

Comparison Of Support Vector Machines (SVM) And Linear Regression (LR) Sentiment-Based Performance Using E-Commerce Product Reviews.

Cyriaque Havugabaramye, Vedaste Nyandwi

Lecturer at Davis College, Rwanda

Assistant Lecturer at Davis College, Rwanda

Department of Information Systems at Davis College, Rwanda.

Date of Submission: 20-06-2023

Date of Acceptance: 29-06-2023

ABSTRACT

In businesses especially online businesses, Millions of online shoppers regularly share their opinions on social media and review sites. By examining this publicly available customer data, you can identify patterns and understand how your customers think. Sentiment analysis uses artificial intelligence (AI) to collect text data from various sources, identify opinions, and convert the results into positive, neutral, or negative responses to a product, service, or brand. Therefore, e-commerce reviews play an important role for business owners to improve their services. This study uses Logistic Regression and Support Vector Machines classifiers to analyze and predict E-Commerce Reviews' polarity using sentiment analysis.

Keywords: E-Commerce, Customers' reviews, customers' polarity, sentiment analysis, Support Vector Machines (SVM), Linear Regression (LR)

I. INTRODUCTION

With the rapid development of e-commerce technology, users enjoy shopping on various e-commerce platforms. Compared to offline shopping, users can shop anytime, anywhere (Yang et al., 2020). In addition, products on e-commerce platforms are rich in types and styles, allowing consumers to purchase the products they want without leaving their homes (Jiang et al., 2022). Online shopping brings convenience to consumers, but due to the virtual nature of e-commerce platforms, there are many problems with products sold on the platform, such as product sales. Therefore, it is very important to perform sentiment analysis to evaluate products purchased on electronic trading platforms (Kumara et al., 2022).

The trend of e-commerce is pushing people to buy products online for their busy lifestyle. People read all the comments to get a clear idea about the product before buying it. Reading all the comments to get product ideas is impossible, labor intensive and time consuming (Kumara et al., 2022). The challenge of e-commerce product review sentiment analysis (SA) lies in assigning dimensions and disambiguating sentiment words. The problem of dimension mapping is related to mapping the scored blocks of text with the correct dimensions. The term word disambiguation problem refers to situations in which mood words are associated with more than one dimension. Therefore, sentiment analysis of e-commerce reviews is considered a multidimensional classification process (Liu et al., 2020).

Sentiment analysis (SA) provides a way to gain meaningful insights from large volumes of customer reviews. It is used to extract people's attitudes, feelings and perceptions of the product described in the text (Kumara et al., 2022). Analyzing e-commerce data helps online retailers understand customer expectations, provide better shopping experiences, and increase sales. Sentiment analysis can be used to identify positive, negative, and neutral information from customer reviews. Researchers have developed many techniques in sentiment analysis (Mehtab, 2021).

The main of this paper is to analyze e-commerce customers' feedback, predict customers' polarity, score, classify and visualize the feelings and opinions that your customers display in their reviews of hotel using Logistic Regression classifier and Support Vector Machines.

II. RELATED WORKS

Various researches have been conducted on sentiment analysis and machine learning concepts to determine e-commerce customers' polarity from their comments. Below are some of the researchers reviewed in this study.

Product reviews are a determinant of a customer's trust relationship with your store, helping build credibility and trust, and giving potential buyers a perspective on your item and what sets it apart from other items. It helps you understand more clearly(Thomas, 2019). In this study, the researcher recommended the complex strategies for opinion and product feature extraction, and new classification models that can consider name properties ordered in rating predictions.

Advances in e-commerce are influencing customer buying behavior. The buyer makes the desired decision based on the ratings available in the e-commerce(Sasikala & Mary Immaculate Sheela, 2020). The dataset used in this study was composed by 4783 negative words and 2006 positive words with mood values were evaluated for each sentence. The researchers proposed extended system that solves the keyword processing problem and using hybridized algorithms in the future prediction process to improve performance.

E-commerce product review sentiment analysis is considered a multidimensional

classification process that includes three dimensions: document-level, sentence-level, and aspect-level tasks. These three dimensions correspond to specific aspects of the entire document, each sentence, and the entity for sentiment polarity classification(Huang et al., 2022). The researcher proposed the extension of side-level sentiment analysis of e-commerce product reviews and a more in-depth examination of consumers' specific opinions on various attributes of a product, and other consumers.

The previous studies used sentiment analysis and machine learning algorithms to analyze e-commerce customers' feedback. The proposed extension of the studies on side-level sentiment analysis of e-commerce product reviews and a more in-depth examination of consumers' specific opinions on various attributes of a product, and other consumers.

III. METHODOLOGY

This work is expected to analyze for text analysis using Amazon Musical Instruments Reviews dataset revolving around the reviews written by customers. Using target attribute values helps to classify textual data using a Logistic Regression classifier and Support Vector Machines. This dataset contains 10261 rows and 9 columns. Each row corresponds to a customer rating.

Table 1: Description of the dataset

	reviewerID	asin	...	unixReviewTime	reviewTime
0	A2IBPI20UZIR0U	1384719342	...	1393545600	02 28, 2014
1	A14VAT5EAX3D9S	1384719342	...	1363392000	03 16, 2013
2	A195EZSQDW3E21	1384719342	...	1377648000	08 28, 2013
3	A2C00NNG1ZQQG2	1384719342	...	1392336000	02 14, 2014
4	A94QU4C90B1AX	1384719342	...	1392940800	02 21, 2014
...
10256	A14B2YH83ZXMPP	B00JBIVXGC	...	1405814400	07 20, 2014
10257	A1RPTVW5VEOSI	B00JBIVXGC	...	1404259200	07 2, 2014
10258	AWCJ12KBO5VII	B00JBIVXGC	...	1405987200	07 22, 2014
10259	A2Z7S8B5U4PAKJ	B00JBIVXGC	...	1404172800	07 1, 2014
10260	A2WA8TDCTGUADI	B00JBIVXGC	...	1405468800	07 16, 2014

[10261 rows x 9 columns]

The data was downloaded from Kaggle.com whose data will be taken and fit into the machine learning-based sentiment analysis algorithm to meet our specific objectives. The

researcher used python programming language and its important libraries to analyze and predict e-commerce customers' reviews. The data are organized in excel whose extension is csv file.

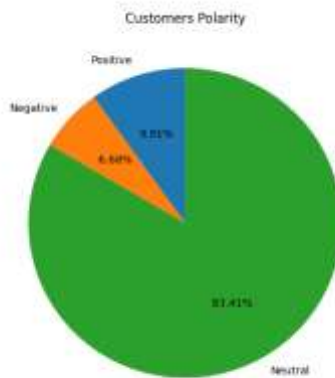


Figure 1: Customers' Polarity about Musical Instruments Reviews.

IV. RESULTS AND DISCUSSION/FINDINGS

The researcher has applied **Support Vector Machines** and **Linear Regression** machine learning algorithms to predict and analyze music instruments customers' polarity.

Support Vector Machines

After applying Support Vector Machines, machine learning algorithm produced 97% of accuracy, 54.02% of area under curve, 96% of precision, 97% of recall and 96% of F1-Score.

	precision	recall	f1-score	support
0	0.21	0.07	0.10	72
1	0.98	0.99	0.99	3005
accuracy			0.97	3077
macro avg	0.59	0.53	0.54	3077
weighted avg	0.96	0.97	0.97	3077

Figure 2: Support Vector Machines Classification Report.



Figure 3: Confusion Matrix of SVM

Linear Regression

After applying Linear Regression, machine learning algorithm produced 97% of accuracy,

53.13% of area under curve, 96% of precision, 97% of recall and 96% of F1-Score.

	precision	recall	f1-score	support
0	0.21	0.07	0.10	72
1	0.98	0.99	0.99	3005
accuracy			0.97	3077
macro avg	0.59	0.53	0.54	3077
weighted avg	0.96	0.97	0.97	3077

Figure 4: Classification Report of Linear Regression
 Confusion Matrix for Logistic Regression



Figure 5: Confusion Matrix of Linear Regression

The following table summarized the performance in terms accuracy, area under curve, precision, recall and f1-score metrics.

Table 2: The summary of two machine learning algorithms in term of metrics

Training Size (%)	MLA	Accuracy	AUC	Precision	Recall	F1-Score
70	Support Vector Machines	97	54.02	96	97	96
	Linear Regression	97	53.13	96	97	97

V. CONCLUSION

This study envisioned to analyze and predict the music instruments customers' polarity from the customers' reviews downloaded from Kaggle.com. It visualized the polarities that can help the managers and the owners or investors to take the right decision. Logistic Regression and SVM classifiers were used, they produced good accuracy, recall, f1-score, precision and area under the curve. The outcomes presented that the 83.41% of the reviews are neutral, 6.68% of the reviews are negative while 9.91% of review are positive. This means that a lot of effort has to be putted in the customers' satisfactions and the quality of the product. The machine learning algorithms used produced better results in terms of performance metrics. The performance metrics considered were accuracy, recall, f1-score, precision and area under the curve. To present data graphically, different python libraries were used such as Matplotlib, Numpy, sklearn, pandas, etc... Word cloud, bigram and trigram highlighted the common used words or set or words as shown with the graphs. One further direction will be designing advanced machine learning that combine all advantages of strong learners to produce good results.

REFERENCE

- [1]. Huang, W., Lin, M., & Wang, Y. (2022). Sentiment Analysis of Chinese E-Commerce Product Reviews Using ERNIE Word Embedding and Attention Mechanism. *Applied Sciences* 2022, Vol. 12, Page 7182, 12(14), 7182. <https://doi.org/10.3390/APP12147182>
- [2]. Jiang, H., Hu, C., & Jiang, F. (2022). Text Sentiment Analysis of Movie Reviews Based on Word2Vec-LSTM. 2022 14th International Conference on Advanced Computational Intelligence (ICACI), 129–134. <https://doi.org/10.1109/ICACI55529.2022.9837505>
- [3]. Kumara, W., Kumara, P., & Sourjah, F. (2022). Value Customer Reviews in E-Commerce Platforms using Sentiment Analysis Tactics: A Review. 2022 3rd International Conference for Emerging Technology (INCET), 1–6. <https://doi.org/10.1109/INCET54531.2022.9824693>
- [4]. Mehtab, A. M. (2021, Jan 13). Sentiment Analysis in E-Commerce. Retrieved from medium.com: <https://medium.com/federatedai/sentiment-analysis-in-e-commerce-e8a06a498a75>
- [5]. Liu, Y., Lu, J., Yang, J., & Mao, F. (2020). Sentiment analysis for e-commerce product reviews by deep learning model of Bert-BiGRU-Softmax. *Mathematical Biosciences and Engineering*, 17(6), 7819–7837. <https://doi.org/10.3934/MBE.2020398>
- [6]. Sasikala, P., & Mary Immaculate Sheela, L. (2020). Sentiment analysis of online product reviews using DLMNN and future prediction of online product using IANFIS. *Journal of Big Data*, 7(1), 1–20. <https://doi.org/10.1186/s40537-020-00308-7>
- [7]. Thomas, K. K. (2019). Sentiment Analysis in Product Reviews using Natural Language Processing and Machine Learning. *International Journal of Information Systems and Computer Sciences*, 8(2), 147–151. <https://doi.org/10.30534/ijiscs/2019/35822019>
- [8]. Yang, L., Li, Y., Wang, J., & Sherratt, R. S. (2020). Sentiment Analysis for E-Commerce Product Reviews in Chinese Based on Sentiment Lexicon and Deep Learning. *IEEE Access*, 8, 23522–23530. <https://doi.org/10.1109/ACCESS.2020.2969854>