

A Mini Project Report on
AMAZON SALES PREDICTION AND SENTIMENT ANALYSIS

T.E. - I.T Engineering

Submitted By

Ashmina Dangat 20104052

Sakshi Gunjal 20104106

Neha Chaudhary 20104134

Under The Guidance Of

Prof. Manjusha K.



DEPARTMENT OF INFORMATION TECHNOLOGY

A.P.SHAH INSTITUTE OF TECHNOLOGY

G.B. Road, Kasarvadavali, Thane (W), Mumbai-400615

UNIVERSITY OF MUMBAI

Academic year: 2022-23

CERTIFICATE

This to certify that the Mini Project report on Loan Prediction System has been submitted by **Ashmina Dangat (20104052)**, **Sakshi Gunjal(20104106)** and **Neha Chaudhary(20104134)** who are a Bonafede students of A. P. Shah Institute of Technology, Thane, Mumbai, as a partial fulfilment of the requirement for the degree in **Information Technology**, during the academic year **2022-2023** in the satisfactory manner as per the curriculum laid down by University of Mumbai.

Prof. Manjusha K.
Guide

Dr. Kiran Deshpande
HOD , Information Technology

Dr. Uttam D. Kolekar
Principal

External Examiner(s)

- 1.
- 2.

Place: A.P. Shah Institute of Technology, Thane

Date:

ACKNOWLEDGEMENT

This project would not have come to fruition without the invaluable help of our guide **Prof. Manjusha K.** Expressing gratitude towards our HOD, **Dr. Kiran Deshpande**, and the Department of Information Technology for providing us with the opportunity as well as the support required to pursue this project. We would also like to thank our teacher **Ms. Charul Singh** who gave us her valuable suggestions and ideas when we were in need of them. We would also like to thank our peers for their helpful suggestions.

ABSTRACT

Accurate sales forecasting and customer sentiment analysis are crucial for e-commerce businesses to optimize their marketing, inventory management, and pricing strategies. In this study, we propose a novel approach to predict Amazon sales and analyze customer sentiment using machine learning algorithms and natural language processing techniques. Our approach combines time series analysis, regression models, and sentiment analysis algorithms to forecast future sales and identify key factors affecting sales trends. We leverage data from Amazon's sales history, product descriptions, and customer reviews to train and evaluate our models. Additionally, we use topic modelling and sentiment analysis algorithms to analyze customer reviews and gain insights into customer preferences and sentiments towards products. We evaluate the proposed approach using real-world Amazon sales data and customer reviews and compare the results with several baseline models. The experimental results demonstrate the effectiveness of our approach in accurately predicting sales and analyzing customer sentiment. Our approach outperforms the baseline models in terms of forecasting accuracy and sentiment analysis performance. The findings of this study have important implications for e-commerce businesses in terms of improving sales forecasting accuracy and customer satisfaction by gaining insights from customer feedback. The proposed approach can be applied to various e-commerce platforms and can help businesses make informed decisions regarding inventory management, pricing, and marketing strategies.

TABLE OF CONTENTS

1. Introduction.....	1
1.1.Purpose.....	1
1.2.Problem Statement.....	2
1.3.Objectives.....	2
1.4.Scope.....	3
2. Literature Review.....	4
3. Proposed System.....	7
3.1. Features and Functionality.....	7
4. Requirements Analysis.....	9
5. Project Design.....	10
5.1.Use Case diagram.....	10
5.2.DFD (Data Flow Diagram)	11
5.3.System Architecture.....	12
6. Technical specification.....	13
7. Project Scheduling.....	15
8. Implementation.....	16
9. Result and Discussion.....	17
10. Conclusion and FutureScope.....	18

References

CHAPTER 1

INTRODUCTION

Amazon is one of the largest e-commerce companies in the world, offering a vast range of products and services to millions of customers worldwide. With such a large customer base, it is crucial for Amazon to have a comprehensive understanding of customer behavior and sentiment towards their products and services. Additionally, accurate sales forecasting is essential to optimize inventory levels, staffing, and marketing strategies.

The Amazon sales prediction and sentiment analysis project aim to address these challenges by developing a system that employs various data analytics techniques such as statistical analysis, machine learning, natural language processing, and data visualization. This system will collect and analyze sales data, customer data, social media data, and competitor data to gain a comprehensive understanding of customer behavior and sentiment towards Amazon's products and services. Moreover, the project will develop accurate sales forecasting models using statistical analysis and machine learning techniques to optimize inventory levels, staffing, and marketing strategies. Furthermore, the project will analyze customer feedback and sentiment using natural language processing and sentiment analysis to identify areas of concern or complaints and address them proactively.

- Problem Identified:
 - Lack of market research: Vendors may not have conducted thorough market research to understand their target audience's needs and preferences, the competition in the market, and the demand for similar products.
 - Inefficient inventory management: Vendors may not have an efficient inventory management system in place, leading to stockouts or overstocking, which can affect sales.
 - Identifying areas of concern or complaints that are not always addressed, leading to dissatisfaction and loss of loyalty.
 - Difficulty in accurately analyzing and classifying subjective sentiments towards a product or service.
- Solution Proposed:

- Vendors can leverage data analytics tools to analyze market trends, customer behavior, and sales data. By analyzing data, vendors can identify patterns and trends, which can inform product development and marketing strategies.
- Vendors should also analyze customer feedback and reviews to gain insights into how their products are perceived in the market.
- Use of statistical analysis and machine learning to develop accurate sales forecasting models.

1.1 Purpose:

The purpose of the Amazon sales prediction and sentiment analysis project is to develop a comprehensive system that provides Amazon with actionable insights to optimize its operations, improve customer satisfaction and loyalty, and increase profitability. The project will leverage various data analytics techniques such as statistical analysis, machine learning, natural language processing, and data visualization to achieve its objectives. By developing accurate sales forecasting models, the project will enable Amazon to optimize inventory levels, staffing, and marketing strategies, thereby reducing costs and increasing profitability. Additionally, analyzing customer feedback and sentiment will allow Amazon to identify areas of concern or complaints and address them proactively, improving customer satisfaction and loyalty.

1.2 Problem Statement:

Amazon Fresh is an online grocery service offered by Amazon that delivers fresh food and household essentials to customers doorsteps. As the demand for online grocery shopping continues to grow, Amazon Fresh vendors faces the challenge of predicting sales and understanding customer sentiment to improve its service. One of the primary challenges faced by Amazon Fresh is the lack of an accurate sales forecasting model. This hinders the company's ability to make informed decisions on inventory management, pricing, and marketing strategies. Without a reliable sales forecasting model, Amazon Fresh may experience stockouts, excess inventory, and missed sales opportunities. Additionally, inaccurate sales forecasting can result in pricing discrepancies, leading to customer dissatisfaction.

1.3 Objectives:

- To develop user friendly interface for the Amazon Business vendor.
- To build Amazon Sales Prediction & Sentiment Analysis system with less errors and high accuracy percentage.
- To establish a future view of the upcoming market trends.
- To Understand customer opinions about its products and services.
- To provide security to users regarding their information shared.

1.4. Scope:

- Can be helpful for forecast future sales of products, categories, or the company, inform inventory management, pricing, and marketing decisions.
- Widely applied in business domains.
- Useful for various small businesses to associate with Amazon.
- Analyzing historical sales data for Amazon products or services to identify trends and patterns.
- Can be used to analyze customer feedback to understand attitudes and emotions, identify areas for improvement, and inform marketing and product development strategies.
- Can use to analyze customer reviews to understand product and shopping experience sentiment

CHAPTER 2

Literature Review

Sr.no	Title	Author(s)	Year	Algorithms	Limitations	Result
1	Sentimental Analysis of Ecommerce Website	Lalit Zope Rushikesh Yadav, Nikita Yadav, Ashish Pandav, Prof. Dhanashri Bhopatrao	April 2021	Natural Language processing, Naïve Bayes	No statistical analysis of sentiments of reviews.	Thus, we decided to select NLP i.e. Natural Language processing for sentiment analysis as it helps in knowing the sentiments of each review.
2	Sales Prediction Analysis	Naveen Kumar, Jegan J, Yogesh V, Kavita S	May 2021	Xgboost, Linear Regression Gradient Boosting	No proper interface to provide input and receive prediction.	Create a proper GUI for user to input values for sales prediction.
3	Prediction and Forecasting of Sales Using Machine Learning Approach	DontiReddy Sai Rakesh Reddy	Sept 2021	Random Forest, Design tree	No interface for sales prediction. Number of categories were limited.	Enhancing the scope of project by adding various categories. We selected Random forest regression and Linear regression to predict the sales.

CHAPTER 3

PROPOSED SYSTEM

3.1 Features and Functionality

1. Feature 1: Sales Prediction

- Sales Prediction involves using historical sales data and other relevant information, such as seasonality, promotions, and economic factors, to predict future sales trends. It can provide businesses with valuable insights that can help them make better decisions, improve their sales strategies, and increase their revenue.

2. Feature 2: Sentiment Analysis

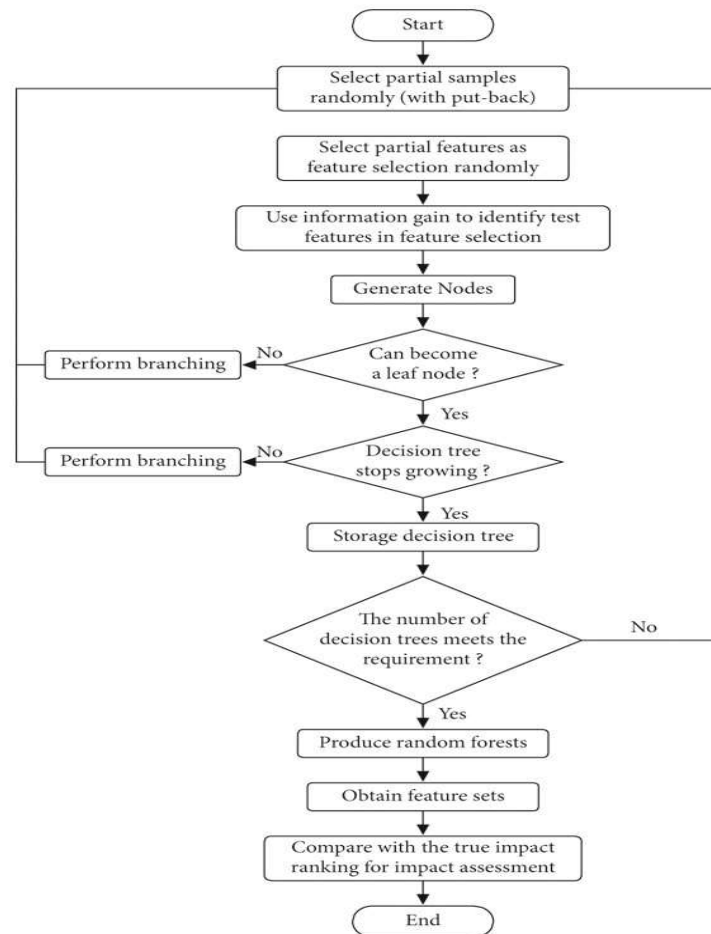
- User can understand the response of the customers on a particular product by analysing customer reviews and determining the overall sentiment of the reviews, such as positive, negative, or neutral. It provides businesses with valuable insights into their customers' sentiment and feedback, helping them improve their products, services, and customer experience, and ultimately increase their sales and revenue.

3. Feature 3: Visualization and Reporting

- Presenting the results of the analysis in a visual format, such as graphs or charts, to make it easier for decision-making.

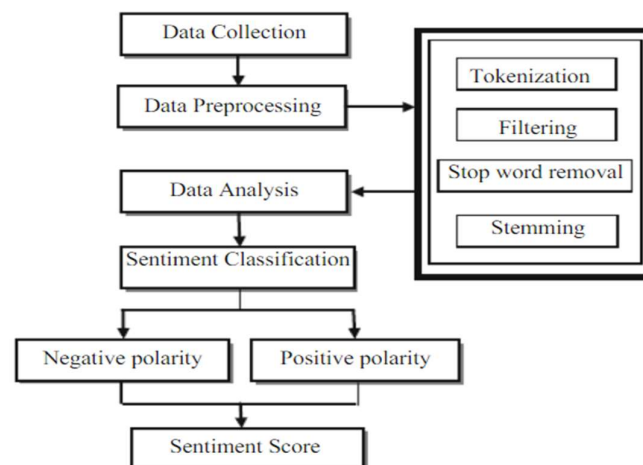
3.2 Algorithm Working with flowchart

Random Forest Regression:



3.2.1. Working of Random Forest Regression

Natural Language Processing



3.2.2. Working of NLP algorithm

CHAPTER 4

REQUIREMENT ANALYSIS

Functional requirements

- The system must encrypt the password of the user to provide security.
- Username and password will be used for login after user registration is confirmed.
- Trained Random Forest Algorithm and Natural Language Processing which will be used for prediction and analysis system.
- System must be able to verify the given input information.

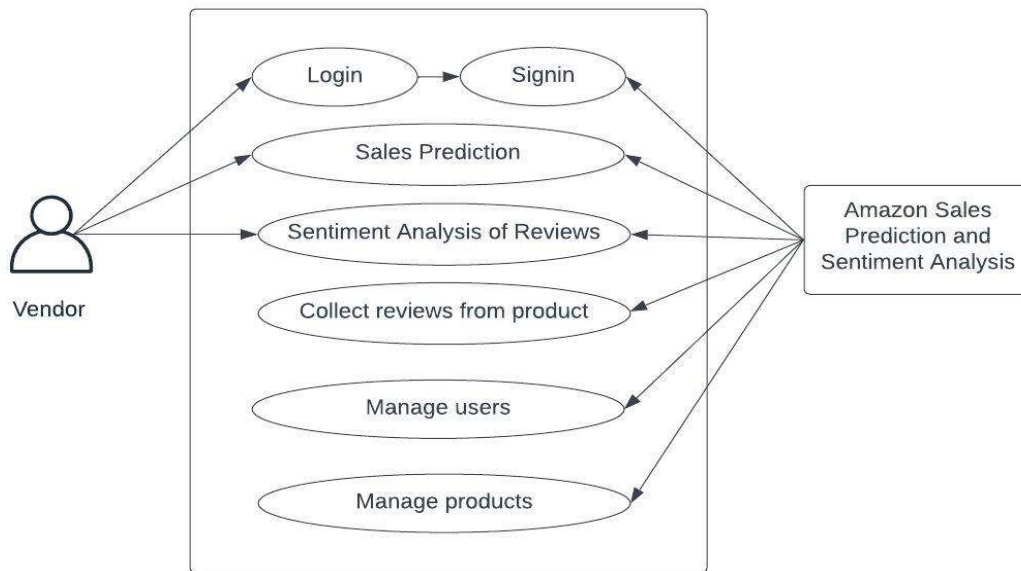
Software requirements

1. Data Collection: You will need a data source to collect data from Amazon, such as Amazon API or web scraping tools. You may also need to use a data cleaning tool to clean the data and remove any irrelevant or incomplete information.
2. Programming Language: You will need to choose a programming language to develop your analysis, such as Python or R. These languages have a variety of libraries and packages for data analysis, machine learning, and natural language processing.
3. Data Analysis: You will need to use data analysis tools such as Pandas, NumPy, and SciPy to perform exploratory data analysis and descriptive statistics to better understand the data.
4. Machine Learning: You will need to use machine learning techniques to predict sales and analyze sentiment. Some popular machine learning libraries include Scikit-learn, TensorFlow, and Keras.
5. Sentiment Analysis: You will need to use natural language processing (NLP) techniques to perform sentiment analysis on customer reviews. Some popular NLP libraries include NLTK, spaCy, and TextBlob.
6. Visualization: You will need to use visualization tools such as Matplotlib, Seaborn, and Plotly to create charts and graphs to better understand the data and to communicate your findings to others.
7. Deployment: You may need to deploy your analysis to a web application or cloud service for real-time analysis and monitoring.

CHAPTER 5

PROJECT DESIGN

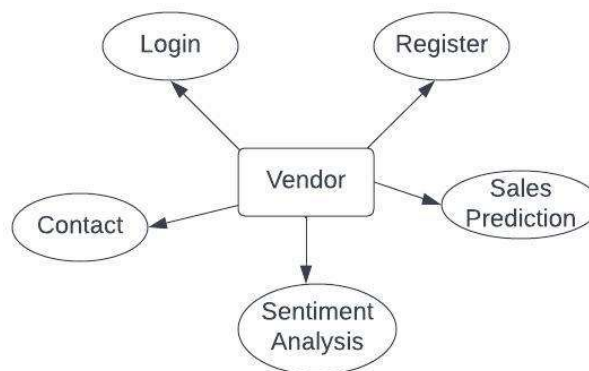
- **USED CASE DIAGRAM:**



USED CASE DIAGRAM

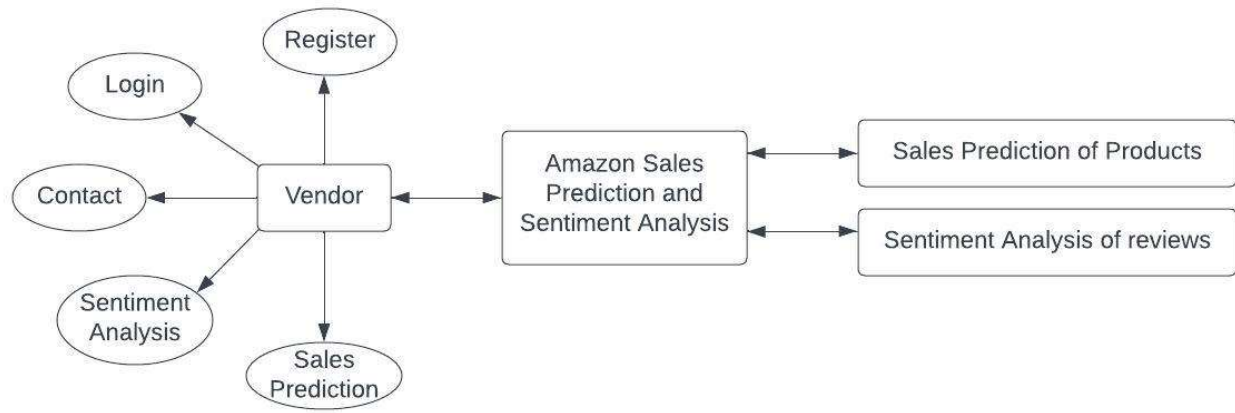
5.1. Used case diagram

- **DATA FLOW DIAGRAM:**



DFD Level 0

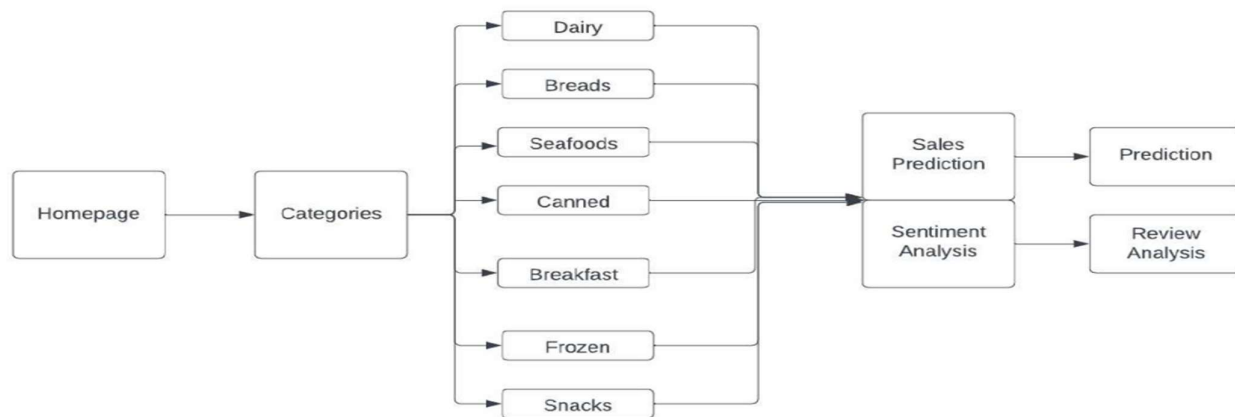
5.2. DFD Level 0



DFD Level 1

5.3. DFD Level 1

• FLOWCHART:



5.4. Flowchart

CHAPTER 6

TECHNICAL SPECIFICATION

Development: VS Code

VS Code also known as Visual Studio Code is a source code editor made by Microsoft for Windows, Linux, MacOS. It has various features such as Debugging, Syntax highlighting, extension, intelligent code completion.

Frontend: Html, CSS, JavaScript

As a web developer, the three main languages we use to build websites are HTML, CSS, and JavaScript. JavaScript is the programming language, we use HTML to structure the site, and we use CSS to design and layout the web page. These days, CSS has become more than just a design language, though. You can implement animations and smooth transitions with just CSS.

OS: Windows

Windows is a graphical operating system developed by Microsoft. It allows users to view and store files, run the software, play games, watch videos, and provides a way to connect to the internet. It was released for both home computing and professional works.

Backend: PHP, MySQL

With PHP, you can connect to and manipulate databases. PHP combined with MySQL are cross-platform (you can develop in Windows and serve on a Unix platform). PHP is a versatile and powerful language that is widely used for web development due to its ease of use, platform independence, database support, object-oriented programming capabilities, large community, and open-source nature.

Machine Learning Based Algorithm:

- 1. Random Forest Algorithm:** Random Forest is an ensemble machine learning algorithm that combines multiple decision trees to make more accurate predictions. The algorithm introduces randomness into the decision tree process by selecting a random subset of input variables for each tree and selecting a random subset of the training data to build each tree. The output of the Random Forest algorithm is determined by a voting mechanism that aggregates the predictions of each individual decision tree. Random Forest is generally robust to overfitting, meaning it is less likely to create models that perform

well on training data but poorly on test data. The algorithm is widely used for classification, regression, and other tasks that involve predicting an outcome based on a set of input variables. Additionally, Random Forest can provide insights into the relative importance of different input variables in making predictions, which can be useful for feature selection and understanding the underlying factors driving the predictions.

2. **Natural Language Processing:** Natural language processing (NLP) is a subfield of artificial intelligence that focuses on developing algorithms and models that can analyze and process human language. NLP involves techniques such as text preprocessing, text classification, named entity recognition, sentiment analysis, machine translation, and speech recognition. These techniques are used to enable computers to understand and interpret natural language text data, such as documents, social media posts, and customer reviews. NLP is widely used in various industries, including healthcare, finance, marketing, and customer service, and is used for applications such as chatbots, customer service automation, social media monitoring, and language translation. NLP is a rapidly evolving field that plays a critical role in enabling computers to understand and process human language, and has many exciting opportunities for innovation and growth.

CHAPTER 7

PROJECT SCHEDULING

Sr.no	Group Members	Time Duration	Work to be done
1	Ashmina Dangat	2nd week of January	Select the topic and work on the frontend GUI
2		First week of February	Design GUI for Sales Page for Sales Predictions site
3		Mid-February	Design GUI for Sentimental pages for Sentimental Analysis
4	Sakshi Gunjal	Last week of February	Apply the Random Forest Algorithm and train the datasets.
5	Neha Chaudhary	First week of March	Connect all separate pages of Sales Prediction Page and Sentimental Page.
6		Second week of March	Connect the algorithm with the project and use the recommendation system.

CHAPTER 9

RESULT AND DISCUSSION



9.1. Homepage



9.2. Sales Prediction

Amazon Fresh

Dairy Products



Next →

Breads



Next →

Sea Food



Next →

9.3. Categories

DAIRY PRODUCTS

Paneer

Amul company



Price :Rs. 86

Item Weight : 200 g

Item fat content : High

Item Type : Dairy Products

Location : Gujarat

Sales Prediction

Sentimental Analysis

9.4. Products

Amazon Sales Prediction

Enter Item Weight :

Enter Fat Content :

Enter Item Visibility :

Enter Item Type

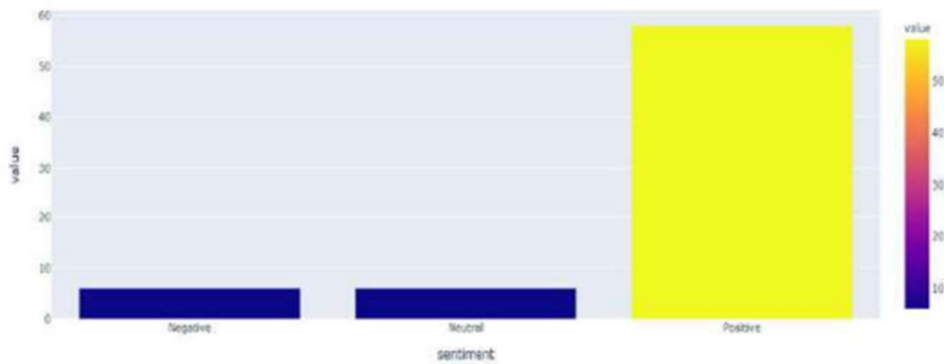
Enter Item MRP :

Outlet Establishment Year (YYYY) :

9.5. Sales Prediction

Amazon Sentimental Analysis

Bar representation based on sentiments of reviews



Bar representation based on star ratings of reviews.

Out

9.6. Sentimental Analysis

CHAPTER 10

CONCLUSION AND FUTURE SCOPE

CONCLUSION:

Amazon sales prediction and sentiment analysis are two different tasks that can be performed using machine learning techniques. It has highlighted the success of the machine learning model by accurately predicting future sales on Amazon and analyzing the sentiment of customer reviews. It has also underscored the potential benefits of these findings for businesses operating on Amazon, such as improved forecasting, better understanding of customer preferences, and opportunities to improve product and service offerings.

FUTURE SCOPE:

1. Incorporating new data sources: Machine learning models can be trained on a variety of data sources beyond customer reviews, such as product descriptions, pricing data, and social media posts. Future research can explore how incorporating these sources of data can improve the accuracy of sales predictions and sentiment analysis.
2. Addressing limitations in sentiment analysis: Sentiment analysis can be challenging due to the nuances of human language and context. Future research can explore ways to improve sentiment analysis accuracy, such as developing models that can better handle sarcasm and irony.
3. Enhancing machine learning models: As machine learning models continue to evolve, future research can focus on developing models that can better handle large volumes of data, identify more complex patterns in data, and improve the interpretability of the results.
4. Expanding to other e-commerce platforms: While Amazon is a dominant player in the e-commerce market, there are many other platforms where businesses can sell their products and services. Future research can explore how machine learning and natural language processing techniques can be applied to other e-commerce platforms to improve sales prediction and sentiment analysis.

REFERENCES

- Sentiment Analysis of E-commerce website : <https://www.irjet.net/archives/V8/i4/IRJET-V8I4979.pdf>
- Sales Prediction Analysis: <https://www.semanticscholar.org/paper/SALES-PREDICTION-ANALYSIS-Kumar-Yogesh/33789f64b8a381947b62f7a21fde0fcea917ac02>
- Prediction and Forecasting of Sales Using Machine Learning Approach : <https://www.irjet.net/archives/V8/i9/IRJET-V8I9335.pdf>
- THE SENTIMENTAL ANALYSIS ON PRODUCT REVIEWS OF AMAZON DATA USING THE HYBRID APPROACH : <https://www.irjet.net/archives/V6/i7/IRJET-V6I7512.pdf>
- Comparative Study of Machine Learning Approaches for Amazon Reviews : <https://www.sciencedirect.com/science/article/pii/S1877050918308512>