.31% .

**Limitations and Suggestions for Future Work:**

This work depicts the implementation of many machines learning models, including Logistic Regression, Neive Bayes, Random Forest, and Support Vector Machine, using various feature extraction methodologies for a text categorization job. Furthermore, we applied the pre-trained BERT model to the sentiment analysis job on the Amazon mobile phone reviews dataset and fine-tuned it. We want to employ word2vec in the future to extract features from our models and identify bogus reviews. Aside from the classifiers previously discussed, others such as Gated Recurrent Unit (GRU) and Bi-LSTM might be utilised. Furthermore, the data set is exclusive to Amazon mobile reviews; nevertheless, it may be expanded to the examination of Amazon reviews in general. I utilized CUDA and cuDNN to enhance computational power; however, Google Colab remains a valuable option for accelerating the study for the limited capabilities of other laptop.

# References [14]

|  |  |
| --- | --- |
| [1] | S. A. a. A. N. S. Aljuhani, "International Journal of Advanced Computer Science and Applications," *A Comparison of Sentiment Analysis Methods on Amazon Reviews of Mobile Phones,* vol. 10, 2019. |
| [2] | K. S. a. D. J. a. M. J. Kumar, "IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)," *Opinion mining and sentiment analysis on online customer review,* 2016. |
| [3] | A. S. a. A. A. a. D. P. Rathor, "Procedia computer science," *Comparative study of machine learning approaches for Amazon reviews,* vol. 132, 2018. |
| [4] | B. a. S. S. Bansal, "Procedia computer science," *“Sentiment classification of online consumer reviews using word vector representations,* vol. 132, pp. 1147--1153, 2018. |
| [5] | A. a. S. V. a. M. B. ernian, "Computers and Artificial Intelligence," in *Sentiment analysis from product reviews using SentiWordNet as lexical resource*, Bucharest, 2015. |
| [6] | J. F. V. W. P. G. N. R. Moraes, " Expert Systems with Applications," *“Document-level sentiment classification: An empirical comparison between SVM and ANN,* vol. 40, pp. 621--633, 2013. |
| [7] | D. a. X. H. a. S. Z. a. X. Y. Zhang, "“Chinese comments sentiment classification based on word2vec and SVMperf," *Expert Systems with Applications,* vol. 42, pp. 857--1863, 2015. |
| [8] | Y. a. L. M. a. E. K. K. E. A. Amrani, "“Random Forest and Support Vector Machine based Hybrid Random Forest and Support Vector Machine based Hybrid," *Procedia Computer Science,* pp. 511-520, 2018. |
| [9] | Y. a. K. V. Saito, "Classifying User Reviews at Sentence and Review Levels Utilizing Naïve Bayes," in *PyeongChang Kwangwoon Do*, Korea (South), 2019. |
| [10] | Y. a. L. M. a. E. K. K. E. A. Amrani, "Random Forest and Support Vector Machine based Hybrid Approach to Sentiment Analysis," *Procedia Computer Science,* pp. 511-520, 2018. |
| [11] | Bahrawi, "Sentiment Analysis Using Random Forest Algorithm-Online Social Media Based.," *JOURNAL OF INFORMATION TECHNOLOGY AND ITS UTILIZATION,* vol. 2, pp. 29-33, 2019. |
| [12] | M. a. S. R. Fikri, "A Comparative Study of Sentiment Analysis using SVM and SentiWordNet," *Indonesian Journal of Electrical Engineering and Computer Science,* vol. 13, pp. 902-909, 2019. |
| [13] | H. a. K. A. Sinha, "A Detailed Survey and Comparative Study of sentiment analysis algorithms," in *2016 2nd International Conference on Communication Control and Intelligent Systems (CCIS)*, Mathura, India, 2016. |
| [14] | R. S. C. D. M. J. Pennington, "GloVe: Global Vectors for Word Representation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar, 2014. |
| [15] | N. T. a. Milievi, "Comparing sentiment analysis and document representation methods of Amazon reviews," *2018 IEEE 16th International Symposium on Intelligent Systems and Informatics,* pp. 000283--000286, 2018. |
| [16] | K. a. M. W. a. C. W. Ogada, "N-gram Based Text Categorization Method for Improved Data Mining," *Journal of Information Engineering and Applications,* vol. 5, pp. 35--43, 2015. |
| [17] | Y. a. A. G. a. J. P. a. K. T. Sharma, "Vector representation of words for sentiment analysis using GloVe," in *n 2017 international conference on intelligent communication and computational techniques*, Jaipur, 2017. |
| [18] | A. H. C. H. H. B. G. M. Naili, "“Comparative study of word embedding methods in topic segmentation," *Procedia Computer Science,* vol. 112, pp. 340-349, 2017. |