An Efficient Machine Learning-Based Sentiment Analysis for Amazon Product Reviews

K.Sravana Kumari[1], Dr. B. Manjula[2]
[1]Research scholar, Department of Computer Science,KakatiyaUniversity, Warangal, TS, India
[2] Professor in Department of Computer Science,Kakatiya University, Warangal, TS, India

Abstract: Web portals such as Amazon collect a large quantity of consumer input regularly. It might be a time-consuming and tedious task to read through all of the criticism. They must classify people's thoughts as they are presented in feedback forums. A feedback management system is an example of one application that could benefit from this. They classify each comment and review according to one of these classes and use these classes to inform our purpose of an overall rating for the product. It is necessary for the company to have a comprehensive picture of the feedback provided by customers and to focus its attention on the appropriate areas. This results in an increase in the number of customers who are loyal to the company, as well as a growth in business, reputation, and the value of the brand, as well as profits. As a result, they suggest that we use the aspect terms, also referred to as the targets from the texts, to extract the users' feelings regarding particular qualities associated with the products. This study built a model to predict the comment's sentiment based on the comment declaration using Python and Support vector machine (SVM), random forest (RF), and Improved logistic regression (ILR) are the three different machine learning techniques. The dataset used in this study was compiled from customer reviews of musical instruments sold on Amazon.com. SMOTE is utilized so that the unbalanced dataset can be managed, and AUC and ROC are used so that the optimal approach can be determined. For the review text's classification, the solution they have given is based on a logistic regression that uses the Kernel Density Estimation approach. Classification evaluation indicators like F1-score, precision, recall, and accuracy are utilized in the evaluation process.

Index Terms: Sentiment analysis, Text classification, kernel function, logistic regression
DOI Number: 10.48047/NQ.2022.20.20.NQ109170 **NeuroQuantology2022;20(20): 1661-1668**

1. INTRODUCTION

E-commerce and modern logistics are advancing at a breakneck pace these days. One of the defining characteristics of this age of the Internet is the growing trend of people shifting their buying preferences toward acquiring goods through e-commerce platforms [1-3]. Shopping online offers a substantial improvement in both convenience and productivity. It is currently well-liked among people of all ages. Nevertheless, due to the nature of online buying, which only enables the perusal of virtual goods because it cannot physically inspect the product before purchase, many issues are caused by inconsistencies between the item description and the thing itself. Customers rely heavily on product reviews as a reference when making purchases to ensure that the item matches the description provided. As a direct consequence, product reviews ought to be an integral component of the evaluation process for commodities. The evaluations, for instance, are amenable to being analyzed using sentiment analysis [4, 5].

Quantifiable research like this might benefit both the consumers and the producers. Web portals like Amazon gather a large amount of feedback from their users. In recent years, most of the effort in sentiment analysis has been to develop more accurate classification algorithms and constantly tackle substantial obstacles and constraints in this field. Natural language processing (NLP) did little research on sentiments until 2000 when humans began to have a vast volume of online text [6]. Natural language processing (NLP) provides the required technology for text mining to automatically extract knowledge from these documents [7]. The objective of natural language processing, also known as NLP, is to create a language that will enable computers to comprehend unstructured text and assist them in doing so. One of the Python libraries that are utilized the most, NLTK, stands for Natural Language Toolkit, and it is used in the field of natural language processing for Python.

It might be a time-consuming and tedious task to read through all of the criticism. They must classify people's thoughts as they are presented in feedback forums. One possible application for this is a feedback management system. It is possible to categorize individual comments and reviews and assign an overall rating based on those individual comments and reviews. It is necessary for the company to have a comprehensive picture of the feedback supplied by consumers and to focus its attention on the appropriate areas. This results in an increase in the number of customers who are loyal to the company, as well as a growth in business, fame, and the value of the brand, as well as profits. In the most common use of the currently known regression technique, support vector machines solve problems involving more than one categorization [8, 9], but this algorithm has some drawbacks. The issue of dichotomy and linear classification is the only kind of problem that the logistic regression algorithm can solve. In most cases, support vector machines can only handle limited training data, and they struggle mightily when confronted with various classification issues. The accuracy of the classification will be significantly impacted as soon as the dataset can no longer satisfy this assumption. [10] developed a model of uncertainty to constraint conditions to overcome the problem described above, which is characterized by the fact that it is challenging to that it is difficult to forecast, that it is impractical to implement large-scale samples, and that it is irrelevant to multi-classification.(DLR) is an acronym for density-based logistic regression, a statistical method that performs well in practical applications. Kernel density-based logistic regression is the foundation of our model. In addition, they developed an original kernel function with the express purpose of applying it to multi-classification problems.

eISSN 1303-5150

This article uses the "Amazon Musical Instruments Reviews" dataset, which can be found on the website Kaggle. Its purpose is to conduct sentiment analysis on the recommendations using Python as its platform that customers have left for various musical instruments sold on Amazon. They will classify positive and negative feedback, depict it using a word cloud, and then use SVM, RF, and an improved logistic regression model to predict ratings based on reviews. Finally, they will evaluate the accuracy of various machine-learning techniques.

The remainder of the paper is composed of the following structure shown below. In the next section, they will conduct a brief evaluation of the relevant prior research. The following section, "Section 3, demonstrates our approach to improving the framework for sentiment analysis. The results of the tests are discussed in Section 4, and they show that our technique is at the lead of what is considered modern methodology. The Section 5 of the paper serves as the paper's conclusion.

2. RELATED WORKS

In this section we discussed some past state-of-the-art works that use machine learning for sentiment analysis. This type of research aims to find out how people think, feel, evaluate, and feel about things and their characteristics from the text. These can be a wide range of products, services, institutions, people, events, situations, or topics.

Using the MapReduce processing framework, the authors of [11] proposed a scalable approach for conducting Twitter sentiment analysis. To achieve this goal, a MapReduce-based Naive Bayes training method has been presented. Unlike other works, this technique requires only a single MapReduce job, which distinguishes it from those other works. The trained model is currently being used to classify many tweets. Experiments are carried out with the help of an Amazon EMR cluster, and the outcomes demonstrate that the suggested strategy is highly scalable and cost-effective for larger data sets. It is 1.5–1.7 times faster than one of the prior studies and 2.7–3.4 times faster than the Naive Bayes Classifier that is available through Apache Mahout. To evaluate the framework's scalability, a substantial training corpus has been acquired from Twitter. When trained with this new dataset, the accuracy of the classifier came close to reaching 79%.

The data collected from social media is filtered, stored, and with technology that utilizes natural language processing, a sentiment analysis was performed on the content. The authors of [12] constructed a sentiment analysis model to improve sentiment analysis performance through the utilization of various portions of speech, in addition to reducing the dimensionality of the data. Using the machine learning techniques of Naive Bayes, Support Vector Machines, and K-Nearest Neighbor, the model's performance is compared with that of two more sentiment analysis models. The results of the tests show that the model can improve how well sentiment analysis works by using techniques from machine learning.

In [13] authors realized learning using the data sets gathered from interpretations made on the social platforms of the determined brands and to communicate the topic of emotion analysis to researchers in the most effective way possible. Because of the disadvantages, such as not paying attention to the norms of writing on social media or other digital platforms, the range of accuracy rates that can be achieved is quite broad. In their investigation, an accuracy rate of 70% was reached. This highlights the usefulness of machine learning in interpretation, classification, and emotion analysis.

In [14], the findings were analyzed and summarised by the authors using the recommended reporting components for systematic reviews and meta-analyses (PRISMA) format. To make this resource available to researchers and academics, the objective is to compile a database of the methods and approaches utilized in the investigations. Before moving on, they proceeded to analyze the implications, trends, and problems that the corpus offered—the suggested directions for further research in applying SA in creating and analyzing social media campaigns.

Instead of having the user supply the subjective text described in [15], the authors introduced a system that could extract a neutral written description of the photographs generated automatically from the visual information and the utilization of such a report. The proposed method would extract three views of an image shared on social media: a visual view, a subjective text view, and an objective text view. Based on the hypothesis table, the system will give a sentiment polarity that is either neutral, negative, or positive. Regarding the remarks, Naive Bayes. This method analyses the comments to determine the Sentiment.

In [16], the authors introduced a hybrid method that predicts the Sentiment using the lexical Sentiment VADER and the machine learning algorithm should both be approached. The Naive Bayes classifier has a significant impact on determining an accurate prediction of the comment's attitude. The classifier was found to have an accuracy of 79.78% and an F1-score of 83.72%.

The authors of the paper proposed a method of transfer learning based on the multi-layer convolutional neural network [17]. It is interesting to note that to extract features from the source domaincreated a convolutional neural network model. The weights in the convolutional layer and the pooling layer are then distributed between the samples of the source domain and the target domain samples. This step completes the deep learning process. Next, they go through the final layer, which added the fully connected layer, and make minor adjustments to the weights. After that, In this step, the models are transferred from the source domain to the destination domain. The method recommended the research does not require the network to be retrained for the environment it will be used to, which contrasts with the typical transfer learning methods currently in use. The empirical analysis of the data set, which contains data from many different domains, shows that the proposed strategy works well.

3. PROPOSED WORK

The proposed approach is a fast, efficient, and difficult-to-overfit classification method, particularly for high-dimensional data. Nevertheless, it can only deliver label categorization. The performance of the SVM algorithm with a linear kernel function can offer a hyperplane representing malware detection. However, the efficacy of this method is contingent on the accuracy of feature selection. The Logistic Regression model is the most accurate, as related to Support Vector Machine (SVM). This article aims to conduct a

eISSN 1303-5150

sentiment analysis of reviews based on products. The data presented in this article is from online product reviews gathered from the website "amazon.com." Figure 1 shows the strategy to use review data that shows promising results to do review-level classification.

Dataset description:

The name of the dataset that they utilize is the Amazon Musical Instruments Reviews database. The data for this specific dataset was obtained from the Amazon website and includes 10,261 individual reviews of various products. The dataset contains suitable text and many other elements that can be used for analysis.

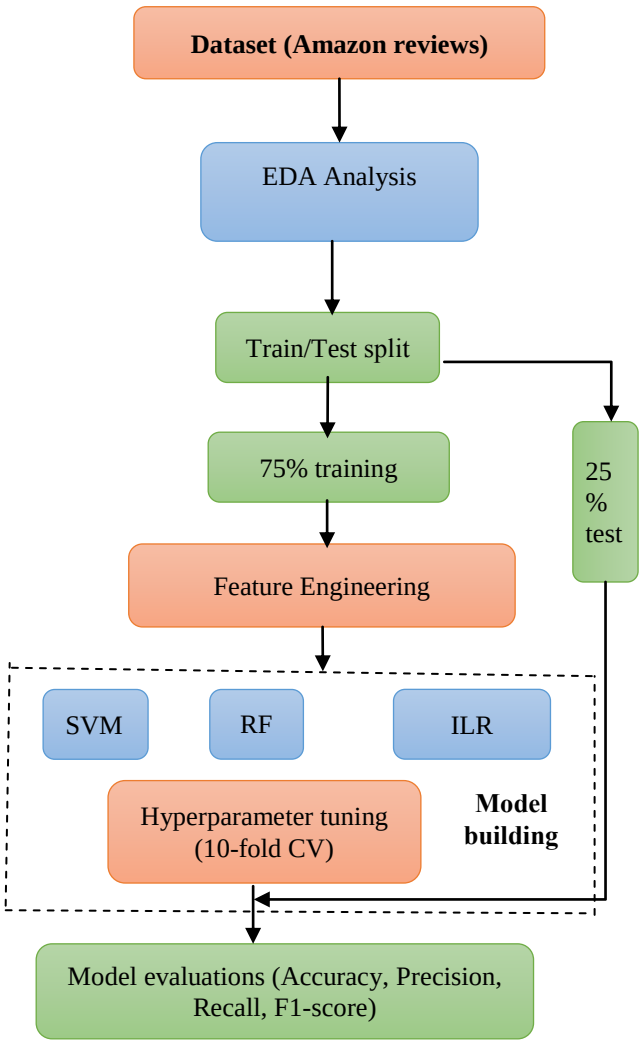


Figure 1: System of Sentiment Analysis framework

The dataset contains multiple features concerning user reviews of musical instruments. However, they only irregularly require those qualities as model variables since sentiment analysis does not emphasize those features significantly. They may need to ignore our part of eliminating stopwords in our preprocessing phase because there are important words in determining user attitudes in our model. According to our text analysis, most of the deals that have been struck concern guitars or other string-based instruments. They can claim that the guitar learnt great attention from the buyers' pool, and the merchants can focus their products on this instrument.

This document contains the reviewer ID, the user ID, the reviewer's name, the reviewer text, a summary that was derived from the reviewer text, an overall rating out of five stars, and the review time.

- reviewerID — ID of the reviewer, e.g. A2SUAM1J3GNN3B
- asin — ID of the product, e.g., 0000013714
- reviewerName — name of the reviewer
- helpful — helpfulness rating of the review, e.g., 2/3
- reviewText — text of the review
- overall — rating of the product
- summary — summary of the review
- unixReviewTime — time of the review (unix time)
- reviewTime — time of the review (raw).

A. Data Preprocessing

As an illustration of how our data is preprocessed, they will utilize the Amazon Musical Instruments Reviews as a case study in this article. First, they used the columns in the dataset labeled "review Text" and "overall" as the objectives for our analysis. After that, they get rid of the anagram sign,

eISSN 1303-5150

change all of the capital letters to lowercase, and then use "stopwords" at API NLTK of Python to eliminate meaningless words like conjunctions and pronouns.

B. Feature Engineering

The next step is to obtain the aspect-term set by selecting the top 160 frequency nouns. Word frequency is calculated for every line using the TFIDF vectorizer included in the scikit-learn package for Python [27], and then They sort the results in descending order. The aspect-term set contains no words that were derived from any other source. As a result, they can avoid the human labor associated with labeling the item.

TF-IDF Vectorizer:
From a shape, we successfully transformed our reviews with a TF-IDF Vectorizer of 7000 top bigram words. Now, as they know from before, our data is imbalanced with minimal neutral and negative values compared to positive sentiments. They need to balance our dataset before going into the modeling process.

Resampling Dataset:
There are many ways to resampling to an imbalanced dataset, such as SMOTE and Bootstrap methods. They will use SMOTE (Synthetic Minority Oversampling Technique) to randomly generate new replicates of our undersampling data to balance our dataset. The data is already balanced, as they can see from the counter of each sentiment class before and after the resampling with SMOTE.

Splitting Dataset:
We split our dataset into 75:25 portions for the training and test set.

Model Selection and Evaluation
We need to find the best model that fits our data well. First, they should do cross-validation techniques to find the best model. Because of that, they will need to try every classification model available and find the best models using the Confusion Matrix and F1 Score as our main metrics and the rest of the metrics as our support.

Random forest. It was determined that the random forest classifier would provide the highest level of accuracy compared to a single decision tree. Hence that is the method that will be used. In its most basic form, it is an ensemble method that relies on bagging. The following is how the classifier works: The classifier begins by generating *k* samples of the data set *D* using the bootstrap method, with each model denoted as D_i. There are the same number of tuples in a D_i as there are in a *D*, and these tuples are sampled with replacement from *D*. It is possible that some of the original tuples of *D* will not be included in *D*i, while others may appear more than once due to the sampling method known as sampling with replacement. After that, the classifier constructed a decision tree based on each D_i. Consequently, a "forest" of *k* different decision trees is produced.

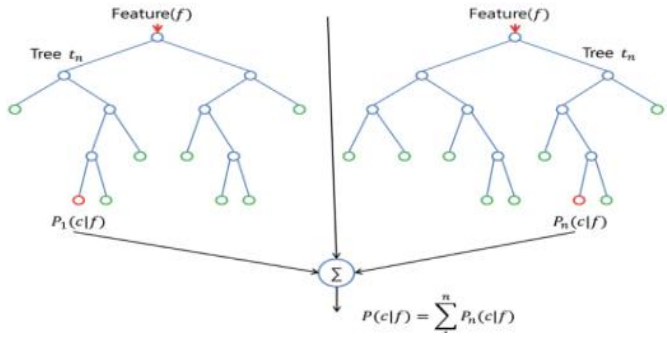


Figure 2: Tree structure of Random Forest model

CART is the name of the decision tree algorithm that is implemented in scikit-learn (Classification and Regression Trees). To determine the category of an unknown tuple, X, each tree contributes one vote consisting of its class prediction. The student in X's class who finishes in first place in the vote count gets to make the ultimate choice. The Gini index is utilized for CART's tree induction process. Calculating the Gini index for D looks like this:

$$gini(D) = 1 - \sum_{i=1}^{m} p_i^2 \quad (1)$$

Where p_i represents the likelihood that a tuple in *D* is a member of class C_i, the Gini index is used to determine the degree of *D*'s impureness. When the index value is smaller, this indicates that *D* was partitioned more effectively.

Support vector machine. The support vector machine, sometimes known as SVM, is a technique for classifying data that might be linear or nonlinear. If the data can be separated linearly, the support vector machine will look for the linear optimal separating hyperplane, also known as the linear kernel. This is a decision boundary that divides data belonging to one class from data belonging to another class. A separating hyperplane can be described using the following mathematical equation: $WX+b=0$, where *W* is a weight vector and $W=w_1,w_2,...,w_n$, and *X* is a training tuple. The value *b* is a scalar. To maximize the efficiency of the hyperplane, the problem effectively reframes itself as one of minimizing "*W*," which is ultimately computed as follows:

$$\sum_{i=1}^{n} \alpha_i y_i x_i \quad (2)$$

where α_i are a set of numeric parameters, and the labels, yi, are determined by the support vectors, Xi.

$$\sum_{i=1}^{n} w_i x_i \geq 1 \quad (3)$$

If y_i= -1 then

$$\sum_{i=1}^{n} w_i x_i \geq -1 \quad (4)$$

If the data cannot be separated linearly, the SVM will use nonlinear mapping to increase the dimension of the data. After that, it finds a linear hyperplane and uses that to solve the problem. Kernel functions are the functions that are used to conduct transformations such as these. For this

eISSN 1303-5150

experiment, the kernel function was chosen as the Gaussian Radial Basis Function (RBF).

$$K(X_i, X_j) = e^{-\gamma \|X_i - X_j\|^2/2} \quad (5)$$

here X_i is support vectors, X_j are testing tuples and is a free parameter that, in our experiment, utilizes the default value that scikit-learn provides. The next page's figure shows a classification example using the SVM based on the linear kernel and the RBF kernel.

Improved Logistic Regression:

We improve upon the existing multi-classification technique and then expand the DLR model so that it can tackle the multi-classification problem. Assuming there are *C* classes, the DLR model is defined as follows, for *k* = 1, 2,....., *C*:

$$p(y = k|x) = \frac{e^{w_k^T \phi_k(x)}}{\sum_{j=1}^{C}\left(e^{w_j^T \phi_j(x)}\right)} \quad (6)$$

$w_k = (w_{k1}, w_{k2}, \ldots, w_{kD})$ is the class k's characteristic weighting parameter, and $\phi_k = (\phi_{k1}, \phi_{k2}, \ldots \phi_{kD})$ is the class *k*'s characteristic transformation function.

$$\phi_{kd}(x) = \ln p(y = k|x_d) - \frac{D-1}{D} \ln p(y = k) \quad (7)$$

The probability formula for class *k* can be calculated in the following manner, in accordance with the Nadaraya-Watson estimator:

$$\mathrm{p}(y = k|x_d) = ln \frac{\sum_{i \in D_k} e^{-\frac{(x_d - x_{id})^2}{h_d^2}}}{\sum_{i=1}^{N} e^{-\frac{(x_d - x_{id})^2}{h_d^2}}} \quad (8)$$

Finally,Theyneed to make the loss function as little as possible.

$$E(w,h) = -\sum_{i=1}^{N}\sum_{i=1}^{N}\left(1_{yi=k} \ln p(y_i = k|x_i)\right) \quad (9)$$

$$= -\sum_{i=1}^{N}\sum_{i=1}^{N}\left(1_{yi=k} \ln \frac{e^{w_k^T \phi_k(x_i)}}{\sum_{j=1}^{C}\left(e^{w_j^T \phi_j(x_i)}\right)}\right)$$

$$-\sum_{i=1}^{N}\sum_{i=1}^{N}\left(1_{yi=k}\left(w_k^T \phi_k(x_i) - ln\sum_{j=1}^{C}\left(e^{w_j^T \phi_j(x_i)}\right)\right)\right) \quad (10)$$

where, $1_{yi=k}$is 1 if and only if y_i= *k*, otherwise it takes value 0. The procedure for calculating the gradient of the Loss function with respect to wk is next on our agenda, and it will be presented here.

$$\nabla_w E = -\sum_{i=1}^{N}\sum_{k=1}^{C}\left(1_{yi=k}\left(x_i - \frac{e^{w_k^T \phi_k(x_i)}}{\sum_{j=1}^{C}\left(e^{w_j^T \phi_j(x_i)}\right)} x_i\right)\right) \quad (11)$$

$$= -\sum_{i=1}^{N}\sum_{k=1}^{C}\left(x_i\left(1_{yi=k} - \frac{e^{w_k^T \phi_k(x_i)}}{\sum_{j=1}^{C}\left(e^{w_j^T \phi_j(x_i)}\right)}\right)\right)$$

$$= -\sum_{i=1}^{N}\sum_{k=1}^{C} x_i\left(\left(1_{yi=k} - p(y_i = k|x_i)\right)\right) \quad (12)$$

During the testing, the testing data are subjected to the identical kernel function transformation as before. We change the weight wk according to the direction in which the gradient is descending. This continues until the wk converges, at which point they consider the wk in the model to be trained appropriately. The equation replaces the modified version of wk and the converted x (6). The next step is to evaluate the probabilities associated with the various classes, after which they select the category with the highest probability that best represents the data. At this stage, the generalization of the logistic regression to multi-classification based on the kernel density function has been finished and is ready for use.

Before calculating the probability using Eq, the DLR algorithm applies a feature transformation to the input x to obtain (6). The input to the probability formula should therefore be changed to read rather than x. At the same time, the Sigmoid function in the probability formula is replaced by the SoftMax function.

4. RESULTS AND DISCUSSION

This section demonstrates the performance of our proposed framework for sentiment analysis using extensive experimental results. The Scikit-Learn module generated results with accurate breakdowns of each class and confusion matrices for binary and multi-class classifiers. The training set and a testing set are split to analyze their feelings. The testing set has 25% of the whole number of sentiments, whereas the training set has 75% of the samples. They used five performance indicators to assess how well our suggested sentiment analysis framework performs: Accuracy, Precision, Recall, F1-measure, and AUC (Area under ROC Curve). The performance measurements can be stated in terms of True Positive, False Positive, True Negative, and False Negative (TP, FP, TN, and FN, respectively).

Exploratory Data Analysis:

For EDA, consider Text Polarity, Text Vectorizer and Word Cloud for Sentiment Analysis on Amazon Musical Instruments Reviews.

eISSN 1303-5150

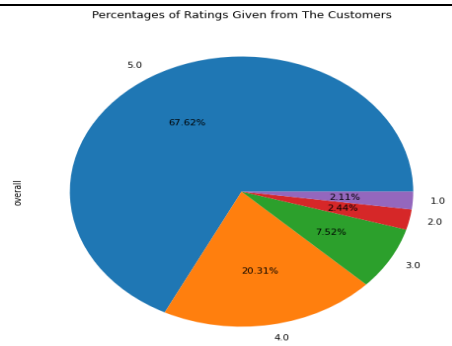


Figure 3: Percentages of ratings given from the customers

From the chart above, the majority of musical instruments sold on Amazon have perfect ratings of 5.0, meaning the condition of the products is good. To denote those ratings above three are positive, ratings equal to 3 are neutral, and ratings under 3 are negative, we that the number of negative reviews given in the dataset is relatively small.

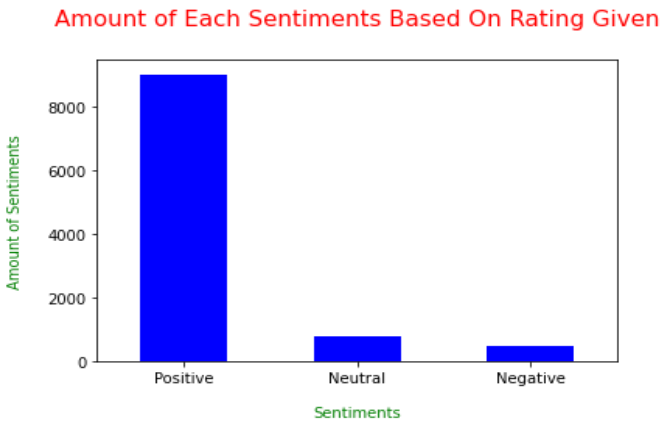


Figure 4:Amount of each sentiment based on rating given

Figure 5 shows the Reviews with negative polarity will be in the range of [-1, 0), neutral ones will be 0.0, and positive reviews will have the range of (0, 1]. From the histogram above, it is clear that most reviews are distributed with positive sentiments, meaning that what they extracted from our analysis before is true. Statistically, this histogram shows that our data is normally distributed but not with the standard distribution. In conclusion, our analysis of the number of sentiments from the reviews is correct and corresponds to the histogram above.

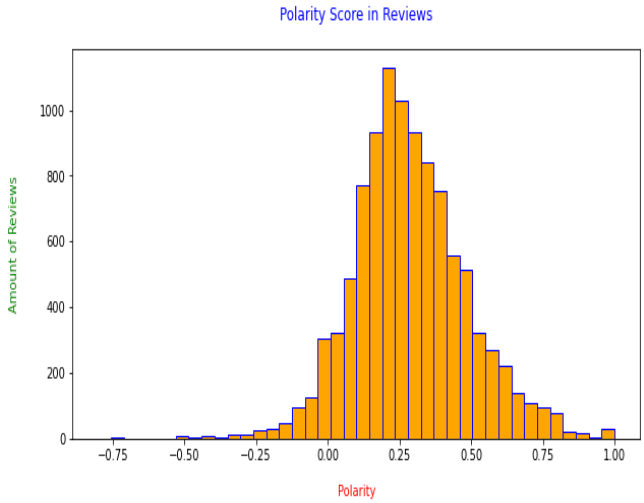


Figure 5: Polarity Score in Reviews

Figure 6 shows that our review has a text length between 0-1000 characters. The distribution has positive skewness; in other words, it is skewed right, which means that our reviews rarely had more considerable lengths than 1000 characters. Of course, the review they use here is affected by the text preprocessing phase, so the size might not be the actual value of the review itself, as some words might have been omitted already. This will also have the same effect when they count the total words in our reviews.

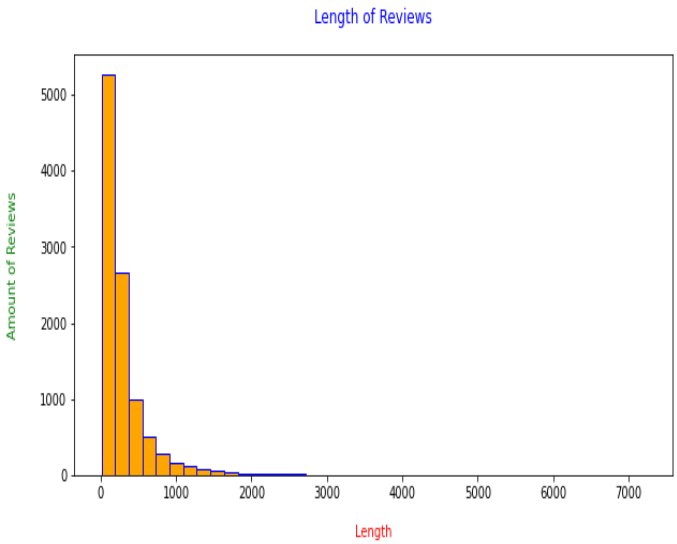
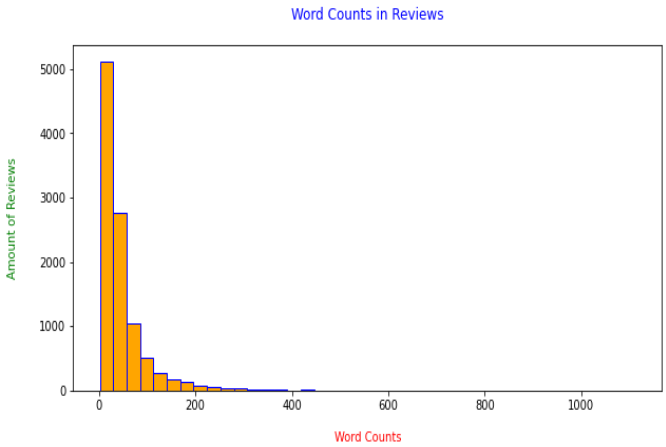


Figure 6: Length of Reviews



eISSN 1303-5150

Figure 7: Word Counts in Reviews

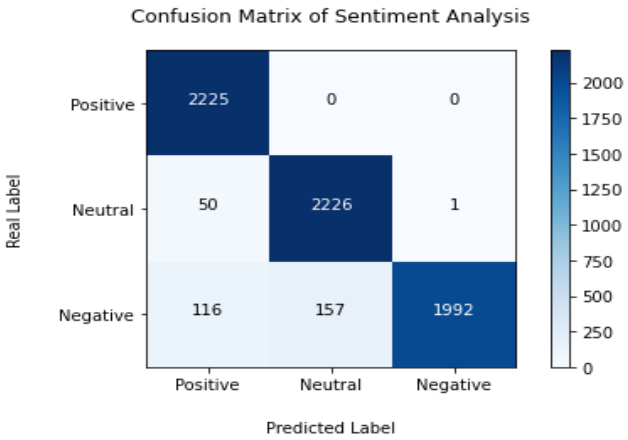


Figure 8: Confusion Matrix Sentiment Analysis

Table 1: Evaluation results of SVM, RF and Our work

Models	Accuracy	Precision	Recall	F1-score
SVM	94.21	91.57	96.57	94.71
RF	94.80	92.54	97.56	95.98
ILR (Our work)	95.22	93.40	98.21	96.17

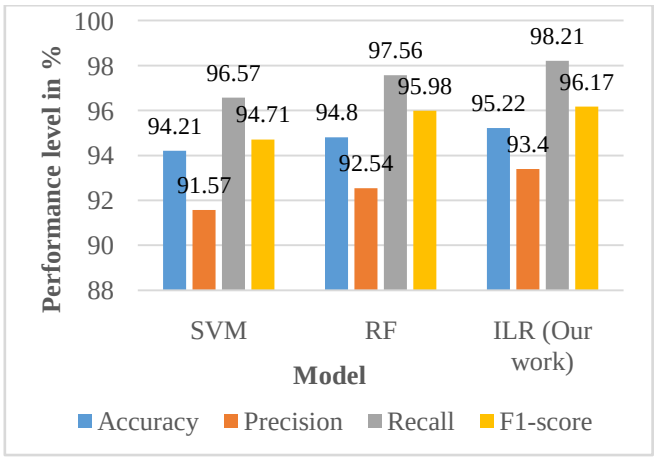


Figure 9: Comparison of the classification accuracy with a SVM, RF and proposed work.

Based on Table 2, the receiver operating characteristic curve (ROC) area under the curve proposed for the model has a value of 93.22% for the test data, and the testing time is 145 msec. An overview of the model's performance may be gleaned from the analysis of the metrics table and the ROC curve visualization.

Table 2: Evaluation results of training data and test data with a baseline and proposed method

Evaluation metric	LR Baseline		ILR Proposed	
	Training Data	Test Data	Training Data	Test Data
ROC AUC	0.8909	0.9175	0.9211	0.9322
Time (msec)	250.109	167.68	189.21	145.71

Figure 10 shows that our method produces less training and testing time of 189.21 msec and 145.71 msec related to the baseline model. The computational time of our model is around 60 sec for training on 300 samples and less than 50 ms for testing on one surface.

Figure 10: Comparison of the training and testing time using baseline and proposed model

5. CONCLUSIONS

Most customers will get a reliable initial impression of a product based on the reviews available on many online buying platforms. As a result, sentiment analysis plays an increasingly essential role in helping sellers identify whether or not their products are satisfying customers and in assisting customers in gaining a better image of the products. Our dataset contains many features about user reviews on musical instruments. But we rarely need those features as our model variables because those features are not important for sentiment analysis. We need to remove stopwords in our preprocessing phase because there might be some important words in determining user sentiments in our model. Our text analysis shows that most transactions are related to guitars or other string-based instruments. The guitar got great attention from the customer pool, and the sellers can emphasize their products on this instrument. This paper presents an improved logistic regression model based on kernel density estimation. This model has the capability of being utilized in the resolution of nonlinear multi-classification issues. We tried almost all classification models available. Using 10-Fold Cross Validation, we get that the Logistic Regression Model got the best accuracy and decided to use this model and tune it. In our attempt to predict our test set, we also received excellent accuracy and a high F1 Score. This means that our model works well on sentiment analysis.

REFERENCES

[1] Wu, Fangyu& Shi, Zhenjie& Dong, Zhaowei& Pang, Chaoyi& Zhang, Bailing. (2020). Sentiment analysis of online product reviews based on SenBERT-CNN. 10.1109/ICMLC51923.2020.9469551.

eISSN 1303-5150

[2] Wang, Binhui& Wang, Ruiqi& Liu, Shujun& Chai, Yanyu& Xing, Shusong. (2020). Aspect-Level Sentiment Analysis of Online Product Reviews Based on Multi-features. 10.1007/978-981-15-3412-6_16.

[3] Bose, Rajesh & Dey, Raktim& Roy, Sandip &Sarddar, Debabrata. (2019). Sentiment Analysis on Online Product Reviews. 10.1007/978-981-13-7166-0_56.

[4] Nandhini, T. &Nivetha, S. & Pavithra, R. & Veena, Dr.S.T.. (2020). Multimodal Sentimental Analysis for Tweets. International journal of recent trends in engineering & research. Special Issue 7. 198-206. 10.23883/ijrter.conf.20200315.031.nlm7o.

[5] Kaur, Ramandeep &Kautish, Sandeep. (2019). Multimodal Sentiment Analysis: A Survey and Comparison. International Journal of Service Science, Management, Engineering, and Technology. 10. 38-58. 10.4018/IJSSMET.2019040103.

[6] Papakitsos, Evangelos&Karanikolas, Nikitas&Samaridi, Nikoletta. (2020). Lexicographic Environments in Natural Language Processing (NLP). 10.1145/3437120.3437310.

[7] Patel, Jay. (2020). Natural Language Processing (NLP) and Text Analytics. 10.1007/978-1-4842-6576-5_4.

[8] Hao, Zhifeng& Liu, Bo & Yang, Xiaowei& Liang, Yanchun& Zhao, Feng. (2005). Twi-Map Support Vector Machine for Multi-classification Problems. 869-874. 10.1007/11427391_139.

[9] Deng, Cai-Xia & Xu, Li-Xiang & Li, Shuai. (2010). Classification of Support Vector Machine and Regression Algorithm. 10.5772/9392.

[10] Chen, Wenlin& Mao, Yi & Guo, Baolong. (2013). Density-based logistic regression. 140-148. 10.1145/2487575.2487583.

[11] Nasir, Noshaba& Zafar, Kashif & Alamgir, Zareen. (2017). Sentiment Analysis of social media Using MapReduce.

[12] Omuya, Erick &Okeyo, George &Kimwele, Michael. (2021). Sentiment Analysis on Social Media using Machine Learning Approach. 10.22541/au.163620143.37655829/v1.

[13] Çelik, Özer&Osmanoğlu, Usame& ÇANAKÇI, Büşra. (2020). Sentiment analysis from social media comments. 8. 366-374. 10.21923/jesd.546224.

[14] Gupta, S. &Sandhane, Raghav. (2022). Use of sentiment analysis in social media campaign design and analysis. Cardiometry. 351-363. 10.18137/cardiometry.2022.22.351363.

[15] Bhoir, Harshala&Jayamalini, K. (2021). Visual Sentiment Analysis on Social Media Data. International Journal of Scientific Research in Computer Science, Engineering and Information Technology. 366-372. 10.32628/CSEIT2174101.

[16] D, Chaithra. (2019). Hybrid approach: naive bayes and sentiment VADER for analyzing sentiment of mobile unboxing video comments. International Journal of Electrical and Computer Engineering (IJECE). 9. 4452. 10.11591/ijece.v9i5.pp4452-4459.

[17] Meng, Zhenghua& Long, & Yu, & Zhao, & Liu, (2019). Cross-Domain Text Sentiment Analysis Based on CNN_FT Method. Information. 10. 162. 10.3390/info10050162.

eISSN 1303-5150