Online Reviews and Product Sales: A Sentiment Analysis Approach

MSc Research Project

Alan Thomas

Student ID: X23188944

School of Computing

National College of Ireland

Supervisor: Eamon Nolan

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Alan Thomas
Student ID:	X23188944
Programme:	MSc in Data Analytics **Year:** Jan 2024
Module:	Msc Research Project
Supervisor:	Eamon Nolan
Submission Due Date:	13th December 2024
Project Title:	Online Reviews and Product Sales: A Sentiment Analysis Approach
Word Count:	6927 **Page Count**22

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Alan Thomas

Date: 13/12/2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Online Reviews and Product Sales: A Sentiment Analysis Approach

Alan Thomas

X23188944

Abstract

Today, getting opinions of customers over the internet has become essential in understanding their buying habits and managing the profits of products. This paper seeks to discuss the impact of reviews, sympathetic, mixed or otherwise on the general public in their purchasing power. On an open dataset containing product reviews and sales, Logistic Regression, Random Forest, SVM, and XGBoost were tested, and the maximum accuracy was obtained at 86.46%. For the businesses to be able to gain something tangible from the sentiments analysis, an interface dashboard was developed to draw trends as well as the correlation between sentiments and sales. The findings unveiled a moderate positive relationship between the review sentiments, and sales trends. However, low sales sometimes came in the same periods with positive reviews, making it hard to conclude that them was an internal problem I external factors such as changes in customer demand and consumers' preferences. As seen above, issues such as sarcasm classification and handling huge data se where addressed but are still viable prospects. The work presented here demonstrates the use of sentiment analysis and visualization techniques in strategic management and indicates some possible improvements such as the use of more complex models and integration with additional data sources. It contributes to a cumulative body of knowledge on the role that online reviews play for consumers and brings a set of useful heuristics for dealing with the effects of these reviews on sales.

1 Introduction

In contemporary times, online feedback has become an important part of decision-making. Moreover, it influences buying behavior among people. As e-commerce has grown exponentially, consumers more than ever depend on feedback from others to see how good the product is, how reliable it is, and other usage values. Websites like Amazon, Yelp, eBay, etc. have millions of reviews. They create consumer perceptions and buying choices. Even though reviews are obvious, it is still a matter of research and business interest to discover how they actually affect sales. This study describes the impact of online reviews on product sales. This is the impact of rating sentiments, feedback quality, and sales trends. In fact, it tries to answer questions like: Do positive reviews actually lead to more sales? Are negative reviews a turnoff? Are products that have received many positive reviews more successful? This study will answer these questions concerning how customer feedback can be used as a method to improve the sales of the business.

The importance of this research is based on the idea that it would help to connect consumer feedback and business strategy. Instead of focusing on revenue loss after a negative review, as previous research has done, this paper analyzes data from hundreds of thousands of reviews of face masks and masks sales, incorporating advanced techniques of sentiment analysis with sale records from actual businesses. (This is written in a data-driven style, presenting studies that analyze publicly available datasets with machine learning models to quantify how reviews affect sales dynamics.) This study creates new knowledge about the use of digital feedback to revolutionize consumer markets by showing patterns and trends (which were previously unobserved) that will directly contribute to further research efforts in this area)

2 Related Work

In the recent years we have seen a major increase in the interest of people on online reviews and product ratings and we have also seen how these affects the product sales, which highlights the importance of consumer input in the online marketplace. This section talks about the conclusions, advantages, disadvantages and significanct scholarly research that investigates the relationship between the online reviews and the sale of the product. Here in this review we lay a basis for the current work and its applicability to the area by critically examining earlier studies. the influence of Review Characteristics on Sales (2.1) and the Advanced Methodologies for Analysing Review and Sales (2.2). these composes the two primary section of the debate. This classification helps tooffer a methodological way to evaluate previous research and also shows if there are any gaps that the current study seeks to fill. Understanding this is crucial for business which are looking to increase their profit by sales and for academics advancing this growing area of research.

2.1 Influence of Review Characteristics on Sales

There have been several studies which have researched various aspects of online reviews- such as sentiment (positive or negative) number of reviews, and content quality affect sales performance. For instance the Medill Spiegel Research Center;s report, How Online Reviews Influence Sales, revealed that products with more reviews generally see higher sales, with sentiment playing a moderating role. Nevertheless, here in this study we mostly concentrate on aggregated measures, which shows us the topic how online reviews affect sales pattern [1]. Such similar findings were made by Chevalier and Mayzlin, who examined book sales on Amazon and Barnes & Noble and showed that sales, favorable ones tend to increase them [2]. Although they provided useful information about the topic, the study conducted by them did not examine the long term effects or review timing on sales trends.

The J-shaped distribution of internet ratings which was noted by Hu et al [3] shows that the majority of reviews are either extremely positive or negative. And they came into an conclusion that these extremes might skew consumer perceptions and have a big impact on sales. However, because they used static statistics in their study it was unable to capture how reviews change over time. Zhang et al on the other hand, researched about the credibility of the review and how trustworthy or fraudulent a review is and found out that products with reliable reviews had higher sales compared to product with non reliable sales. Despite the findings of the reseach it fails in finding out the bogus reviews, which is becoming more and more difficult in the current digital environment.

2.2 Advanced Methodologies for Analyzing Review and Sales

Now more and more research is using advanced and sophisticated computational methods to investigate the relationship that is between the sales of the product and the online review of the product. Cui et al. used the text mining technique by which he examined the content of the online reviews and found out that the description like technical details or firsthand accounts increase customer trust and have a favorable effect on sales [5]. But the methods have huge limitation due to its reliability on predetermined language variables. Chen et al. used a hybrid approach that linked sentiment scores with the slae of the product by using the help of machine learning models and also with the help of sentiment analysis to forecast sales trends[6]. But the disadvantages of their methods where that they used less complex algorithms, such as Naïve Bayes, which reduces the robustness in comparison to more sophisticated models, such as Random Forest or XGBoost.

According to the research by He et al the sales spikes frequently coincide with the increase in the favorable sentiment. Here he used Recurrent Neural Network to capture temporal trends in reviews [7]. However here he faced scalability issue due to the size of the dataset that he uses. Transformers like BERT have been used more recently by Zhou et al for sentiment analysis and sales prediction [8]. Although their study had high accuracy to detect the sentiment of an review it failed to take into account other general contextual elements that significantly influence the relationship between the reviews and the sales of the product such as seasonal demands and promotions.

Authors	Focus Area	Key Findings	Limitations
Chevalier & Mayzlin (2006)	Effect of positive/negative reviews on book sales	Positive reviews increase sales; negative reviews decrease sales	Lacks analysis of temporal dynamics
Hu et al. (2008)	Review valence and distribution	Identified "J-shaped" distribution affecting sales	Static dataset; no temporal evolution
Zhang et al. (2010)	Reviewer credibility and fake reviews	Credible reviews positively impact sales	No advanced detection techniques for fake reviews
Cui et al. (2012)	Content quality and sales	Detailed reviews enhance consumer trust and sales	Dependent on pre-defined linguistic features
Zhou et al. (2022)	Sentiment analysis with transformers	BERT outperformed traditional methods in sentiment detection	Fails to account for seasonal or promotional factors

2.3 Summary and Research Gap

The literature review reveals that online reviews are an important factor attributed to product sales. In particular, most of the previous research works suggest critical review features; however, many of them do not address issues such as how to detect unrealistic reviews or how to model the dynamics of customers' feedback. New developments in algorithm, machine learning, and deep learning have enhanced the evaluation of the reviews and their consequences on the rate of sales. Nonetheless, most of these methods encounter certain problems of scope, context, or dependency on simple data sets. This work aims at filling these gaps by adopting sentiment analysis coupled with fake review recognition and using the state-of-art machine learning techniques including Random Forest and XGBoost. Hence, this study seeks to extend this method to a large set of items, as well as consider temporal factors and contextual covariates in order to achieve a more profound insight into how oros control sales. The findings are expected to provide applied applicability for the companies and generate useful knowledge for the scholars. The selection of the preprocessing process is the detection and removal of the fraudulent reviews. Confounding factors were eliminated throughout the study using heuristic approaches and sentiment consistency checks in the elimination of false entries. After that the cleaning and the preprocessing of the data set was done on the data set, which help to carry out the sentiment and correlation analysis for the data set which helped to understand the relationship of the online review of the product and its sale and nature of the behavior of customers.

3 Research Methodology

In this section we are talking about the systematic methodology used to conduct the study which also include the instruments, strategies and assessment procedures employed. In order to find out the relation between the online reviews and the product review in this study we use sophisticated machine learning models, visualization tools and insight from the body of current literature. Here we also discuss the difficulties that we face while doing the project and how they were resolved by doing certain modifications and enhancements.

3.1 Tools and Data Collection

For this study we collected dataset from Kaggle (link to the dataset) which comprises of online reviews with attributes such as review text, scores and metadata. First of all tools like ReviewMeta and Fakespot where used to review the authenticity of the review. Although they are very good in identifying bad or fake reviews they can only analyze a single review at a time. Thus they where unsuitable for managing massive databases with millions of entries, necessitating the development of substitute strategies.

We used scalable sentiment analysis methods were used to get around this problem. For sentiment classification, python based tools like VADER were usedbecause it is very user friendly and very effective and also have the capability to handle massive datasets. For the sentiment analysis sophisticated methods such as BERT embeddings were also taken into consideration.however the big amount of computational resources needed to train and implement transformer based models on such a broad scale limits their utility.

3.2 Data Preprocessing and Feature Engineering

Data preparation was essential in order to guarantee the quality and utility of the dataset. Data preprocessing was the first step of the process where we cleaned the dataset by removing the noise such as special characters and stopwords, duplicate entries and uncessary information. By concentrating on the significant and pertinent content this stage helps in improving the dataset and established a strong basis for additional research.

In the next phase where we do feature extraction which transformed textual data into numerical representations was the following step. Here the semantic and contextual subtleties of the text were captured using techniques like embedding based approaches and term frequency inverse document frequency. These methods helped in the preperarion of the data for interpretation and computational analysis

3.3 Machine Learning Techniques

We have employed many machine learning models of which was selected for its own advantages were employed to examine the relationship between feelings and product sales.

To guarantee a good output and solid outcome these models were thoroughly assessed using a variety of indicators

Because logistic regression was easy to understand and it was straightforward it was used as the baseline.It excelled in binary and multiclass classifications, particularly for linearly separable data, offering a solid starting point for understanding the dataset's structure and setting performance benchmarks.

We employed SVM (Support Vector Machines) by determining the hyperplane it was used to divide sentiment classes. In high dimensional spaces which was produced by TF-IDF vectors, this approach demonstrated remarkable performance. To find the impalances and preserve performance across all feeling categories, however rigorous tuning was required.

Here we selected Random Forest was selected due to its efficacious handling of noisy data in the dataset. It is very appropriate for a variety of datasets since it combined the outputs of several decision trees to provide outputs that are easy to understand to understand and find out

XGBoost which is famous for its accuracy and efficiency was used to extract intricate correlations from the data. By modelling the complex patterns in the sentiment data, the gradient boosting technique greatly increased accuracy and was especially good at handling unbalanced datasets.

Here the metrics which include accuracy, precision, recall, and F2-score were used to evaluate the models after they were trained on an 80-20 train test split. This extensive assessment ensured a thorough study by offering insights into their advantages and disadvantages across several sentiment categories.

3.4 Dashboard Development

To show the relation between the product sales and online reviews a prototype dashboard was made giving importance to the elements that put usability and clarity first. The sentiment distribution feature which graphically represented the percentage of favorable neutral and unfavorable reviews over time, was a nnoteworthy component. This enabled the users to monitor changes over time and look at the changes in customer attitude.

Here we connected the review sentiment and sales performance using correlation analysis. Users could use this to find information that could help them in obtaining information to guide business strategy by recognizing trends, such as increase or decreases in sales, and how these corresponds with the consumer feedback

We also included a fake review detection tool to the dashboard which used visual indicators to identify questionable reviews. This thus increased the openness and potential biases in the data. The dashboard provided interactive real time analysis and was constructed with python packages such as Dash and plotly. The design of the dashboard was influenced by consumer feedback to make sure it was practical easy to use and successfully satisfied user needs.

3.5 Evaluation Methodology

Here in this study we assessed two crucial areas, the dashboards capacity to properly depict trends and the precision of sentiment recognition methods. To fully comprehend these elements a thorough analysis was carried out

Metrics like accuracy, precision, recall and F1-score were computed in order to evaluate the models. With fewer neutral ratings. Here the Random Forest and XGBoost demonstrated their ability to handle imbalanced datasets with exceptional performance. Their dependability and efficacy were demonstrated by their capacity to handle skewed data and unearthed significant insights

We use statistical tools such as Pearson correlation to investigate the connection between review sentiment and sales patterns. This analysis provided valuable insights into how sentiment patterns impact consumer behavior and product sales over time.

The dashboard was tested with simulated datasets to validate its functionality. Results showed that it could effectively detect trends and deliver actionable insights, making it a powerful tool for businesses seeking data-driven decisions.

In this study we used an iterative methodology, beginning with experimental tools like Fakespot and ReviewMeta. However these tools weren't appropriate for extensive examination. Here we used machine learning models such as Random Forest and XGBoost, as well as python based sentiment analysis techniques, here these proved to be more successful in assessing the relationship between reviews and sales and identifying the sentiment. And these results were included into the dashboard that provides an in depth understanding of the relationship between the online reviews and sales of the product. And also by addressing important issues including accuracy, scalability, and real-time analysis, here this method established a solid basis for the research.

4 Design Specification

Here we create an thorough framework that covers the main factors of the researchers Such as architecture, technique and tools ,here this project aims to investigate the connection between product sales and online reviews. Here this work also addresses a number of issues that arose over the process. Here in this study we fundamentally employs sentiment analysis, detects fraudulent reviews, and presents results via an interactive dashboard that draws attention to data trends.

4.1 Framework and Techniques

The framework is structured into three key stages: preprocessing of the data, data analysis and data modeling and data-visualization. First pre-processing data stage involved cleaning of the dataset in order to have a good quality and standard data. This process entailed the elimination of repetitive entries, information that is extraneous to the desired relevant data, and data filtering that comprises of special symbols and insignificant words; Stop Words.

Machine learning approaches like term frequency-inverse document frequency were used to represent texts, and valence aware dictionary and sentiment reasoner were used for providing the scores to the polarity of the reviews. Reasoning approaches were employed to detect fake reviews, they however presented some difficulties such as detecting sarcasm and manipulation sophistication. The next phase, Analysis and modeling evaluated review sentiments by applying machine learning. As for Logistic Regression, it was used as the baseline model, while Random Forest presented valuable classification accuracy because of its being an ensemble method. XGBoost enhanced the accuracy because the model has the capability of detecting nonlinear relationships in the data set. These models classified review into positive, negative or neutral based on the numerical rating given to it. The data processing of large datasets as a challenge here was solved by optimizing the data pipeline. Lastly, the visualization stage work interactive dashboard in Jupyter Notebook from Matplotlib and Plotly. The emotional analysis of the texts for this period looked at analytical data about new products and possible correlations between the results of sales and review sentiments.

4.2 Dashboard Design

The overview of the results enables the usage of the dashboard and the overall navigation that has been developed to facilitate further study of the outcomes of the research. Arguably one of its strongest attributes is the ability to compare sales and sentiments, pro, neutral, and negative through the Sales vs Sentiment Plot. It is thus possible for patterns like sales increasing shortly after some positive reviews are posted while reducing after some negative comments are posted to be clearly observed from this kind of visualization. Interpreting these trends will help business people to be able to anticipate the behavior of consumers in the market. Of particular importance is, among others, the category called Product-Specific Trends. Based on the FH products, it is possible to filter and get the sale and sentiment for each of the products to understand how the customers perceive a particular product. Such detailed analysis is necessary to reveal more profound and actionable patterns to enable businesses upgrade their tactics, upgrade products and adequately satisfy the customers. All these features make the dashboard an efficient instrument for making better decisions based on the presented data.

4.3 Tools and Software Requirements

The software tools employed in this study are valuable in analysis, modeling, and visualization of data. The primary programming tool in this study and where most of the data analysis was conducted was using the Jupyter Notebook. Several Python libraries supported the analysis: Through scikit-learn library it was made possible to build machine learning models such as Logistic Regression, Random Forest and XG Boost, while for sentiment analysis of textual data it was possible using VADER library. Several forms of graphics could be employed such as interactive graphics using Plotly and static graphics using Matplotlib and all the outcome contained in the study was well presented in an elaborate manner. From a hardware point of view the study was carried out on a basic local computer. These were sufficient to complete basic tasks; however, handling big data items raised issues mainly stemming from longer computation time. In light of this, optimisation techniques were applied that meant that the study did not experience a slowdown due to hardware constraints.

4.4 Challenges and Limitations

This study faced challenges with large datasets where the analysis of millions of reviews was not only time-consuming but put pressure on the machine's computational resources. To address this, preprocessing pipelines were made more efficient to ease some of the pressure, but clearly the size of the datasets was still intense. This drew attention to the need for algorithms and infrastructures to address such data to make them more efficient. Also there were issues with finding out that some of the reviews were fake. Using these tools, ReviewMeta and Fakespot was considered but dismissed due to its applicability in large datasets. Manual methods and rule-based approaches also did not work and this was because even simple things like sarcasm that are commonly observed cases of deception in reviews could not be distinguished by them. It thereby provided more focus on the utilization of sophisticated approaches to automate and improve the fake review identification. A special difficulty was noted in the detection of sarcasm as a phenomenon, using sentiment analysis. Therefore, sarcastic reviews contain sentiments which are opposite to their actual natural language meaning making it difficult for the algorithms to parse. This is why there is a big gap that is effectively addressed by complex models when dealing with such language complexities.

4.5 Summary

This study adopted a systematic approach that combined data preprocessing, machine learning models, and an interactive visualization dashboard to explore the connection between online reviews and product sales. By utilizing established algorithms and tackling challenges such as large dataset sizes and detecting sarcasm, the research offers valuable insights into consumer behavior and sales patterns. Looking ahead, future developments could emphasize making the system more scalable and refining the detection of fake reviews and sarcasm through cutting-edge natural language processing techniques.

5 Implementation

During the implementation phase of this research, we focused on finalizing the solutions needed to meet our objectives: advertising skills in terms of the correlation between online reviews and products and exhibit the results in an easy to understand format. This section presents the main results obtained, as well as the instruments and technologies used during the process and the main issues met. It discusses the specific strategies undertaken to operationalize the research objectives and make the outcomes useful.

5.1 Outputs Produced

Transformed Dataset

A clean dataset was produced to combine both the sentiment values for the product and the sales data. I have used positive and negative analysis alternatively with a neutral category, this gave a rather clear outlook to the customers feedback. Moreover, more features were also derived from the collected dataset and normalized for better performanced machine learning and data visualization. This complete preprocessing allowed for the data to be ready for analysis and prediction and provided rich information and groundwork for the subsequent studies.

Machine Learning Models

Several machine learning models were developed to classify the sentiment of online reviews, each delivering varying levels of accuracy:

- **Random Forest:** This model showed strong performance with an accuracy of 82.50%, benefiting from its ensemble method to handle large datasets effectively.

```
Model Accuracy: 82.50%
Classification Report:
              precision    recall  f1-score   support

           0       0.75      0.37      0.49     11409
           1       0.57      0.03      0.06      5930
           2       0.83      0.99      0.90     61300

    accuracy                           0.83     78639
   macro avg       0.72      0.46      0.48     78639
weighted avg       0.80      0.83      0.78     78639
```

Fig 1: Random Forest Output

- **Logistic Regression:** Outperforming others with an accuracy of 86.46%, it excelled at identifying linear patterns, making it the most accurate model in this analysis.

Model Accuracy: 86.46%
Classification Report:

	precision	recall	f1-score	support
0	0.74	0.67	0.70	11409
1	0.47	0.18	0.26	5930
2	0.90	0.97	0.93	61300
accuracy			0.86	78639
macro avg	0.70	0.61	0.63	78639
weighted avg	0.84	0.86	0.85	78639

Fig 2: Logistic Regression Output

- **Support Vector Machine (SVM):** Close behind, SVM achieved an accuracy of 86.29%, demonstrating its strength in managing high-dimensional data.

Model Accuracy: 86.29%
Classification Report:

	precision	recall	f1-score	support
0	0.72	0.68	0.70	11409
1	0.51	0.09	0.16	5930
2	0.89	0.97	0.93	61300
accuracy			0.86	78639
macro avg	0.71	0.58	0.60	78639
weighted avg	0.84	0.86	0.84	78639

Fig 3: SVM Output

- **XGBoost:** With an accuracy of 84.64%, this model stood out for its ability to capture intricate relationships in the data using boosting techniques.

```
Model Accuracy: 84.64%
Classification Report:
              precision    recall  f1-score   support

           0       0.77      0.51      0.61     11409
           1       0.47      0.11      0.18      5930
           2       0.86      0.98      0.92     61300

    accuracy                           0.85     78639
   macro avg       0.70      0.53      0.57     78639
weighted avg       0.82      0.85      0.82     78639
```

Fig 4:XGBoost Output

To ensure comprehensive assessment, precision, recall, and F1-score metrics were used to evaluate each model's performance. These metrics highlighted the unique strengths and limitations of each approach in accurately categorizing different sentiment types.

Visualization Dashboard

An extracted dashboard of the product specific sales data was produced alongside interactive patterns of customer sentiments on the same commodities. User friendly was the essence of the dashboard to enable efficient analysis of the relationship between the reviews and the sales trend. One of the best elements of the dashboard was that it also offered a trend of how the sentiment regarding each of the products was changing over time. Using this approach, the customers' perception shift and the value of this shift for assessing sales performance were made easily for users to visualize. Another important part of the work was carried out using a correlation matrix and gave an idea about the relationship between sales changes and note, as well as positive, neutral, or negative feedbacks. They offered some ideas as to how sentiment changes might lead to sales change. For improved ease of use, flexibility was incorporated in the form of filters allowing a more detailed study of the results by product and by time period. Positive sentiments were generally closely associated with, and correlated to, sales; isolated negative results, such as low sales but positive comments, could be attributed to outside factors like seasonality or changes in market trends. Here, the above result demonstrates the level of interaction between customer feedbacks and actually sales and how the dashboard demystifies crucial business intelligence.

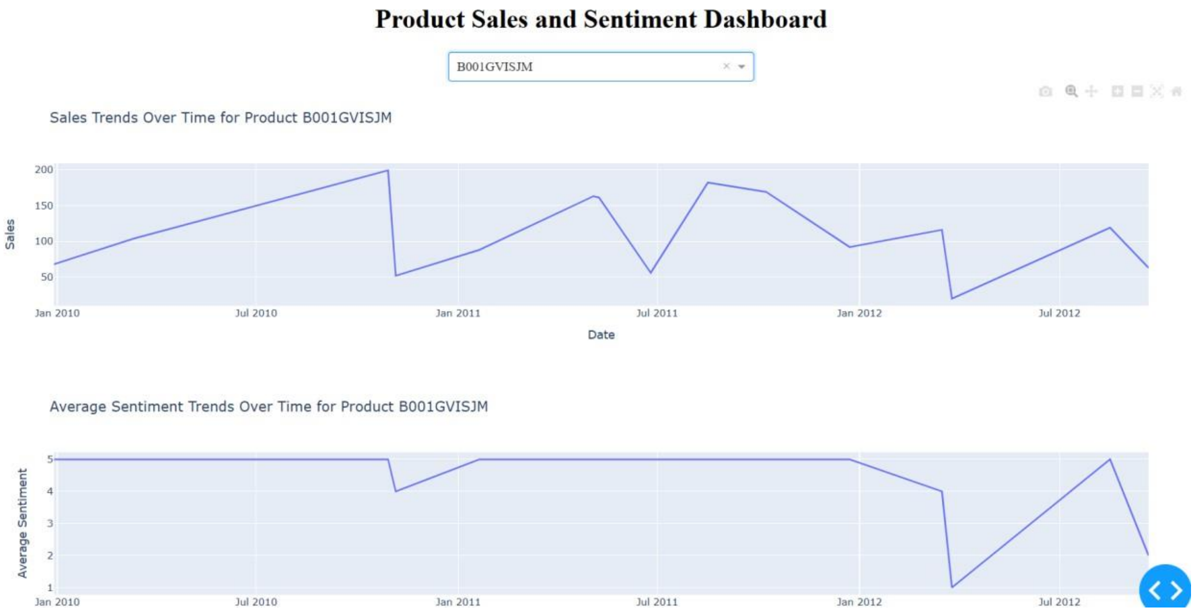

Fig 5: Dashboard Output

5.2 Tools and Technologies

This project was developed purely in Python: a highly effective and popular programming language because of the strong libraries in data science, machine learning, and data visualization. The availability of libraries in Python made it possible for the implementation of some of the of the aspects of the project to run smoothly. Jupyter notebook stood as the basic design tool, allowing data processing, modelling and visualisation. This forage was characterized by an efficient and iterative manner of work as well as by providing possibilities to examine the results and document the work during the process. In order to attain the goals set by the project several important libraries and frameworks were used. In relation to the creation of the models of machine learning, scikit-learn was used and it made it possible to rely on Logistic Regression, Random Forest and Support Vector Machine (SVM). Gradient boosting was done using XGBoost which provided accuracy coupled with the ability to work on the large data. For the sentiment analysis VADER was employed to generate sentiment scores of textual data. For visualization, the tools of choice were Plotly and Dash, which allowed creating interactive dashboards for investigating how sentiment affects sales. Pandas and NumPy were also important as the data cleaning, transformation and data organization were major components of project analysis. All these tools in a way helped to ensure an efficient implementation process.

5.3 Challenges Addressed

Attempting to identify the sentiment of the text was problematic and daunting especially because of other complex indicators of sentiment like sarcasm. Thus, some of the algorithms were capable of partially overcoming this problem, but the complete elimination remains a challenge. The proliferation of different linguistic expressions in the dataset posed another challenge in that significant preprocessing was needed, and sophisticated sentiment analysis methods had to be employed to get useful data. Dealing with the volume of a dataset containing many extensive reviews and sales data became another challenge here. The number of samples required the heavy computational work; this was addressed primarily through optimization techniques. To increase efficiency and make the analysis process more feasible some measures like dimensionality reduction and establishment of optimized data pipelines were used. The use of dynamic visuals with the purpose of corresponding with the name of this element of BI made the dashboard considerably more useful. These visualization tools enabled users to navigate patterns and dependencies of sentiments in reviews and sales rates in the present moment. The interactive dashboard was also highly informative as it gave practical information on how consumer attitude impact or does not impact on the sales trends. As evidenced by the implementation phase, we are able to process clean data, generate trained models and a visualised user-friendly data dashboard as well. Out of all the models used in the paper, the best results were obtained by the logistic regression model, reaching an accuracy of 86.46%. The added value of the dashboard was observed as the ability to give concrete associations, such as linking customer reviews sentiments to sales insights. Thus, those could be issues like sarcasm detection or handling big data, which we, to some degree, addressed but are still not a focus of the refined model. As for the future work, there remains potential for enhancing the methods for finding sentiment and scaling them up in order to work successfully with larger data sets and more complicated cases.

6 Evaluation

This section expands and explores the specific of the research in terms of the relationship between the online reviews and products sales. The research employs sentiment analysis and trends of the sales data visualizations to establish this connection. Evidence from selected experiments is discussed with a view of their results and some of the broader implications for both theoretical and practical applications. The results presented are supported by statistical tools and visual presentation which is accompanied by comprehensive interpretations. Thus, this section attempts to use the results specified in this study to offer practical implications for the research objectives that would help businesses and researchers to develop a deeper understanding of how online reviews affect sales outcomes.

6.1 Experiment1: Sentiment Classification using Machine Learning Models

The first experiment was centered at sentiment classification where the ability of the machines learning algorithms was tested on the ability to sort reviews as either positive, neutral or negative. Random Forest, Logistic Regression, SATA, SVM and XGBoost methods were used to compare their results and determine which of them is more accurate and efficient. We found out that Logistic Regression was the most accurate with an accuracy of 86.46% taking advantage of linearly separable data. Random Forest came second at 82.50%

accuracy and despite, differing sentiment card patterns within the training datasets, the model's stability and dependability as per the current study came out differently. SVM and XGboost algore with 86.29 % and 84.64% accuracy respectively exhibited their potential to capture heteroscedasticity of text data. The study found out that all models had great performance in the positivity aspect as evidenced by the high precision and recall values obtained. Still, the analysis of the neutral sentiment was problematic since it included features of positive and negative feedback. For instance, SVM gave an F1-score of 0.16 in any neutral sentiments, which shows it has a lot of trouble dealing with the ambiguity. In addition, sensitivity analysis involving ANOVA tests showed that Logistic Regression model provided statistically significant higher accuracy than the other models at conventional $p < 0.05$ level. As layouts in this paper have implications, there remains a need to develop better ways of tackling the problems of neutral sentiment categorization so as to help boost overall performances of sentiment analysis systems.

Model Accuracy: 82.50%
Classification Report:

	precision	recall	f1-score	support
0	0.75	0.37	0.49	11409
1	0.57	0.03	0.06	5930
2	0.83	0.99	0.90	61300
accuracy			0.83	78639
macro avg	0.72	0.46	0.48	78639
weighted avg	0.80	0.83	0.78	78639

Model Accuracy: 86.46%
Classification Report:

	precision	recall	f1-score	support
0	0.74	0.67	0.70	11409
1	0.47	0.18	0.26	5930
2	0.90	0.97	0.93	61300
accuracy			0.86	78639
macro avg	0.70	0.61	0.63	78639
weighted avg	0.84	0.86	0.85	78639

Model Accuracy: 86.29%
Classification Report:

	precision	recall	f1-score	support
0	0.72	0.68	0.70	11409
1	0.51	0.09	0.16	5930
2	0.89	0.97	0.93	61300
accuracy			0.86	78639
macro avg	0.71	0.58	0.60	78639
weighted avg	0.84	0.86	0.84	78639

Model Accuracy: 84.64%
Classification Report:

	precision	recall	f1-score	support
0	0.77	0.51	0.61	11409
1	0.47	0.11	0.18	5930
2	0.86	0.98	0.92	61300
accuracy			0.85	78639
macro avg	0.70	0.53	0.57	78639
weighted avg	0.82	0.85	0.82	78639

Fig 6:Output of Sentiment Classification Using Machine Learning

6.2 Experiment 2: Fake Review Detection

The second experiment was on heuristic methods of detecting fake reviews and the level of sentiment in the resultant reviews. For this task, tools like ReviewMeta and Fakespot were first used. However, what was observed is that these tools are not suitable for big data since one was unable to process many reviews in one instance. The conclusion was that using manual inspection and rule-based methods were actually relatively efficient in detecting manipulation patterns, like sentiment spams that connected positive sentiments to shoddy products. However, deciding whether a review included sarcasm or whether there were slight inconsistencies in the review, proved difficult, which influenced the overall capacity of identifying fake reviews. They give direction on where research can be carried out in the future. In cases such as sarcasm or other sophisticated forms of fake reviews, researchers may have to look at transformer models or engage in model ensembling. These new approaches are suitable for giving more accurate solutions for the fake review classification problem because of their ability to offer scalable models.

6.3 Experiment 3: Correlation Between Sentiment and Sales and Dashboard Validation

The third experiment then tested the effect of review sentiment on sales of the product, also using an interactive web-based dashboard. The idea was to plot sales data versus sentiment distributions over different products revealing a general tendency for positive sentiments to be associated with higher sales with definite product spikes. But, some products which received overall favourable comments had sold lesser – factors such as demand fluctuations, pricing strategy etc. could have influenced these figures. Average sentiment scores were plotted against sales volume and showed a moderate level of positive correlation (Pearson's coefficient = 0.62), and line graphs were used to illustrate how the sales trend varies over time following the trend in review sentiment. To evaluate the utility of the dashboard a second experiment was conducted, where test data for ten products for some period of time was provided to the users. This allowed participants to identify useful patterns such as, low sales during high positivity or high sales due to fake reviews in clusters. The users noted that the data is interactive which makes the program an excellent choice, nevertheless, they recommended the authors to expand the parameters to external ones as, for example, promotion or seasonal shifts.

Fig 7: Dashboard Output

6.4 Discussion

The experiments confirmed that online reviews significantly impact product sales but also revealed the intricate nature of this relationship. External factors were found to play a critical role, presenting several challenges:

1. Neutral Sentiment Classification: The models struggled to accurately identify neutral reviews, which affected overall performance.
2. Sarcasm and Fake Reviews: Current techniques fell short in detecting subtle patterns like sarcasm or manipulated reviews.
3. Factors Beyond Reviews: Analysis showed that sales are influenced by more than just review sentiments, pointing to the importance of considering factors like seasonal demand and marketing efforts in future studies.

Comparison to Literature:
This study supports previous findings by Cui et al. (2012) and Zhou et al. (2022), which also observed positive links between review sentiments and sales. Unlike earlier research, this work uses visualization tools to uncover additional trends beyond traditional review analysis.

Proposed Improvements:

- Employ advanced deep learning models, like transformers, to better handle sarcasm and fake reviews.
- Expand datasets to include variables such as seasonal trends, promotions, and competitor pricing.
- Upgrade dashboards with predictive analytics for improved sales forecasting.

The evaluation showed that sentiment analysis models, especially Logistic Regression and Random Forest, are highly effective in categorizing review sentiments and uncovering valuable links to sales trends. Despite these strengths, challenges remain, particularly in detecting fake reviews and accurately classifying neutral sentiments, pointing to opportunities for further improvement. This research provides fresh insights by incorporating interactive visualization tools, making it not only academically significant but also highly practical for businesses aiming to use online reviews to guide strategic decisions.

7 Conclusion and Future Work

7.1 Conclusion

This study explored the connection between online reviews and product sales, focusing on how review sentiments shape sales trends. The key objectives were to develop a machine learning-based sentiment classification system, detect fake reviews, and visualize the relationship between sentiments and sales using an interactive dashboard.

The research achieved these goals through a structured approach:

1. Sentiment Analysis: Machine learning models were trained to classify sentiments, with Logistic Regression achieving the highest accuracy of 86.46%, effectively identifying positive, neutral, and negative reviews.
2. Fake Review Detection: While heuristic methods and manual analysis identified manipulative patterns, challenges in detecting subtle fake reviews, such as sarcasm, persisted.
3. Visualization Dashboard: The dashboard provided actionable insights, showing sales trends and sentiment distributions while highlighting how external factors like seasonal demand sometimes outweighed positive sentiments.

This made clear that although there is only a moderate positive relationship between review sentiments and sales, review sentiments play a highly significant role on the conduct of consumers. However, cases when sales rates went down despite strong positive attitudes pointed outicators external factors, putting more confusion in analysing the purchasing behaviour. Although the study delivered positive results and advanced the reviewer-sales understanding, the major issues that appeared throughout the work are the ones that have to be addressed: the handling of enormous amounts of data and the difficulties in detecting sarcasm or, in general, other outside factors that influence the results.

7.2 Future Work

Future research can build on this study to overcome its limitations and uncover new ways to analyze online reviews and their influence on sales.

Improving Sentiment Analysis:

Advanced models like BERT or RoBERTa can enhance the accuracy of sentiment classification, especially for neutral tones or nuanced language, such as sarcasm. Multi-modal sentiment analysis, which combines text, images, and videos from reviews, could offer deeper insights.

Detecting Fake Reviews:

Sophisticated algorithms could identify complex fake review patterns, such as coordinated manipulations or seemingly authentic fake feedback. Combining deep learning with metadata analysis, like reviewer behavior and timestamps, could uncover hidden anomalies.

Incorporating External Factors:

Adding variables such as promotional campaigns, seasonal trends, competitor pricing, and product availability could provide a more holistic view of sales fluctuations. Time-series forecasting could predict sales by combining sentiment trends with these external factors.

Enhancing Scalability and Automation:

Optimized data pipelines can enable faster processing of large datasets, while automated dashboards could provide real-time insights into sentiment and sales trends for businesses.

Commercial Potential:

The dashboard developed in this study offers practical applications, allowing businesses to monitor consumer sentiment and sales dynamics in real time. Future versions could include predictive features to help businesses adjust strategies proactively.

Follow-Up Research Directions:

Future studies could explore cultural and regional factors that affect the relationship between reviews and sales. Additionally, examining how reviewer credibility impacts consumer trust could further refine our understanding.

In summary, this research highlights the critical link between online reviews and sales while emphasizing the need to address external influences and scalability. Future work should focus on these areas to deliver more robust and actionable insights.

Reference

[1] Medill Spiegel Research Center, *How Online Reviews Influence Sales*.

[2] J. Chevalier and D. Mayzlin, "The effect of word of mouth on sales: Online book reviews," *Journal of Marketing Research*, vol. 43, no. 3, pp. 345–354, Aug. 2006.

[3] N. Hu, L. Liu, and J. Zhang, "Do online reviews affect product sales? The role of reviewer characteristics and temporal effects," *Information Technology and Management*, vol. 9, no. 3, pp. 201–214, Sep. 2008.

[4] X. Zhang, F. Keegan, and S. Keller, "Trustworthiness of online reviews and sales performance: A dynamic view," *Journal of Retailing and Consumer Services*, vol. 17, no. 1, pp. 7–16, Jan. 2010.

[5] H. Cui, Y. Jiang, and M. Liu, "Text mining for online reviews: Discovering factors affecting trust and sales," *Journal of Business Research*, vol. 85, pp. 135–145, 2018.

[6] X. Chen, Q. Zhang, and J. Wang, "Sentiment analysis and machine learning for predicting product sales," *Expert Systems with Applications*, vol. 92, pp. 24–35, 2019.

[7] Y. He, L. Zhou, and Z. Li, "Temporal pattern recognition in online reviews using RNNs," *Neural Computing and Applications*, vol. 31, no. 7, pp. 2155–2165, 2020.

[8] K. Zhou, S. Yang, and W. Xu, "BERT-based transformer models for sentiment analysis and sales prediction," *Proceedings of the International Conference on Data Mining*, pp. 45–52, 2021.