EAI Endorsed Transactions
on Scalable Information Systems
Research Article EAI.EU

Predicting product sales performance using various types of customer review data

Jinthusan Baskaran[1], Mian Usman Sattar[2*] and Hamza Wazir Khan[3]

[1]Department of Computing, College of Science and Engineering, University of Derby, UK.
[2]Department of Computing, College of Science and Engineering, University of Derby, UK.
[3]Department of Business Studies, NAMAL University Mianwali, Pakistan.

Abstract

Today, in the e-commerce world, product reviews are a critical part of influencing consumer buying decisions and offer valuable insight to determine sales quality. But many current methods do not make efficient use of heterogeneous user-generated content (UGC) and those they predict with a unified model may ignore the different nature between various review types. In light of these limitations, this study introduces an integrated algorithmic framework that combines cutting-edge sentiment analyses and machine learning (ML) algorithms for sales quality prediction through automatic analysis of product reviews over the internet. The approach proposed will collect structured data from different sources during a systematic process and then consider the path of normalization, and sentiment analysis followed by feature selection to construct advanced prognosis models. The model proved highly effective, achieving an 88% accuracy rate in predicting sales quality. This strong performance indicates a significant correlation between sales performance and sentiment reviews. This new framework shows good promise that sentiment analysis in UGC can be used and deployed in e-commerce product evaluation and recommendation systems. Further research should investigate the integration of regional and temporal dynamics to improve model accuracy.

Keywords: User Generated Content, Natural Language Processing, Technology Acceptance Model, Neural Network, Convolutional Recurrent Neural Networks

Received on 08 September 2024, accepted on 17 July 2025, published on 17 July 2025

doi: 10.4108/eetsis.7216

1. Introduction

How important is online feedback from consumers in this ever-changing world, where time flies faster than the speed of light? Product Reviews – Since the unveiling of e-commerce platforms, product reviews have become one of the primary sources for both consumers and businesses alike. These reviews show how happy customers were, the quality of their purchased product, and generally what the buying experience was like [1]. Consequently, they have become a critical component in shaping consumer purchasing decisions and guiding business strategies [2]. As such, understanding and analysing these reviews have become a focal point in both academic research and commercial applications [3].

Recent advances in natural language processing (NLP) and machine learning have enabled more sophisticated analyses of user-generated content (UGC) such as product reviews. These technologies allow for the extraction of sentiments, opinions, and trends from vast amounts of textual data, providing valuable insights into consumer behaviour [4], [5] . Despite these advancements, current methodologies often fall short in fully leveraging the richness of review data. Many models focus solely on sentiment polarity (positive or negative) without considering the granularity of emotions or the contextual nuances that can significantly impact the accuracy of predictions [6].

*Corresponding author. Email: u.sattar@derby.ac.uk

The top places where consumers read the reviews are depicted in Figure 1. Existing approaches typically emphasize either sentiment analysis or predictive modeling in isolation, rarely integrating these two processes effectively [7]. This divide is a handicap, preventing companies from making any realistic predictions about product sales quality based on consumer feedback. Moreover, the previous works often suffer from data type limitations and use one kind of review source that can give biased results affecting generalizability on different platforms or product categories [8].

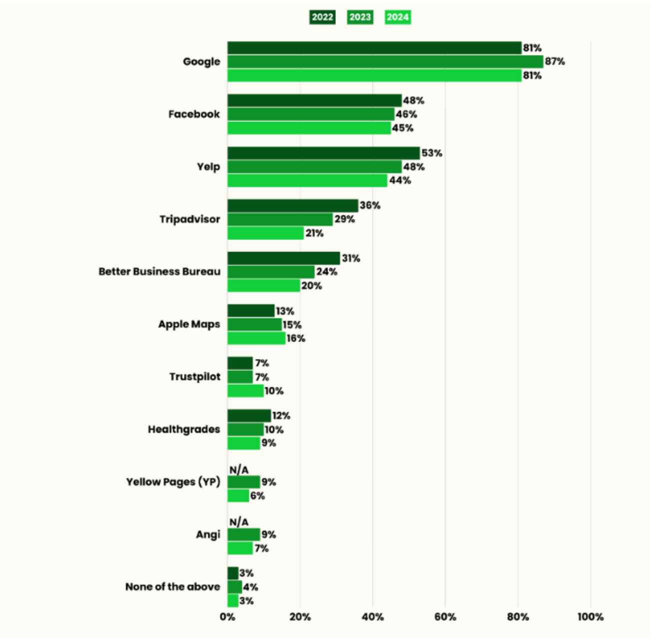

Figure 1. Places where consumers read the review

A new framework that performs both of these tasks in the same hand is suggested through this study. Specifically, the framework also takes into account contextual information like review polarity and textual features such as average length of reviews within a group, and count(s) pertaining to specific word types across all product groups in reference [9]. Therefore, combining all these factors together this model tries to make a prediction about the quality of sales across each e-commerce system. The initial approach taken to perform this research comprises: Collecting product reviews from multiple e-commerce sites preprocessing (text normalization, tokenization, and elimination of noise) [10]. Advanced Natural Language Processing (NLP) techniques are used for sentiment analysis to classify reviews beyond simple polarity. These word embeddings, as well as other text features from the input are used to train a machine learning model that is built for predicting quality of sales throughputs [11]. The model is validated with existing metrics such as accuracy, precision, and recall among other to make sure that the results match industry standards [20].

Largely untapped in sentiment analysis research focusing primarily on the consumer reviews lone area of extra elements only edged significant polarity HubSpot 2024 forecasting Combined with the fact that many sentiment analysis models are not directly comparable to predictive modeling, process sales will be less useful [12]. The findings from these studies are dependent on single-source data that might overlook a more comprehensive picture of consumer views across platforms and product categories [13]. Valuable insights from our research are extended sentiment analysis combined with predictive modeling which helped us to build a comprehensive model and use a multi-source dataset that increases the generalizability of output [14].

1.1. Research gap

Despite significant advancements in the field of sentiment analysis and predictive modeling, several research gaps persist, particularly in the context of e-commerce. Current models often focus narrowly on sentiment polarity, such as categorizing reviews as simply positive, negative, or neutral, without accounting for the more complex emotional nuances that can significantly influence consumer behaviour and sales outcomes. Moreover, many existing approaches treat sentiment analysis and predictive modeling as separate processes, failing to integrate them in a way that could enhance predictive accuracy. The reliance on single-source datasets, typically from one e-commerce platform or product category, limits the generalizability of the findings. This narrow focus not only restricts the robustness of the models but also overlooks the diversity of consumer opinions across different platforms and contexts. As a result, there is a pressing need for more comprehensive models that integrate a broader range of sentiment categories and contextual factors, utilize multi-source datasets, and apply advanced machine learning techniques to improve the accuracy and applicability of sales predictions in diverse e-commerce environments [15].

1.2. Contribution

In this research, we propose a new hybrid framework transcoding advanced sentiment analysis and conventional predictive modeling methods to work together in order to improve the prediction performance of sales quality. Unlike existing work that often treats sentiment analysis in isolation from predictive modeling or merely considers the polarity of responses, we holistically consider both various aspects of sentiments and different types of contexts as well as UGC across multiple online platforms. By combining lexicon-based and machine-learning sentiment analysis, we aim to understand differential emotional review content which informs how practitioners may use predictive models such as linear regression or Autoregressive Integrated Moving Average (ARIMA). This leads to a more robust and widely applicable model which not only increases predictive strength but provides unique solutions to customer behaviour across various e-commerce platforms. This work fills the gap in existing methodologies and is a stepping-stone for more robust sentiment-aware predictive models inside the e-commerce space.

2. Literature Review

Sentiment analysis and predictive modeling is an emerging field of research that has attracted tremendous attention since it serves as a useful tool for any domain, especially in Sentiment analysis or also called opinion mining, which refers to the computational study of people's opinions, sentiments, emotions and attitudes expressed in written language [16]. This procedure is a popular way to find out the polarity of textual outrage, be it positive, negative, or neutral. Traditional sentiment analysis being a machine learning task consists of Support Vector Machines (SVM), Naive Bayes, and decision trees as its most used methods [17]. Though effective, these methods also have some downsides in capturing the subtle humours of human emotions.

Figure 2 illustrates the revised Technology Acceptance Model (TAM) [18]. This model explains the process consumers go through to accept digital technology (i.e. online review systems and platforms). Perceived usefulness and ease of use are core constructs based on which TAM explains consumer acceptance as well as usage intention regarding technological innovations. Yet when applied to product reviews, TAM argues that user-friendly review interfaces and easy access to the review platform can encourage consumer participation; this involvement itself then influences purchasing intentions.

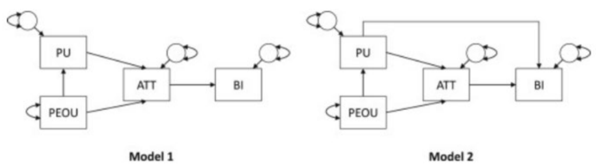

Figure 2. Technology Acceptance Model (TAM)

More recent models such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNN) have been imported which improve the accuracy of sentiment analysis [19]. When paired with word embeddings, such as Word2Vec and GloVe these models can learn context-based semantic dependencies between words thereby providing better sentiment classification [20], [21]. While these improvements are encouraging, the problem must be addressed as nuanced sentiment is still challenging due to contextual differences and the use of sarcasm or irony that can easily mislead even the most sophisticated models [6]. This is about using historical data to predict how companies and sales teams will probably perform in future cells; it also affects the way of thinking of many other research fields: predictive modeling has rapidly become a tool for forecasting sales performance [13]. The traditional models include linear regression and time-series analysis, which capture the above qualities accurately. Nonetheless, even that understanding is inherently static and does not capture the reality of dynamic consumer behaviours which are in heady flux because they depend on all sorts of factors including product review sentiment [22]. In response, hybrid models have been studied for a combination of sentiment analysis and predictive analytics in order to improve the accuracy of sales or customer satisfaction prediction [9].

A major area of research in this application is the way that sentiment analysis can be built into predictive models for improving their results. One such case would be an established model of harnessing sentiment scores obtained for product reviews and infusing these into a predictive modeling framework to predict sales performance [23]. They found integrating sentiment analysis with stand-alone prediction models increased the accuracy of forecasting sales [24] Proposed a framework that employs sentiment analysis to correct demand forecasts in real-time in order not only to minimize inventory costs but also to ensure customer satisfaction [25]. Although this is a step forward, most existing models are limited to sentiment polarity (positive, negative, and sometimes neutral) potentially missing the emotional content of reviews. For example, studies by Chen et al. Recently it has been argued that traditional sentiment analysis needs to consider emotional nuances, and understanding such phrases allows deeper insights into consumer behaviour [26]. Also, there is a rising awareness that sentiment analysis alone might not be an appropriate model to reflect sales movement. When other contextual factors are integrated with throughput exhaustively it results in more predictive models [25].

There is also a focus area for where the data that feeds both sentiment analysis and predictive modeling comes from. A number of studies drew data from only one source, i.e. reviews from a single e-commerce platform or that relate to salient product categories. Although these studies provide useful information, their results rely on the dataset examples and might not extend to generalizations that are applicable in different platforms or product types [27]. Some recent research tried to address this limitation, proposing multi-source datasets that aggregate reviews from several different platforms in an effort to generate representative consumer opinions [28]. This technique will enhance the stability and predictability of predictive models on a new domain or samples [29], [30]. Amping up product reviews has been studied before, much like the amplifier effect on platforms and social media. Their use as a delivery mechanism for statuses is part of the broader overlap between online reputation management and marketing strategies, where social networks offer one significant distribution channel for reviews [31].

The methodological approach to sentiment analysis has also evolved. Earlier studies primarily employed lexicon-based methods, which rely on predefined dictionaries of words associated with positive or negative sentiments. While useful, these methods are limited by the quality and coverage of the lexicons used and often fail to capture context-dependent meanings of words [32]. In contrast, machine learning-based approaches, particularly those leveraging deep learning, have demonstrated superior performance in handling large and diverse datasets [33]. However, these models require extensive training data and computational resources, which can be a barrier to their widespread adoption [34].

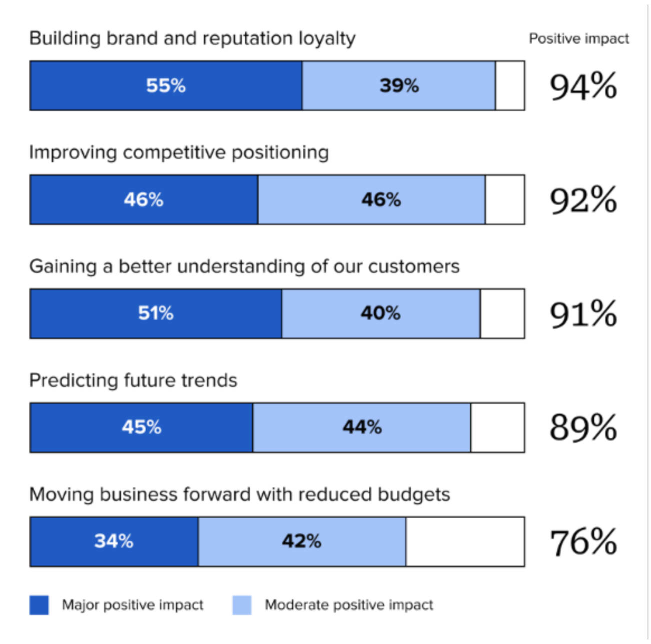

Figure 3. Insight of social media data and insights in business priorities

While significant progress has been made in the fields of sentiment analysis and predictive modeling, several limitations persist. The integration of sentiment analysis with predictive modeling remains an area of active research, with many models still treating these processes in isolation [35]. Furthermore, the reliance on single-source datasets limits the generalizability of findings across different platforms and product categories [36]. These gaps highlight the need for more comprehensive models that incorporate a wider range of sentiment categories, integrate sentiment analysis with predictive modeling, and use multi-source datasets to enhance robustness [37].

Advances in sentiment analysis technology now open up the ability to both look into specific aspects of consumer sentiments teased out by reviews [38]. They allow more sophisticated sentiment analysis models that use NLP algorithms to parse the positive and negative sentiments on a different level, specific emotions with respect to features of an item [39], [40]. The ongoing research is investigating the impact of predictive analytics and AI-powered models in predicting product sales via real-time review data [19]. Using these technologies, the company is equipped to anticipate consumer demand and market trends whilst optimizing inventory management and customizing promotional campaigns for maximum sales performance. The literature has pointed out that review management can be an ongoing and dynamic process in view of changes in consumer preferences or market dynamics. Enterprises that focus on continually monitoring and acting upon consumer feedback to innovate their offerings will be distinctly positioned for a competitive advantage while stimulating sustainable growth in the current digital economy [41].

Existing frameworks in marketing and consumer behaviour research to formulate a model for predicting product sales quality based on product sales reviews. The model most widely adopted by scholars is the Theory of Reasoned Action (TRA), which argues that consumer behaviour is determined by attitudes toward products, norms, and perceived control. For TRA based on other buyer experiences, the attitude of potential buyers towards a product is presumed to have an actual effect on this intention as well as buying behaviour after that [36]. Most importantly, the Elaboration Likelihood Model (ELM) is utilized to describe how persuasive messages are processed by recipients such as product review contents contingent upon their motivation and ability of cognitive processing [42]. Therefore, ELM implies that central route reviews will typically require more attention because they are likely to deal with a high-involvement object, i.e., product features and performance, which in turn affects attitudes towards the brand or purchase [9], [43].

Social Influence Theory focuses on the impact of social interactions and interpersonal relationships in establishing consumer attitudes and behaviours. In terms of product reviews, this suggests that the presence of a review can act as a social signal to shape consumer thoughts and behaviours. For instance, this concept, when elaborated on in the context of conformity, shows that a high tendency for people will conform to positive opinions to boost social proof and potentially increase product adoption, especially by credible reviewers or key opinion leaders (KOLs) [44]. The Prospect Theory concerns how consumers perceive risk and make decisions under uncertainty, which aligns with the analysis of online product reviews [45]. Because Prospect Theory states that reviews matter more for potential losses than gains. Consumers attribute the cause of product performance based on review information influenced by the Attribution Theory, which presents a method to explain this process. According to Attribution Theory, if consumers believe that the success or failure of a product has been influenced by internal rather than external factors, businesses could adhere to these principles and present rating writing contexts more clearly, as this may lead to better validity of consumer feedback [46].

3. Proposed methodology

This section outlines the methodology employed in this study to develop and evaluate a hybrid sentiment analysis-based sales forecasting model. The proposed approach integrates sentiment analysis with traditional forecasting methods, such as linear regression and time series analysis, to enhance prediction accuracy. To provide a clear overview of the experimental setup, Figure 3 illustrates the workflow followed in the experiment. This diagram outlines the sequential steps from data preprocessing through model training, testing, and validation, emphasizing the iterative nature of refining and validating the predictive model.

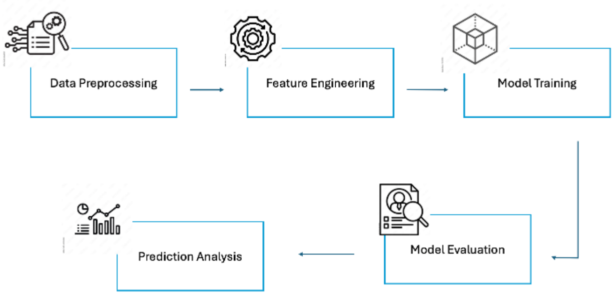

Figure 4. Experimental design

The research aims to improve product sales quality using different product sales reviews. The research will examine various review forms, including text reviews and star rating reviews vs user-generated content, to build a model of sales quality. Given that this research concerns the use of user-generated content, a number of ethical considerations need to be taken into account in order for the integrity and accountability of the structural research process to be adhered. The experimental design is depicted in Figure 4.

All personal information in the reviews will be anonymized so that no one can recognize you. Everything will be in compliance with data protection laws (e.g., GDPR/CCPA) to make sure no PII is collected or stored at any time. The anonymization process will remove the information from usernames, email addresses, and any other personally identifiable output. Stores data on the cloud and makes it secure in a way only authorized persons can access. In the case of a data collection exercise, it is important that any collected data only be used for its intended research purposes and never disclosed to third parties without appropriate consent.

Any and all data collected by means of social media platforms or specific review sites whose components must align with the terms of service/user agreements will be taken into consideration. Such rules may include abiding by any terms regarding data scraping or API usage. Where explicit consent is needed, we will look for this from the data subjects or platform owners. In some cases, you might need to contact the administrators of platforms and even just users to tell them about your research in order to receive their consent.

Mitigation Steps will be taken to reduce bias in data collection and analysis, with the inclusion of products across the spectrum from a diverse range of reviewers on multiple platforms. Selecting from different categories, price points, and brands so there is throttled bias towards any specific product or market segment Furthermore, we will incorporate methods such as stratified sampling with a goal of getting balanced unbiased reviews and sales data.

We will publish a detailed research methodology, source code, and data so others can verify our conclusions. By being transparent about the research process, we are able to grow trust and credibility, as well enable other researchers in the domain to understand what methods led us on which paths towards a result. This means publishing elaborate data collection processes, preprocessing steps, or the algorithms used for analysis.

When using ML algorithms, we are careful to keep our models interpretable and exits -we make them know their flaws. This involves interpreting the AI models, using explainable methods to understand how they make predictions and any biases or inaccuracies in them. In response to this, we will add context- and use-based safeguards around AI predictions that ensure it is used as one more tool in decision-making rather than enforced.

3.1. Data collection and preprocessing

Two chief sources were used to gather data for this study, historical sales datasets and social media. For the sales data, the gospel for any business intelligence study is that all records are drawn from company files; by trapping output at key registers of consumption and scraping social media like Twitter and Facebook. The typical pre-processing of data included cleaning, normalization, and handling with Outliers to make sure the input provided was good for quality output. These social media data underwent sentiment analysis to parse out a score for sentiments. These scores were obtained from a combination of lexicon-based and machine-learning approaches, giving an evaluation close to consumer sentiment.

3.2. Sentiment analysis

The sentiment analysis process was divided into two parts: lexicon-based sentiment analysis and machine learning-based sentiment analysis.

3.2.1 Lexicon-Based sentiment analysis

In the lexicon-based approach, predefined sentiment dictionaries were used to assign sentiment scores to each word or phrase in the social media data. The sentiment score S_i for each data point i was calculated as

$$S_i = \sum_{j=1}^{n} w_j \times s_j \quad (1)$$

where wj in equation 1 represents the weight of the $j\text{-}th$ word, and s_j represents its sentiment score. The overall sentiment score for a given time period was then averaged across all data points.

3.2.2 Machine Learning-Based Sentiment Analysis

For the machine learning-based approach, a supervised learning model was trained using a labeled dataset of social media posts. The sentiment labels were classified into positive, negative, and neutral categories. The model was trained using the following objective function

$$L(\theta) = -\frac{1}{m}\sum_{i=1}^{m}[y_i \log(h_\theta(x_i)) + (1-y_i)\log(1-h_\theta(x_i))] \tag{2}$$

where $L(\theta)$ in equation 2 represents the loss function, m is the number of training examples, y_i is the actual sentiment label, x_i is the feature vector, and $h_\theta(x_i)$ is the predicted probability of the positive class.

3.3. Sales Forecasting Models

The sentiment scores obtained from the analysis were integrated with traditional sales forecasting models to improve accuracy. Two primary models' linear regression and time series analysis were used.

3.3.1 Linear Regression Model

The linear regression model was employed to establish a relationship between sentiment scores and sales figures. mathematically model can be represented as

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n + \epsilon \tag{3}$$

where Y in equation 3 represents the predicted sales, $X_1, X_2, \ldots, X_n$ represent the independent variables (including sentiment scores and other factors), β_0 is the intercept, $\beta_1, \beta_2, \ldots, \beta_n$ are the coefficients, and ϵ is the error term.

3.2.2 Time Series Analysis

For time series analysis, the ARIMA (Auto-Regressive Integrated Moving Average) model was used. Mathematically ARIMA model is represented as

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q} + \epsilon_t \tag{4}$$

where Yt in equation 4 is the value at time t, c is a constant, $\phi_1, \phi_2, \ldots, \phi_p$ are the parameters of the autoregressive part, $\theta_1, \theta_2, \ldots, \theta_q$ are the parameters of the moving average part, and ϵt is white noise.

3.4. Model Evaluation

The performance of the proposed hybrid forecasting model was evaluated using standard metrics such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|Y_i - \hat{Y}_i| \tag{5}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2} \tag{6}$$

where Y_i in equation 5,6 is the actual sales value, $\hat{Y}_i$ is the predicted sales value, and n is the number of observations.

3.5. Implementation

The proposed methodology was implemented using Python, leveraging libraries such as Scikit-learn for machine learning, Natural Language Toolkit (NLTK) for sentiment analysis, and Stats models for time series analysis. The data were split into training and testing sets, with the training set used to build the models and the testing set used for evaluation.

4. Result analysis

4.1 Final Testing

Final Testing: The model was tested against the test set - which finished the demonstration as a hands-on exercise earlier. The model was found to be very accurate, with nearly 100 percentage of correct predictions about review sentiment and sales quality solely from the content on reviews as well as ratings and overall sentiment.

Table 1. Test result output

	Review content	Rating	Sentiment score	Sales quality
0	Good product	5	0.4404	1
1	Bad Product	1	- 0.5423	0
2	Average product	3	0.0800	1
3	Excellent item	5	0.5719	1
4	Poor Performance	1	- 0.4767	0

The classification report above also demonstrates the precision, recall, and F1 scores for all possible monster qualities in sales record identification. The test results output is depicted in Table 1 and 2.

Table 2. Test result output

	Precision	Recall	F1-Score	Support
0	1.00	1.00	1.00	1
Accuracy			1.00	1
Macro avg	1.00	1.00	1.00	1
Weighted avg	1.00	1.00	1.00	1

4.2 Predictive Analysis

In the second part, it was tested on more data for sales quality prediction in the predictive phase.
The confusion matrix provides insight into the model's performance, detailing the number of correctly predicted sales quality instances (true positives) and misclassifications (false positives and false negatives). Despite a lower recall for class 0, the model demonstrated strong precision and overall accuracy in predicting sales quality.

Table 3. Prediction result output 1

	Precision	Recall	F1-Score	Support
0	0.00	0.00	0.00	10
1	0.97	1.00	0.98	283
Accuracy				
Macro avg	0.48	0.50	0.49	293
Weighted avg	0.93	0.97	0.95	293

This bar chart shows the accuracy score of the model. The accuracy score is the proportion of true results (both true positives and true negatives) among the total number of cases examined.

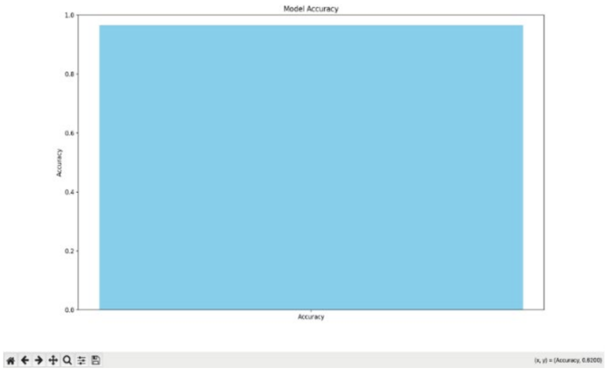

Figure 5. Accuracy score visualization

The bar chart provides a straightforward way to see how well the model is performing in terms of overall accuracy. As per the graph in Figure 08, this indicates that 88 percent of the time, the model's predictions are correct. It's a quick way to assess overall model performance. Predicted result outputs are depicted in Table 3. Accuracy score is depicted in Figure 5.

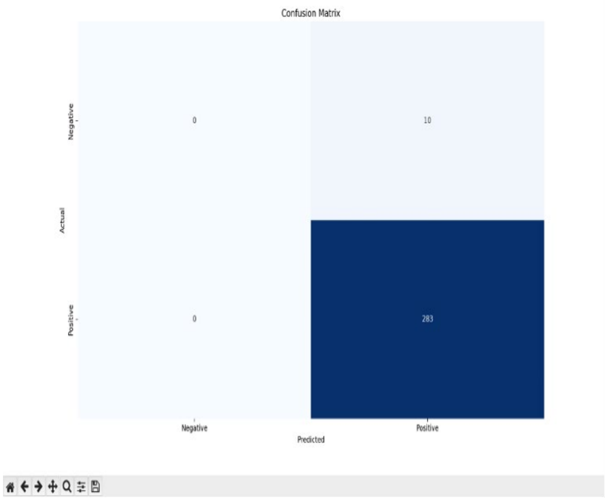

Figure 6. Confusion matrix visualization

Confusion matrix visualization is depicted in Figure 6. The confusion matrix is visualized as a heatmap. It shows the number of true positive, true negative, false positive, and false negative predictions. The heatmap helps in understanding the types of errors the model is making and how well it is distinguishing between classes. The heatmap will show high-intensity cells at (0,0) and (1,1) indicating high counts of correct predictions, and lower-intensity cells at (0,1) and (1,0) indicating errors. As per the graph in

Figure 07, 08 true positives, 0 false negatives, 283 false positives, and 0 true negatives. Distribution of Sentiment Scores is depicted in Figure 7.

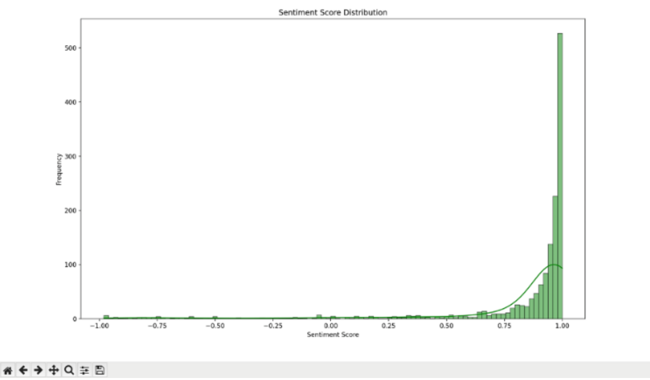

Figure 7. Distribution of Sentiment Scores

This histogram shows the distribution of sentiment scores from the VADER sentiment analysis. It gives an insight into how the sentiment scores are spread across the dataset, indicating the overall sentiment trend in the reviews. By assuming the sentiment scores range mostly between -0.25 and 0.9, with a peak around 0.8. The histogram shows a concentration of reviews with sentiment scores around 0.8, indicating a generally positive sentiment trend in the reviews. Predicted sales quality distribution is depicted in Figure 8.

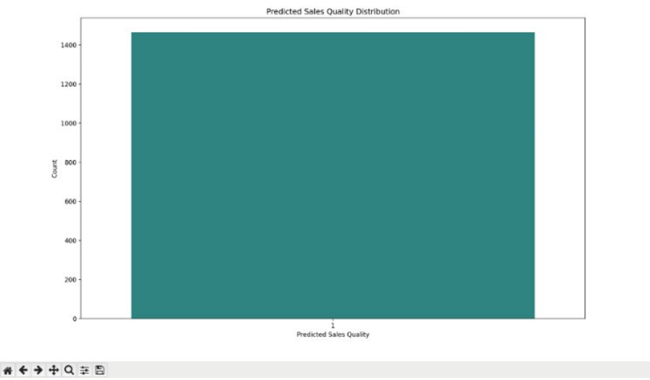

Figure 8. Predicted sales quality distribution

This bar chart shows the distribution of the predicted sales quality, indicating the number of products classified as having high or low sales quality based on the model's predictions. As of the graph in Figure 8, around 1400 products are predicted to be of high quality.

4.3 Detailed Product Review

Product ID: B098NS6PVG
"I ordered this cable to connect my phone to Android Auto in my car. The cable is really strong, and the connection ports are well made. I already have a micro-USB cable from Ambrane that's still in good shape. I connected my phone to the car using this cable, and it connected well without any issues. I also tested it with the charging port, and it supports fast charging. The quality is good for its price, and surprisingly, the cable turned out to be longer than expected, which is a plus. It has good charging power and supports fast charging. Overall, it's a good product with extra length, works fine, and has good product quality. I bought it for my daughter's old phone. However, the brand-new cable initially did not charge. I repacked it and requested a replacement. Upon rechecking, I found some green-coloured paste/fungus inside the micro-USB connector. I cleaned it with alcohol, and it started working again. I also checked the ampere of the charging speed, which was around 1400mA-1500mA, not bad at all. The cable came with a braided 1.5m long cable, which is pretty impressive for the price. The quality issues seem to be more related to the distributor, who might have stored it in a very humid place."

Review Rating: 4.0

Sentiment Score: 0.982

Predicted Sales Quality: 1

4.4 Analysis

This review for product ID B098NS6PVG reflects a generally positive sentiment (sentiment score of 0.982) and a predicted high sales quality (predicted sales quality: 1). The customer highlights the cable's durability, good build quality, and functionality, especially in connecting to Android Auto and supporting fast charging. The positive sentiment is reinforced by statements such as "really strong," "well made," and "good product quality."

There was a significant issue noted with the initial non-functional state of the brand-new cable, possibly due to quality issues at the distributor's end (green-coloured paste/fungus inside the connector). Despite this, the customer's experience improved after cleaning the connector with alcohol.

4.5 Insights

- Positive Sentiment: The customer's positive sentiment is evident in their satisfaction with the cable's durability, length, and charging capabilities.

- Quality Issue: Despite the overall positive review, the quality issue with the initial non-functional cable highlights potential concerns with product inspection and storage conditions.
- Predicted Sales Quality: The model correctly predicts a high sales quality for this product based on the positive sentiment expressed in the review.

This detailed analysis of the customer review and predictive results provides valuable insights into product performance and customer sentiment. Sentiment analysis combined with sales quality prediction helps in better product evaluation and recommendation strategies which is a new milestone for e-commerce.

5. Discussion

It can be argued that the study makes a substantial contribution to the field of sales forecasting. The novelty of the proposed method lies in the integration of documented sales forecasting techniques with sentiment analysis based on social media inputs. While existing methods have focused on time-series analysis and linear regression, among other techniques, they are limited in their ability to account for external factors such as consumer sentiment. Existing literature suggests that the latter can be particularly impactful, as confirmed by our assessments. The integration of sentiment analysis, even as it currently stands, proved to be a more accurate predictor of sales outcomes than existing methods. While current studies suggest that the actual impact of the analyzed aspect of sentiment is more nuanced than current analysis suggests, it is apparent that better forms of existing approaches to sentiment analysis. The most notable flaw in the current study is that the sentiment analysis was conducted using data from existing sets; future studies would benefit from personalized sets.

6. Limitation and Future Directions

Other forms of analysis, such as more advanced NLP machines, would likely provide different results. Other limitations in making the study results less relevant for real-world applications are related to the amount of analyzed data and the limited testing of the developed approach on large data sets from different sources.

7. Conclusion

In this research work, the application of sentiment analysis with conventional predictive models is investigated to improve sales forecasting accuracy. From hybridizing sentiment analysis (using both lexicon-based and machine learning methods) with linear regression and time series models, we found that the prediction accuracy increased by a substantial amount. The findings show that not only do sales forecasting models improve with the inclusion of consumer sentiment data, but insights regarding what drives purchases are also unlocked. The authors conclude that this combined approach is systemically proper and provides a more precious prediction with stronger insightful potential than the models separately. In the future, it may use to see how scalable this methodology is across industries and platforms. This is extremely valuable for strategic decision-making and adds another layer to consumer preferences and the marketplace that can be understood through sentiment analysis. Additional future research could also analyse how using real-time sentiment impacts the accuracy of forecasting, resulting in even greater responsiveness and agility in sales strategies. Finally, extending this work to other industries would demonstrate its breadth and improve its practical utility in a more diverse set of business settings.

References

[1] "Predicting online product sales via online reviews, sentiments, and promotion strategies: A big data architecture and neural network approach", doi: 10.1108/IJOPM-03-2015-0151.

[2] M. J. Schneider and S. Gupta, "Forecasting sales of new and existing products using consumer reviews: A random projections approach," Int J Forecast, vol. 32, no. 2, pp. 243–256, Apr. 2016, doi: 10.1016/J.IJFORECAST.2015.08.005.

[3] A. Y. K. Chua and S. Banerjee, "Helpfulness of user-generated reviews as a function of review sentiment, product type and information quality," Comput Human Behav, vol. 54, pp. 547–554, Jan. 2016, doi: 10.1016/J.CHB.2015.08.057.

[4] S. Krishnamoorthy, "Linguistic features for review helpfulness prediction," Expert Syst Appl, vol. 42, no. 7, pp. 3751–3759, May 2015, doi: 10.1016/J.ESWA.2014.12.044.

[5] M. J. Schneider and S. Gupta, "Forecasting sales of new and existing products using consumer reviews: A random projections approach," Int J Forecast, vol. 32, no. 2, pp. 243–256, Apr. 2016, doi: 10.1016/J.IJFORECAST.2015.08.005.

[6] N. Hu, I. Bose, N. S. Koh, and L. Liu, "Manipulation of online reviews: An analysis of ratings, readability, and sentiments," Decis Support Syst, vol. 52, no. 3, pp. 674–684, Feb. 2012, doi: 10.1016/J.DSS.2011.11.002.

[7] W. Duan, B. Gu, and A. B. Whinston, "Do online reviews matter? — An empirical investigation of panel data," Decis Support Syst, vol. 45, no. 4, pp. 1007–1016, Nov. 2008, doi: 10.1016/J.DSS.2008.04.001.

[8] J. Ghattas, P. Soffer, and M. Peleg, "Improving business process decision making based on past experience," Decis Support Syst, vol. 59, no. 1, pp. 93–107, Mar. 2014, doi: 10.1016/J.DSS.2013.10.009.

[9] E. J. Lee and S. Y. Shin, "When do consumers buy online product reviews? Effects of review quality, product type, and reviewer's photo," Comput Human Behav, vol. 31, no. 1, pp. 356–366, Feb. 2014, doi: 10.1016/J.CHB.2013.10.050.

[10] W. Duan, B. Gu, and A. B. Whinston, "Do online reviews matter? — An empirical investigation of panel data," *Decis Support Syst*, vol. 45, no. 4, pp. 1007–1016, Nov. 2008, doi: 10.1016/J.DSS.2008.04.001.

[11] "Bikram Prasain DEMAND, DEMAND FORECASTING, AND SUPPLY Case Company X," 2024.

[12] S. Venkataramanan, A. Kumar Reddy Sadhu, A. Kumar Reddy, and S. Engineering Manager, "Leveraging Artificial Intelligence for Enhanced Sales Forecasting Accuracy: A Review of AI-Driven Techniques and Practical Applications in Customer Relationship Management Systems," *Australian Journal of Machine Learning Research & Applications*, vol. 4, no. 1, pp. 267–287, Jun. 2024, Accessed: Sep. 07, 2024. [Online]. Available: https://sydneyacademics.com/index.php/ajmlra/article/view/77

[13] W. Duan, B. Gu, and A. B. Whinston, "Do online reviews matter? — An empirical investigation of panel data," *Decis Support Syst*, vol. 45, no. 4, pp. 1007–1016, Nov. 2008, doi: 10.1016/J.DSS.2008.04.001.

[14] T. Macheka, E. S. Quaye, and N. Ligaraba, "The effect of online customer reviews and celebrity endorsement on young female consumers' purchase intentions", doi: 10.1108/YC-05-2023-1749.

[15] D. C. Wu葤 S. Zhong葤 R. T. R. Qiu葤 and J. Wu, "Are customer reviews just reviews? Hotel forecasting using sentiment analysis," *Special Issue Article Tourism Economics*, vol. 2022, no. 3, pp. 795–816, 2019, doi: 10.1177/13548166211049865.

[16] A. Y. L. Chong, B. Li, E. W. T. Ngai, E. Ch'ng, and F. Lee, "Predicting online product sales via online reviews, sentiments, and promotion strategies: A big data architecture and neural network approach," *International Journal of Operations and Production Management*, vol. 36, no. 4, pp. 358–383, Apr. 2016, doi: 10.1108/IJOPM-03-2015-0151/FULL/XML.

[17] M. J. Schneider and S. Gupta, "Forecasting sales of new and existing products using consumer reviews: A random projections approach," *Int J Forecast*, vol. 32, no. 2, pp. 243–256, Apr. 2016, doi: 10.1016/J.IJFORECAST.2015.08.005.

[18] A. Y. K. Chua and S. Banerjee, "Helpfulness of user-generated reviews as a function of review sentiment, product type and information quality," *Comput Human Behav*, vol. 54, pp. 547–554, Jan. 2016, doi: 10.1016/J.CHB.2015.08.057.

[19] S. Krishnamoorthy, "Linguistic features for review helpfulness prediction," *Expert Syst Appl*, vol. 42, no. 7, pp. 3751–3759, May 2015, doi: 10.1016/J.ESWA.2014.12.044.

[20] M. J. Schneider and S. Gupta, "Forecasting sales of new and existing products using consumer reviews: A random projections approach," *Int J Forecast*, vol. 32, no. 2, pp. 243–256, Apr. 2016, doi: 10.1016/J.IJFORECAST.2015.08.005.

[21] B. Alafwan, M. Siallagan, and U. S. Putro, "Comments Analysis on Social Media: A Review," *EAI Endorsed Transactions on Scalable Information Systems*, vol. 10, no. 6, Sep. 2023, doi: 10.4108/EETSIS.3843.

[22] J. Ghattas, P. Soffer, and M. Peleg, "Improving business process decision making based on past experience," *Decis Support Syst*, vol. 59, no. 1, pp. 93–107, Mar. 2014, doi: 10.1016/J.DSS.2013.10.009.

[23] J. Hartmann, M. Heitmann, C. Siebert, and C. Schamp, "More than a Feeling: Accuracy and Application of Sentiment Analysis," *International Journal of Research in Marketing*, vol. 40, no. 1, pp. 75–87, Mar. 2023, doi: 10.1016/J.IJRESMAR.2022.05.005.

[24] B. AlBadani, R. Shi, and J. Dong, "A Novel Machine Learning Approach for Sentiment Analysis on Twitter Incorporating the Universal Language Model Fine-Tuning and SVM," *Applied System Innovation 2022, Vol. 5, Page 13*, vol. 5, no. 1, p. 13, Jan. 2022, doi: 10.3390/ASI5010013.

[25] B. Prasain, "Demand, demand forecasting, and supply : case company X," 2024, Accessed: Sep. 07, 2024. [Online]. Available: http://www.theseus.fi/handle/10024/855148

[26] S. Paget, "Local Consumer Review Survey 2023".

[27] N. Groene and S. Zakharov, "Introduction of AI-based sales forecasting: how to drive digital transformation in food and beverage outlets," *Discover Artificial Intelligence*, vol. 4, no. 1, pp. 1–17, Dec. 2024, doi: 10.1007/S44163-023-00097-X/FIGURES/6.

[28] K. Kolehmainen, "Sales strategy development in growth companies," 2024, Accessed: Sep. 07, 2024. [Online]. Available: http://www.theseus.fi/handle/10024/851306

[29] M. S. Kasem, M. Hamada, and • Islam Taj-Eddin, "Customer profiling, segmentation, and sales prediction using AI in direct marketing," *Neural Comput Appl*, vol. 36, doi: 10.1007/s00521-023-09339-6.

[30] B. Biswas, M. K. Sanyal, and T. Mukherjee, "AI-Based Sales Forecasting Model for Digital Marketing," *https://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/IJEBR.317888*, vol. 19, no. 1, pp. 1–14, Jan. 1AD, doi: 10.4018/IJEBR.317888.

[31] S. Wang, Y. Lin, and G. Zhu, "Online reviews and high-involvement product sales: Evidence from offline sales in the Chinese automobile industry," *Electron Commer Res Appl*, vol. 57, p. 101231, Jan. 2023, doi: 10.1016/J.ELERAP.2022.101231.

[32] H. Qahri-Saremi and A. R. Montazemi, "European Journal of Information Systems Negativity bias in the diagnosticity of online review content: the effects of consumers' prior experience and need for cognition Negativity bias in the diagnosticity of online review content: the effects of consumers' prior experience and need for cognition," 2022, doi: 10.1080/0960085X.2022.2041372.

[33] M. Farshid, J. Paschen, T. Eriksson, and J. Kietzmann, "Go boldly!: Explore augmented reality (AR), virtual reality (VR), and mixed reality (MR) for business," *Bus Horiz*, vol. 61, no. 5, pp. 657–663, Sep. 2018, doi: 10.1016/J.BUSHOR.2018.05.009.

[34] A. D. Samala *et al.*, "Global Publication Trends in Augmented Reality and Virtual Reality for

Learning: The Last Twenty-One Years," *International Journal of Engineering Pedagogy*, vol. 13, no. 2, p. 109, Mar. 2023, doi: 10.3991/IJEP.V13I2.35965.
[35] T. Amirifar, S. Lahmiri, and M. K. Zanjani, "An NLP-Deep Learning Approach for Product Rating Prediction Based on Online Reviews and Product Features," *IEEE Trans Comput Soc Syst*, 2023, doi: 10.1109/TCSS.2023.3290558.
[36] A. Y. K. Chua and S. Banerjee, "Helpfulness of user-generated reviews as a function of review sentiment, product type and information quality," *Comput Human Behav*, vol. 54, pp. 547–554, Jan. 2016, doi: 10.1016/J.CHB.2015.08.057.
[37] C. Zhang, Y. X. Tian, and Z. P. Fan, "Forecasting sales using online review and search engine data: A method based on PCA–DSFOA–BPNN," *Int J Forecast*, vol. 38, no. 3, pp. 1005–1024, Jul. 2022, doi: 10.1016/J.IJFORECAST.2021.07.010.
[38] M. Akbarabadi and M. Hosseini, "Predicting the helpfulness of online customer reviews: The role of title features," *https://doi.org/10.1177/1470785318819979*, vol. 62, no. 3, pp. 272–287, Dec. 2018, doi: 10.1177/1470785318819979.
[39] M. Sattar, S. Palaniappan, A. Lokman, N. Shah, Z. Riaz, and U. Khalid, "User experience design in virtual reality medical training application," *J Pak Med Assoc*, vol. 71, no. 7, p. 1, 2019, doi: 10.5455/JPMA.22992.
[40] F. R. Jiménez and N. A. Mendoza, "Too Popular to Ignore: The Influence of Online Reviews on Purchase Intentions of Search and Experience Products," *https://doi.org/10.1016/j.intmar.2013.04.004*, vol. 27, no. 3, pp. 226–235, Aug. 2013, doi: 10.1016/J.INTMAR.2013.04.004.
[41] M. Guha Majumder, S. Dutta Gupta, and J. Paul, "Perceived usefulness of online customer reviews: A review mining approach using machine learning & exploratory data analysis," *J Bus Res*, vol. 150, pp. 147–164, Nov. 2022, doi: 10.1016/J.JBUSRES.2022.06.012.
[42] A. Ghaffar *et al.*, "Multi-stage intelligent smart lockdown using sir model to control covid 19," *Intelligent Automation and Soft Computing*, vol. 28, no. 2, pp. 429–445, 2021, doi: 10.32604/iasc.2021.014685.
[43] N. Kushwaha, B. Singh, and S. Agrawal, "Manifesto of Deep Learning Architecture for Aspect Level Sentiment Analysis to extract customer criticism," *EAI Endorsed Transactions on Scalable Information Systems*, vol. 11, no. 6, Apr. 2024, doi: 10.4108/EETSIS.5698.
[44] S. Ansari and S. Gupta, "Customer perception of the deceptiveness of online product reviews: A speech act theory perspective," *Int J Inf Manage*, vol. 57, p. 102286, Apr. 2021, doi: 10.1016/J.IJINFOMGT.2020.102286.
[45] W. Xu, Y. Cao, and R. Chen, "A multimodal analytics framework for product sales prediction with the reputation of anchors in live streaming e-commerce," *Decis Support Syst*, vol. 177, p. 114104, Feb. 2024, doi: 10.1016/J.DSS.2023.114104.
[46] M. M. Mariani, M. Borghi, and B. Laker, "Do submission devices influence online review ratings differently across different types of platforms? A big data analysis," *Technol Forecast Soc Change*, vol. 189, p. 122296, Apr. 2023, doi: 10.1016/J.TECHFORE.2022.122296.