Sales Forecasting Using Machine Leaning

BY

MD Injamul Hosen Shuvo
ID: 191-25-741

This Report Presented in Partial Fulfillment of the Requirements for

The Degree of Masters of Science in Computer Science and Engineering

Supervised By

Dr. Sheak Rashed Haider Noori

Professor & Head.

Department of CSE

Daffodil International University

Co Supervised By

Abdus Sattar

Assistant Professor & Coordinator MSc.

Department of CSE

Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JULY 2024

APPROVAL

This Thesis titled **"Sales Forecasting Using Machine Learning"**, submitted by MD Injamul Hosen Shuvo, ID No: 191-25-741 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of M.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 06-07-2024.

BOARD OF EXAMINERS

______________________ **Chairman**

Narayan Ranjan Chakraborty
Associate Professor and Associate Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

______________________ **Internal Examiner**

Md. Sadekur Rahman
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

______________________ **Internal Examiner**

Dr. Ohidujjaman, PhD
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

______________________ **External Examiner**

Mr. Sadat Hossain
Data Scientist
Risk Management Division,
BRAC Bank Limited

DECLARATION

We hereby declare that, this thesis has been done by us under the supervision of **Dr. Sheak Rashed Haider Noori, Professor and Head, Department of CSE** Daffodil International University. We also declare that neither this thesis nor any part of this thesis has been submitted elsewhere for award of any degree or diploma.

Supervised by:

Dr. Sheak Rashed Haider Noori
Professor & Head
Department of CSE
Daffodil International University

Co Supervised by:

Abdus Sattar
Assistant Professor & Coordinator MSc.
Department of CSE
Daffodil International University

Submitted by:

MD Injamul Hosen Shuvo
ID: 191-25-741
Department of CSE
Daffodil International University

©Daffodil International University
ii

ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final thesis successfully.

We really grateful and wish our profound our indebtedness to **Dr. Sheak Rashed Haider Noori**, Professor & Head, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of “Data Mining” to carry out this thesis. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior draft and correcting them at all stage have made it possible to complete this thesis.

We would like to express our heartiest gratitude to **Dr. Sheak Rashed Haider Noori, Professor and Head, Department of CSE**, for his kind help to finish our thesis and also to other faculty members and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, we must acknowledge with due respect the constant support and passion of our parents.

ABSTRACT

Today's dynamic and complicated market environment makes accurate sales forecasting increasingly important for resource allocation and strategic planning. Conventional sales forecasting techniques frequently fail to capture the complex and dynamic nature of market dynamics since they primarily depend on past data and expert opinion. In order to improve accuracy and offer real-time predictive capabilities, this article investigates the incorporation of machine learning (ML) into sales forecasting. Through the utilization of extensive datasets from many sources, such as historical sales data and social media trends, machine learning algorithms are able to provide accurate and thorough projections. The paper goes over the benefits of machine learning (ML) over traditional techniques, such as its quick detection of developing patterns, capacity to analyze massive amounts of data, and flexibility in response to changing situations. We also discuss issues like data quality and system integration, and we look at a variety of machine learning techniques, including ensemble approaches, neural networks, and regression analysis. This article provides a comprehensive overview of how machine learning (ML) may transform sales forecasting and provide organizations a competitive advantage by facilitating proactive management and more informed decision-making.

LIST OF FIGURES

FIGURES	PAGE NO
Figure 3.3.1: Sample Dataset	14
Figure 3.4.1: Sales from 2015 - 2019	14
Figure 3.4.2: Average Daily Sales per Week	15
Figure 3.4.3: Average Daily Sales per Month	15
Figure 3.5.1: Proposed Methodology	16
Figure 4.2.1: Mean of Training Data vs Prediction of ARIMA on Training Data	22
Figure 4.2.2: Mean of Testing Data vs Prediction of ARIMA on Testing Data	23

LIST OF TABLES

TABLE	PAGE NO
Table 4.2.1: Performance of Used Algorithms	22

TABLE OF CONTENTS

CONTENTS **PAGE**

CONTENTS	PAGE
Approval Page	i
Declaration	ii
Acknowledgements	iii
Abstract	iv
List of Figures	vii
List of Tables	vii
CHAPTER	
CHAPTER 1: INTRODUCTION	**1-5**
1.1 Introduction	1-2
1.2 Motivation	2-3
1.3 Research Objective	3-4
1.4 Research Question	4-5
1.5 Report Layout	5
CHAPTER 2: BACKGROUND	**6-11**
2.1 Introduction	6
2.2 Related Work	6-10
2.3 Scope of the problem	10
2.4 Challenges	10-11
CHAPTER 3: RESEARCH METHODOLOGY	**12-20**
3.1 Introduction	12
3.2 Research Subject and Instrumentation	12-13
3.3 Data Collection	13-14

3.4 Statistical Analysis 14-15
3.5 Proposed Methodology 15-19
3.6 Implementation Requirements 19-20

CHAPTER 4: EXPERIMENTAL RESULTS AND DISCUSSION **21-23**
4.1 Introduction 21
4.2 Experimental Results 21-23
4.3 Discussion 23

CHAPTER 5: IMPACT ON SOCIETY, ENVIRONMENT, AND SUSTAINABILITY **24-29**
5.1 Impact on Society 24-25
5.2 Impact on Environment 25-26
5.3 Ethical Aspects 26-28
5.4 Sustainability Plan 28-29

CHAPTER 6: CONCLUSION AND FUTURE WORK **30-25**
6.1 Summary of the study 30
6.2 Conclusions 30-31
6.3 Implication for further study 31-32

REFERENCES **33-34**

CHAPTER 1

INTRODUCTION

1.1 Introduction

Accurate sales forecasting is more important than ever in the fast-paced, constantly-changing corporate environment of today. Strategic planning relies heavily on sales forecasting, which helps businesses decide how best to allocate resources, manage inventories, create budgets, and target markets. These projections have been supported by traditional sales forecasting techniques, which frequently depend on the examination of previous data and professional opinion. But often, they are unable to adequately convey the intricate and ever-changing character of contemporary marketplaces.

Now introduce machine learning, a game-changing technology that has transformed a number of sectors by offering more accurate and useful insights. As a branch of artificial intelligence, machine learning makes use of statistical models and algorithms to find patterns and correlations in massive datasets. Because of this feature, it is the perfect tool for sales forecasting, where it is necessary to understand and anticipate the interactions of several aspects, including seasonality, market trends, consumer behavior, and economic indicators.

Compared to conventional techniques, the use of machine learning into sales forecasting offers several benefits. First off, by continually learning from fresh data and adjusting to shifting circumstances, it provides improved accuracy. Large volumes of data from many sources, such as past sales records, social media trends, site traffic, and even weather patterns, may be processed using machine learning algorithms. This all-encompassing method enables the development of more accurate and thorough projections.

Second, real-time forecasting made possible by machine learning gives companies the most recent information necessary to make quick decisions. Having access to precise and fast projections can provide businesses with a competitive advantage in a world where consumer tastes and market conditions are subject to quick changes. Furthermore, organizations may take proactive measures to address possible difficulties and opportunities as machine learning models can detect abnormalities and developing patterns faster than traditional techniques.

Furthermore, companies of all sizes may utilize machine learning solutions due to their scalability. Businesses of all size, from startups to multinational conglomerates, may customize machine learning models to fit their unique requirements and budgets. The democratization of sophisticated forecasting techniques enables companies to increase customer happiness, streamline processes, and spur development.

In-depth discussions of the benefits, uses, and techniques of machine learning for sales forecasting will be covered in this extensive study. We will look at the many machine learning models that may be used to increase forecasting accuracy, including ensemble approaches, neural networks, and regression analysis. We will also go over the difficulties and factors to be taken into account when putting machine learning solutions into practice, such as data quality, model selection, and integration with current business procedures.

Businesses may improve their forecasting skills and open up new avenues for innovation and operational efficiency by comprehending and utilizing the potential of machine learning. As we go through the complexities of this topic, it becomes clear that machine learning is a catalyst for change in the field of sales forecasting rather than merely a tool for prediction.

1.2 Motivation

Machine learning offers a transformative solution to the limitations of traditional forecasting. By utilizing algorithms that learn from data, machine learning can uncover complex patterns and relationships that are not immediately apparent to human analysts. This capability allows for forecasts that are not only more accurate but also capable of adapting to new trends as more data becomes available.

Motivation:

- Enhanced Accuracy: Machine learning algorithms can analyze large datasets and complex variables to provide more accurate forecasts than traditional methods.

- Dynamic Adaptation: These models improve with exposure to more data, allowing them to adapt to new market trends, consumer behaviors, and external conditions quickly.

- Efficiency Gains: Machine learning automates the forecasting process, significantly reducing the time and effort needed for data analysis and enabling more frequent and timely forecasts.

- Handling Complex Interactions: Machine learning can identify and learn from the intricate interactions between various influencing factors that are often overlooked in traditional forecasting methods.

- Scalability: As businesses grow and data volumes increase, machine learning models scale efficiently to maintain high performance without requiring linear increases in processing resources.

- Integration of Diverse Data Sources: Machine learning can incorporate multiple types of data (e.g., economic indicators, social media sentiment, weather conditions) to provide a comprehensive view of factors affecting sales.

- Improved Decision Making: With more accurate and timely forecasts, businesses can make better-informed decisions regarding inventory management, resource allocation, and strategic planning.

- Competitive Advantage: Leveraging machine learning for sales forecasting can provide a significant edge over competitors by enabling more agile responses to market changes and customer needs.

- Cost Reduction: By optimizing inventory and reducing the risk of overstocking or stockouts, machine learning helps minimize costs associated with wasted resources and lost sales opportunities.

1.3 Research Objective

The main objective of our project to develop a reliable machine learning based model to forecasting product sales.

The objectives of this research are:

- Develop Accurate Sales Forecasting Models: By taking into account past sales data, market trends, and outside variables, machine learning techniques may be used to develop models that accurately anticipate future sales.

- Identify Key Sales Drivers: Examine and measure the effects of the many variables that affect sales, including economic indicators, seasonal trends, promotions, and consumer behavior.
- Enhance Demand Planning: By offering dependable sales projections, assisting companies in optimizing inventory levels, lowering stockouts, and minimizing overstock scenarios, you may boost the effectiveness of demand planning procedures.
- Support Strategic Decision-Making: By offering useful insights from sales predictions, you may facilitate data-driven decision-making for marketing, pricing, and product development initiatives.

Automate Forecasting Processes: Create automated systems that will enhance and update sales predictions continuously, decreasing the need for human interaction and boosting update frequency.

1.4 Research Questions

Based on the objectives you've provided for sales forecasting using machine learning, here are some potential research questions that could guide your investigation:

RQ1. How accurately can machine learning models predict future sales using historical sales data combined with external market trends?

- This question aims to evaluate the predictive power of machine learning models and identify the most influential data points.

RQ2. What are the key drivers of sales in the target market, and how do different variables influence sales outcomes?

- This involves identifying and quantifying the impact of various factors like economic indicators, seasonal trends, promotions, and consumer behaviors on sales.

RQ3. How can machine learning improve demand planning and inventory management for businesses?

- This question focuses on the practical application of machine learning forecasts in reducing stockouts and overstock situations by optimizing inventory levels.

RQ4. In what ways can machine learning-driven sales forecasts support strategic decision-making in areas such as marketing, pricing, and product development?

- The goal here is to explore how predictive insights from machine learning models can inform broader business strategies and operational decisions.

RQ5. What challenges and limitations are commonly encountered when implementing machine learning models for sales forecasting, and how can these be mitigated?

- This question addresses potential hurdles in the practical deployment of machine learning models and seeks solutions to enhance their effectiveness and integration into business processes.

1.5 Report Layout

The report has total 6 Chapters which will be followed given by instructions:

In Chapter 1, we have discussed about the introduction of this research in this part of this research, we discussed in details about the importance of sales forecasting. Research question, motivation of this research, rationale and expected output has also discussed in this part.

In Chapter 2, we reviewed existing work on leaf disease prediction in this chapter. We discussed about other authors approaches, limitations, results, methods in this part. Research scope and challenges also have discussed here.

In Chapter 3, research methodology has mainly discussed here. This chapter shows the data collection procedure, statistics of data, classifiers, figures of stats. Implementation requirements also have discussed in this chapter.

In Chapter 4, results of this research have mainly showed. In this chapter we have showed the result of all the classifiers we have used in this research.

In Chapter 5, shows how this research can impact in our society. Why this work is important and how it will sustain in this arena, this chapter shows mainly.

In Chapter 6, discussion and conclusion of this research have shown. In this part we have also discussed about the future work, and the limitations of this work.

CHAPTER 2

BACKGROUND

2.1 Introduction

The topic of demand and sales forecasting has been greatly affected by recent developments in machine learning (ML). This review of the literature compiles several studies that have used various machine learning approaches to improve forecasting efficiency and accuracy in a variety of industries, such as general goods, retail, and the food business.
Forecasting sales is crucial for supply chain and production management. Planning, strategy, marketing, logistics, warehousing, and resource management are all impacted by it for businesses. In addition to previous behavior and trends, causal forecasting techniques may also anticipate future sales behavior based on correlations between factors. outlines a framework for applying genetic programming, an artificial intelligence method based on the theory of biological evolution, to estimate and simulate export sales. An export sales forecasting model is proposed after an empirical case study of an export firm is examined. Additionally, a six-week sales prediction is created and its results are compared to actual sales data. Lastly, a causal forecasting model variable sensitivity analysis is shown.

2.2 Related Works

The performance of nine cutting-edge machine learning and three traditional forecasting algorithms was objectively examined for the first time in horticulture sales projections by the author of [2]. In every trial, they demonstrated the superiority of machine learning techniques, with the gradient boosted ensemble learner XGBoost emerging as the best performer in 14 of the 15 comparisons. When dealing with datasets that have many seasons, this benefit above traditional forecasting techniques grew. They also demonstrated how adding meta-features and other external variables, such holiday and weather information, improved prediction performance. Furthermore, we looked at whether the algorithms could have predicted the sharp spike in horticulture product demand that occurred in 2020 during the SARS-CoV-2 epidemic. Moreover, XGBoost worked better in this particular situation.

In recent years, there has been an increase in interest in the subject of neural networks for prediction. In order to predict furniture sales, a public dataset that contains a retail store's sales history is examined in [3]. Several forecasting models are used to achieve this goal. Initially, a few traditional time-series forecasting methods are applied, including Triple Exponential Smoothing and Seasonal Autoregressive Integrated Moving Average (SARIMA). Next, more sophisticated techniques like Convolutional Neural Networks (CNN), Prophet, and Long Short-Term Memory (LSTM) are used. Various accuracy measuring techniques, such as Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE), are used to compare the model performances. The outcomes demonstrate the Stacked LSTM method's superiority over the alternative approaches.

In order to predict furniture sales, a public dataset that contains a retail store's sales history is examined in [4]. Several forecasting models are used to achieve this goal. Initially, a few traditional time-series forecasting methods are applied, including Triple Exponential Smoothing and Seasonal Autoregressive Integrated Moving Average (SARIMA). Next, more sophisticated techniques like Convolutional Neural Networks (CNN), Prophet, and Long Short-Term Memory (LSTM) are used. Various accuracy measuring techniques, such as Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE), are used to compare the model performances. The outcomes demonstrate the Stacked LSTM method's superiority over the alternative approaches. Furthermore, the outcomes demonstrate the Prophet and CNN models' strong performances.

Business intelligence is the process that decision makers utilize to choose the best course of action by processing, analyzing, and interpreting the data to acquire insights. Building a model that will provide a forecast based on adjusted sales figures utilizing certain customer-centric qualities was the primary goal of [5]. They divide the consumer base into discrete groups according to their purchasing habits using the RFM model first, then eliminate the groups that are unimportant to the company. After that, they anticipate sales for the remaining data using the ARIMA model. When applied following RFM analysis, they were able to obtain a greater accuracy and better fit of the data for the prediction model.

In [6], the goal was to identify the optimal algorithm for their specific problem statement while attempting to forecast a retail store's sales using various machine learning approaches. They have used both boosting and normal regression approaches in our approach, and they have discovered that the boosting algorithms produce superior outcomes than the regular regression algorithms.

In [7], the sales pattern of Power Bumi products during the COVID-19 pandemic was analyzed, and the forecasting approach that produces the minimum error value in Power Bumi product sales forecasting was compared to PT. Zamrud Bumi Indonesia. Two approaches are used in this study: the least square trend model and exponential smoothing. to use MAD, MSE, and MAPE to determine the error rate. In comparison to other forecasting techniques, the exponential smoothing alpha 0.9 approach has the lowest error value, according to the data. With a forecast of 627.628 boxes, the MAD value is 130.329, the MSE is 28251.23, and the MAPE is 22.00% when it comes to product sales forecasting.

In today's corporate world, managing the retail supply chain is essential to success. Supply chain management greatly depends on the ability to forecast consumer demand. The ability to foresee every detail accurately affects sales volume, customer attractiveness, profit margin, storage, and lost profits. [8] will develop a novel technique that makes use of machine learning to aid in precise prediction. This technique gathers and analyzes a store's historical data. obtaining the pertinent data, processing it, and becoming ready to use it in the appropriate way. utilizing relevant algorithms on the processed data. We are aware that several algorithms have been utilized recently for prediction in K-Nearest Neighbor, Support Vector Machine, Gaussian Nave Bayes, Random Forest, Decision Tree Classifier, and regressions. We get actual data from the marketplace. This document was created using the store position, the month, the event for that month, and additional relevant data. The geographic location of their nation affects forecasting, as our research explains. Their model generates a preliminary demand for a certain good. This estimation benefits retails and their enterprises.

For computer-retailing sales forecasting, a clustering-based forecasting model incorporating clustering and machine-learning techniques is presented in [9]. The suggested approach divided the training data into groups by first using the clustering technique, which groups data with comparable traits or patterns. The forecasting models of each group are then trained using machine-learning techniques. The trained forecasting model of the cluster was used for sales forecasting once the cluster whose data patterns were most comparable to the test data was identified. Given that computer merchants' sales data exhibits comparable data patterns or features throughout a range of time periods, applying the suggested clustering-based forecasting model can improve prediction accuracy. In this work, two machine learning approaches, support vector regression (SVR) and extreme learning machine (ELM), and three clustering techniques, self-organizing map (SOM), growing hierarchical self-organizing map

(GHSOM), and K-means, are employed. There were six forecasting models based on clustering that were put out. The empirical examples include actual sales statistics for laptops, liquid crystal displays, and personal computers.

Businesses may more effectively manage cash flow, production, and create more informed company plans by using sales forecasting. They provide a machine learning-based, accurate, and efficient sales forecasting model in [10]. To extract characteristics from historical sales data, feature engineering is first carried out. They also employed eXtreme Gradient Boosting (XGBoost) to make use of these traits in order to predict the quantity of sales in the future. Our suggested model performs very well for sales prediction with fewer computational time and memory resources, as demonstrated by the experiment results on a publicly available Walmart retail items dataset provided by the Kaggle competition.

[12] assesses and contrasts a number of machines learning models, including ARIMA, XGBoost, SVM, Auto Regressive Neural Network (ARNN), and Hybrid Rossmann, a drugstore corporation, uses models such as Hybrid ARIMA-ARNN, Hybrid ARIMA-XGBoost, Hybrid ARIMA-SVM, and STL Decomposition (using ARIMA, Naive, XGBoost) to anticipate sales. The training data set includes additional information on medicine retailers as well as historical sales data. Metrics like RMSE and MAE are used to gauge how accurate these models are. Originally, sales forecasts were made using linear models like ARIMA. Since ARIMA was unable to accurately capture nonlinear patterns, nonlinear models like SVM, XGBoost, and neural networks were employed. Nonlinear models yielded lower RMSE and outperformed ARIMA. Composite models were created utilizing a combination of decomposition and hybrid techniques in order to further enhance performance. The three hybrid models—Hybrid ARIMA-ARNN, Hybrid ARIMA-XGBoost, and Hybrid ARIMA-SVM—all outperformed the corresponding single models. Subsequently, Snaive, ARIMA, and XGBoost were used to forecast the seasonal, trend, and residual components of the decomposed model, which was then built using STL Decomposition. Compared to hybrid and individual models, STL produced superior outcomes.

In a business setting, the supermarket sales prediction contributes to increased sales. The method aids in problem-domain decision-making. There are several forecasting tools available, including the logistic exponential model and regression model. The most recent technology to demonstrate better performance in terms of forecast accuracy is Facebook's (FB) Prophet. A Facebook Prophet tool has been presented in [14] study to anticipate supermarket sales using data. A few forecasting models, including the additive model, the Autoregressive integrated moving average (ARIMA) model, and the Facebook Prophet model, have been studied in the proposed research project. Based on the suggested study, it can be said that FB Prophet is a more accurate prediction model in terms of fitting, low error, and better prediction.

To anticipate sales volume, a composite GRU–Prophet model with an attention mechanism was built in [15]. Linear and nonlinear characteristics were captured in this composite model using the Prophet model and the GRU model with attention mechanism, respectively. It was discovered through experiments that the composite model outperformed the recurrent neural network, long short-term memory, gate recurrent unit, Prophet, and autoregressive integrated moving average models in terms of applicability and prediction accuracy. The composite model used in this study helps businesses become more competitive in the field of smart manufacturing and is adaptable to quick changes in consumer demand.

2.3 Scope of the problem

Inclusions:

- Development and testing of the machine learning model using a designated training and validation dataset.
- Exploration and selection of appropriate machine learning algorithms based on data characteristics and business needs.
- Feature engineering to extract relevant features from raw data for model training.
- Model performance evaluation through well-defined metrics and comparison with traditional methods.
- Basic deployment strategy outlining model integration and user access.

Exclusions:

- Acquisition of new data sources beyond existing company data.
- Development of a full-fledged production-grade application for deploying the model.
- Ongoing maintenance and monitoring of the model after deployment (may be considered in a future phase).

2.4 Challenges

Sales forecasting using machine learning comes with a set of challenges that businesses need to address to effectively implement and leverage these technologies. Here are some of the common challenges:

1. Data Quality and Availability: Machine learning models are heavily dependent on the quality and quantity of the data they are trained on. Inaccurate, incomplete, or biased data can lead to poor forecasts. Additionally, small or newly established businesses might not have sufficient historical data for effective training of models.

2. Integration with Existing Systems: Integrating machine learning models into existing IT and business systems can be complex and costly. Compatibility issues between new machine learning solutions and legacy systems can hinder smooth deployment and functionality.

3. Changing Market Dynamics: The assumption that historical patterns will predict future trends doesn't always hold true, particularly in volatile markets or industries experiencing rapid change. Machine learning models need continuous updates and retraining to stay relevant, which can be resource-intensive.

4. Overfitting and Generalization: There's a risk of models being overfitted to the training data, making them perform well on seen data but poorly on unseen data. Ensuring that models generalize well to new data is a critical challenge.

5. Cost and Resource Allocation: Developing, training, and deploying machine learning models require significant investments in technology and skilled personnel. The cost and effort to maintain and update these systems can be substantial, which might be a barrier, especially for smaller businesses.

Addressing these challenges requires a careful balance of technical expertise, strategic planning, and ongoing management of machine learning systems. Ensuring that the forecasts generated are actionable and align with business goals is essential for leveraging machine learning effectively in sales forecasting.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

The methodology for “Sales Forecasting using Machine Learning” begins with data collection, where historical sales data and relevant variables such as dates, product categories, prices, promotions, and economic indicators are gathered. This data serves as the foundation for the forecasting model. Once collected, the data undergoes preprocessing to handle missing values, outliers, and inconsistencies. This step ensures the data is clean and reliable, facilitating accurate model training. Exploratory data analysis (EDA) is also conducted to understand data distributions and identify key features influencing sales patterns.

Following preprocessing, the next phase involves splitting the data into training and test datasets. The training dataset is used to develop and train multiple regression models, such as linear regression, decision trees, and neural networks. During this stage, hyperparameter tuning is performed to optimize model performance. Once the models are trained, they are evaluated using metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). Cross-validation is employed to ensure the robustness and generalizability of the models, enabling the selection of the highest-performing model based on evaluation metrics.

The final phase is model selection and deployment. The best-performing model is integrated into the existing sales management system, providing a user-friendly interface for sales teams to input data and receive forecasts. This system is designed to be scalable, handling real-time data input and forecasting. Continuous improvement is facilitated through a feedback mechanism that collects data on forecast accuracy and user satisfaction, ensuring regular model updates to adapt to changing market conditions. Additionally, comprehensive reporting tools and visualizations are developed to present sales forecasts in an easily interpretable manner, aiding stakeholders in understanding sales trends and making informed decisions.

3.2 Research Subject and Instrumentation

As sales forecasting becomes increasingly critical for businesses, finding a more reliable method to predict sales trends is essential. Given the vast amount of sales data and influencing external factors, traditional forecasting methods often fall short in accuracy. It's challenging to predict sales outcomes effectively just by analyzing historical data, as market dynamics and consumer behaviors are continually evolving. Without precise forecasts, businesses cannot strategize effectively, potentially leading to overstocking, missed opportunities, and financial losses.

Considering these challenges, we have chosen to focus our study on enhancing sales forecasting using machine learning techniques. Convolutional Neural Networks (CNNs) and other advanced machine learning models offer promising solutions. These models are capable of filtering out noise and irrelevant data from vast datasets and translating them into meaningful predictions. They create a framework within which multiple neural networks can be tailored to address specific forecasting challenges.

To ensure the effectiveness of these models, it is crucial to process and refine them before comparing their performance against traditional forecasting methods. This allows for more systematic improvements and a successful implementation, ultimately aiming to enhance the precision and reliability of sales forecasts. Our experiments include various iterations and refinements of these models to align with our goal of improving sales forecasting accuracy, thus supporting businesses in making better-informed decisions.

3.3 Data Collection

Data collection is the most important part for any research. We used a fully raw dataset for this research. We collected it from a shop from USA where the sales were listed from 2015 - 2019. It is a time series dataset which has 9800 rows and 18 columns. It is a regression dataset.

Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Na	Segment	Country	City	State	Postal Code	Region	Product ID	Category	Sub-Category	Product Nam	Sales
1	CA-2017-152	8/11/17	11/11/17	Second Class	CG-12520	Claire Gute	Consumer	United State	Henderson	Kentucky	42420	South	FUR-BO-100	Furniture	Bookcases	Bush Somers	261.96
2	CA-2017-152	8/11/17	11/11/17	Second Class	CG-12520	Claire Gute	Consumer	United State	Henderson	Kentucky	42420	South	FUR-CH-100	Furniture	Chairs	Hon Deluxe F	731.94
3	CA-2017-138	12/6/17	16/6/17	Second Class	DV-13045	Darrin Van H	Corporate	United State	Los Angeles	California	90036	West	OFF-LA-100	Office Suppli	Labels	Self-Adhesiv	14.62
4	US-2016-108	11/10/16	18/10/16	Standard Cla	SO-20335	Sean O'Donn	Consumer	United State	Fort Lauderd	Florida	33311	South	FUR-TA-100	Furniture	Tables	Bretford CR4	957.5775
5	US-2016-108	11/10/16	18/10/16	Standard Cla	SO-20335	Sean O'Donn	Consumer	United State	Fort Lauderd	Florida	33311	South	OFF-ST-100	Office Suppli	Storage	Eldon Fold 'N	22.368
6	CA-2015-115	9/6/15	14/6/15	Standard Cla	BH-11710	Brosina Hoff	Consumer	United State	Los Angeles	California	90032	West	FUR-FU-100	Furniture	Furnishings	Eldon Expres	48.86
7	CA-2015-115	9/6/15	14/6/15	Standard Cla	BH-11710	Brosina Hoff	Consumer	United State	Los Angeles	California	90032	West	OFF-AR-100	Office Suppli	Art	Newell 322	7.28
8	CA-2015-115	9/6/15	14/6/15	Standard Cla	BH-11710	Brosina Hoff	Consumer	United State	Los Angeles	California	90032	West	TEC-PH-1000	Technology	Phones	Mitel 5320 IF	907.152
9	CA-2015-115	9/6/15	14/6/15	Standard Cla	BH-11710	Brosina Hoff	Consumer	United State	Los Angeles	California	90032	West	OFF-BI-1000	Office Suppli	Binders	DXL Angle-Vi	18.504
10	CA-2015-115	9/6/15	14/6/15	Standard Cla	BH-11710	Brosina Hoff	Consumer	United State	Los Angeles	California	90032	West	OFF-AP-100	Office Suppli	Appliances	Belkin F5C20	114.9
11	CA-2015-115	9/6/15	14/6/15	Standard Cla	BH-11710	Brosina Hoff	Consumer	United State	Los Angeles	California	90032	West	FUR-TA-100	Furniture	Tables	Chromcraft F	1706.184
12	CA-2015-115	9/6/15	14/6/15	Standard Cla	BH-11710	Brosina Hoff	Consumer	United State	Los Angeles	California	90032	West	TEC-PH-1000	Technology	Phones	Konftel 250 (	911.424
13	CA-2018-114	15/4/18	20/4/18	Standard Cla	AA-10480	Andrew Alle	Consumer	United State	Concord	North Carolir	28027	South	OFF-PA-100	Office Suppli	Paper	Xerox 1967	15.552
14	CA-2017-161	5/12/17	10/12/17	Standard Cla	IM-15070	Irene Maddo	Consumer	United State	Seattle	Washington	98103	West	OFF-BI-1000	Office Suppli	Binders	Fellowes PB	407.976
15	US-2016-118	22/11/16	26/11/16	Standard Cla	HP-14815	Harold Pawla	Home Office	United State	Fort Worth	Texas	76106	Central	OFF-AP-100	Office Suppli	Appliances	Holmes Repl	68.81
16	US-2016-118	22/11/16	26/11/16	Standard Cla	HP-14815	Harold Pawla	Home Office	United State	Fort Worth	Texas	76106	Central	OFF-BI-1000	Office Suppli	Binders	Storex DuraT	2.544
17	CA-2015-105	11/11/15	18/11/15	Standard Cla	PK-19075	Pete Kriz	Consumer	United State	Madison	Wisconsin	53711	Central	OFF-ST-100	Office Suppli	Storage	Stur-D-Stor S	665.88
18	CA-2015-167	13/5/15	15/5/15	Second Class	AG-10270	Alejandro Gr	Consumer	United State	West Jordan	Utah	84084	West	OFF-ST-100	Office Suppli	Storage	Fellowes Su	55.5
19	CA-2015-143	27/8/15	1/9/15	Second Class	ZD-21925	Zuschuss Do	Consumer	United State	San Francisc	California	94109	West	OFF-AR-100	Office Suppli	Art	Newell 341	8.56
20	CA-2015-143	27/8/15	1/9/15	Second Class	ZD-21925	Zuschuss Do	Consumer	United State	San Francisc	California	94109	West	TEC-PH-1000	Technology	Phones	Cisco SPA 50	213.48
21	CA-2015-143	27/8/15	1/9/15	Second Class	ZD-21925	Zuschuss Do	Consumer	United State	San Francisc	California	94109	West	OFF-BI-1000	Office Suppli	Binders	Wilson Jones	22.72
22	CA-2017-137	9/12/17	13/12/17	Standard Cla	KB-16585	Ken Black	Corporate	United State	Fremont	Nebraska	68025	Central	OFF-AR-100	Office Suppli	Art	Newell 318	19.46
23	CA-2017-137	9/12/17	13/12/17	Standard Cla	KB-16585	Ken Black	Corporate	United State	Fremont	Nebraska	68025	Central	OFF-AP-100	Office Suppli	Appliances	Acco Six-Out	60.34
24	US-2018-156	16/7/18	18/7/18	Second Class	SF-20065	Sandra Flana	Consumer	United State	Philadelphia	Pennsylvania	19140	East	FUR-CH-100	Furniture	Chairs	Global Delux	71.372
25	CA-2016-106	25/9/16	30/9/16	Standard Cla	EB-13870	Emily Burns	Consumer	United State	Orem	Utah	84057	West	FUR-TA-100	Furniture	Tables	Bretford CR4	1044.63
26	CA-2017-121	16/1/17	20/1/17	Second Class	EH-13945	Eric Hoffmar	Consumer	United State	Los Angeles	California	90049	West	OFF-BI-1000	Office Suppli	Binders	Wilson Jones	11.648
27	CA-2017-121	16/1/17	20/1/17	Second Class	EH-13945	Eric Hoffmar	Consumer	United State	Los Angeles	California	90049	West	TEC-AC-1000	Technology	Accessories	Imation¬†8G	90.57
28	US-2016-150	17/9/16	21/9/16	Standard Cla	TB-21520	Tracy Blums	Consumer	United State	Philadelphia	Pennsylvania	19140	East	FUR-BO-100	Furniture	Bookcases	Riverside Pal	3083.43
29	US-2016-150	17/9/16	21/9/16	Standard Cla	TB-21520	Tracy Blums	Consumer	United State	Philadelphia	Pennsylvania	19140	East	OFF-BI-1000	Office Suppli	Binders	Avery Recycl	9.618
30	US-2016-150	17/9/16	21/9/16	Standard Cla	TB-21520	Tracy Blums	Consumer	United State	Philadelphia	Pennsylvania	19140	East	FUR-FU-100	Furniture	Furnishings	Howard Mill	124.2

Figure 3.3.1: Sample Dataset

3.4 Statistical Analysis

In this part, we can see the sales on the basis of per month and per week from 2015-2019 in Figure 3 and Figure 4.

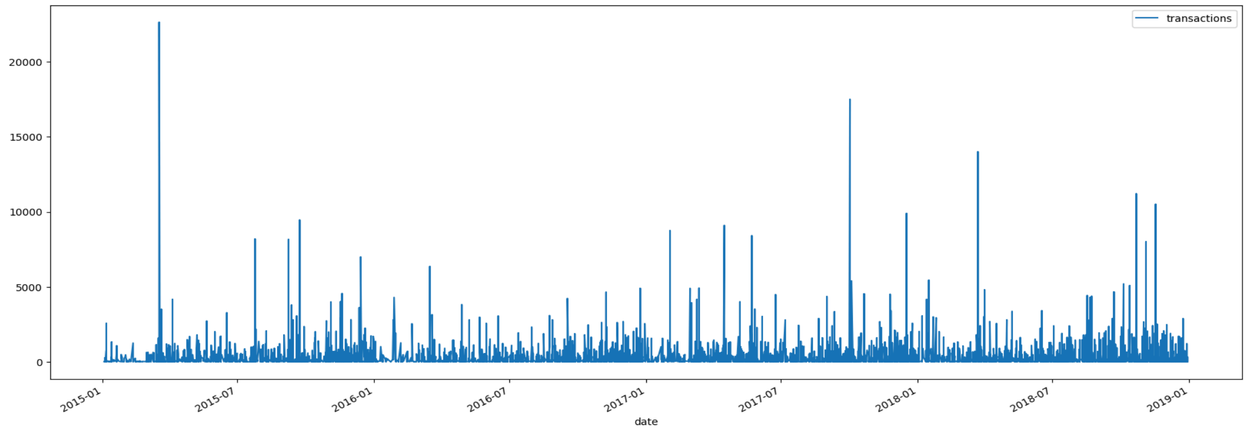

Figure 3.4.1: Sales from 2015 - 2019

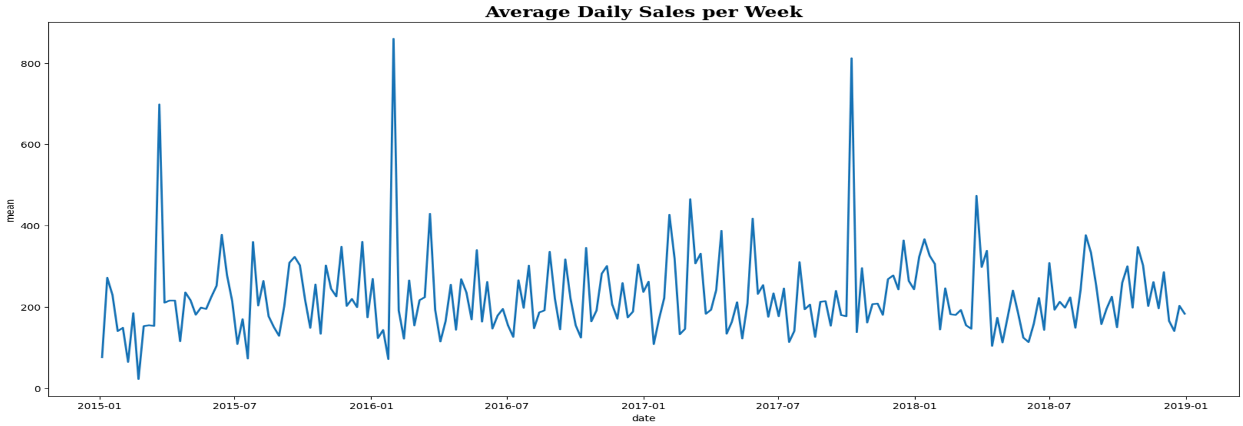

Figure 3.4.2: Average Daily Sales per Week

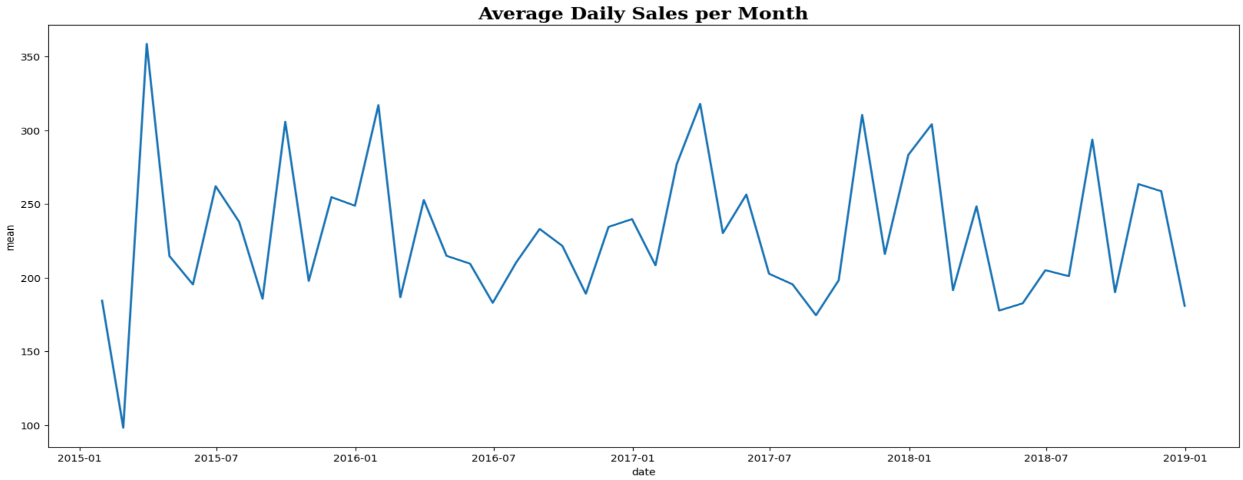

Figure 3.4.3: Average Daily Sales per Month

3.5 Proposed Methodology

In this section we are going to elaborate the proposed methodology for our work which includes data collection, preprocessing, model training and evaluation. Diagram of the overall methodology is shown in Figure 1 and every step of the methodology is further discussed in the following subsections.

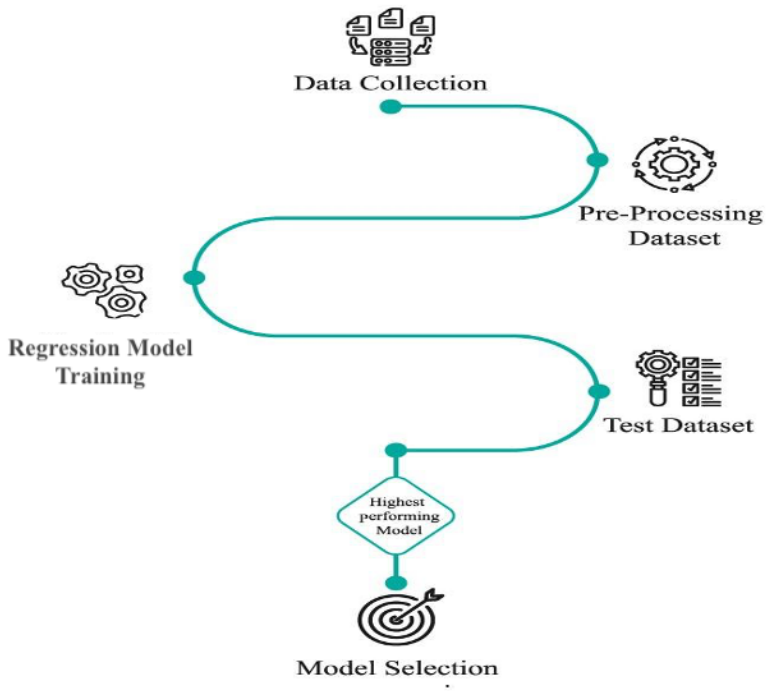

Figure 3.5.1: Proposed Methodology

3.5.1 Data Collection

We discuss about data collection in 3.3.

3.5.2 Data Preprocessing

At first, we checked the null value. There were some null values. We dropped them. Most of the column was string type. So, we had to encoded them into integer value by using label encoder. After null value handling and encoding the dataset was ready to fit in the model.

3.5.3 Model Training

In machine learning and statistics, regression algorithms are a kind of supervised learning that are used to model and examine the correlations between variables. Using one or more input factors (independent variables), the objective is to predict a continuous output variable (dependent variable).

Linear Regression: One of the most straightforward and popular statistical methods for simulating the connection between a dependent variable and one or more independent variables is linear regression. The main goal is to establish a linear connection between the target (output variable) and the input variables (features) [24].

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \epsilon$$

where:

- y is the dependent variable (target).
- $\beta 0$ is the intercept.
- $\beta 1$, $\beta 2$,…,βn are the coefficients of the independent variables.
- $x1$, $x2$…, xn are the independent variables (features).
- ϵ is the error term.

Lasso: A regularization term (L1 penalty) is included in Lasso (Least Absolute Shrinkage and Selection Operator) regression, a kind of linear regression, to enforce sparsity in the coefficients, meaning that certain coefficients may be precisely zero. This aids in choosing features [24].

$$\text{Loss} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \alpha \sum_{j=1}^{p} |\beta_j|$$

where:

- $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ is the residual sum of squares.
- $\alpha \sum_{j=1}^{p} |\beta_j|$ is the L1 regularization term.
- α is the regularization parameter.

Elastic Net: The L1 and L2 penalties of the Lasso and Ridge techniques are combined linearly in Elastic Net, a regularized regression technique. When there are several aspects that are connected with one another, it is helpful [24].

$$\text{Loss} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \alpha_1 \sum_{j=1}^{p} |\beta_j| + \alpha_2 \sum_{j=1}^{p} \beta_j^2$$

where:

- $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ is the residual sum of squares.
- $\alpha_1 \sum_{j=1}^{p} |\beta_j|$ is the L1 regularization term.
- $\alpha_2 \sum_{j=1}^{p} \beta_j^2$ is the L2 regularization term.
- α_1 and α_2 are regularization parameters.

Support Vector Regression (SVR): While Support Vector Machines (SVM) are used for classification, Support Vector Regression (SVR) use the same ideas for regression. In order to optimize for a tolerance margin (epsilon), it attempts to fit the optimal line inside a threshold value [24].

$$\text{Loss} = \frac{1}{2} \sum_{j=1}^{p} \beta_j^2 + C \sum_{i=1}^{n} \max(0, |y_i - \hat{y}_i| - \epsilon)$$

where:

- $\frac{1}{2} \sum_{j=1}^{p} \beta_j^2$ is the regularization term.
- C is a regularization parameter.
- $\max(0, |y_i - \hat{y}_i| - \epsilon)$ is the epsilon-insensitive loss function.
- ϵ defines the margin of tolerance.

Decision Tree Regressor: Decision Tree Regression creates a tree with each leaf representing a projected value by dividing the data into subsets according to the input feature values. This process models the connection between features and the goal [24].

Random Forest Regressor: An ensemble technique called Random Forest Regression creates many decision trees and averages each one's forecast. When compared to a single decision tree, it decreases overfitting and increases accuracy [24].

$$\hat{y} = \frac{1}{T} \sum_{t=1}^{T} \hat{y}^{(t)}$$

where:

- T is the number of trees.
- $\hat{y}^{(t)}$ is the prediction from the t-th tree.

ARIMA (Auto Regressive Integrated Moving Average): Autoregression (AR), differencing (I), and moving average (MA) are the three components of the widely used ARIMA time series forecasting technique. It represents the residual errors from a moving average model applied to lagged data (MA) and the connection between an observation and many lagged observations (AR) [24].

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q} + \epsilon_t$$

where:

- y_t is the value at time t.
- c is a constant term.
- $\phi_1, \phi_2, \ldots, \phi_p$ are the autoregressive coefficients.
- $\theta_1, \theta_2, \ldots, \theta_q$ are the moving average coefficients.
- ϵ_t is the error term at time t.
- p is the order of the autoregressive part.
- q is the order of the moving average part.

3.5.4 Model Evaluation

Mean Absolute Error (MAE): Without taking into account their direction, the MAE calculates the average size of the mistakes in a set of forecasts. It is the mean, with each individual difference carrying equal weight, of the absolute differences between the observed and predicted values over the test sample.

$$\mathrm{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|$$

Mean Squared Error (MSE): The average squared difference between the estimated values and what is estimated is measured by the Mean Squared Error (MSE). Larger mistakes are assigned a higher weight.

$$\mathrm{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

3.6 Implementation Requirements

There are some requirements to implemented this project.

Hardware or Software Requirements

- Windows 10 operating system

- Hard Disk 512 GB
- 4 GB RAM
- Google Chrome or Microsoft Edge

Developing Tools

- Python 3.9
- Tensor flow
- Jupiter or Google Colab

CHAPTER 4

EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Introduction

In this section, we present the findings from our application of machine learning (ML) techniques to improve sales forecasting. The focus of our research has been on assessing the effectiveness of various ML models, including Convolutional Neural Networks (CNNs) and other advanced algorithms, in accurately predicting future sales based on historical data and external market factors.

Our results are derived from a comprehensive series of experiments designed to compare the predictive power of these machine learning models against traditional sales forecasting methods. The data set used encompasses a wide range of variables, from past sales data and seasonal trends to economic indicators and consumer behavior metrics, allowing for a nuanced analysis of model performance.

The outcomes highlighted in this section are structured to first detail the experimental setup, including the data preparation, model training, and validation processes. Following this, we discuss the performance metrics used to evaluate the models, such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and others, providing a quantitative basis for comparison.

Key findings and insights from the experiments will be outlined, demonstrating the capacity of machine learning to transform sales forecasting practices by offering more accurate, efficient, and adaptable predictions. This section aims to substantiate the superiority of machine learning models over traditional forecasting techniques and discuss the implications of these results for future business strategies and decision-making processes.

4.2 Experimental Results

There are a number of different machine learning algorithms that can be used for regression dataset. We used 7 algorithms: Linear Regression, Lasso Regression, Elastic Net Regression, Support Vector Regression, Decision Tree Regression, Random Forest Regression and ARIMA. At first, we used first 6 algorithms but the result was so bad. Then we used forecasting algorithm

ARIMA which gives us much better result. The performance of all algorithms is shown below in Table 1.

Table 4.2.1: Performance of Used Algorithms

SL	Model	Mean Squared Error	Mean Absolute Error
1	Linear Regression	298122.463	268.026
2	Lasso Regression	298113.648	267.989
3	Elastic Net Regression	298729.908	267.545
4	Support Vector Regression	330885.301	204.165
5	Decision Tree Regression	372022.723	206.285
6	Random Forest Regression	265424.179	183.198
7	ARIMA	44.796	41.840

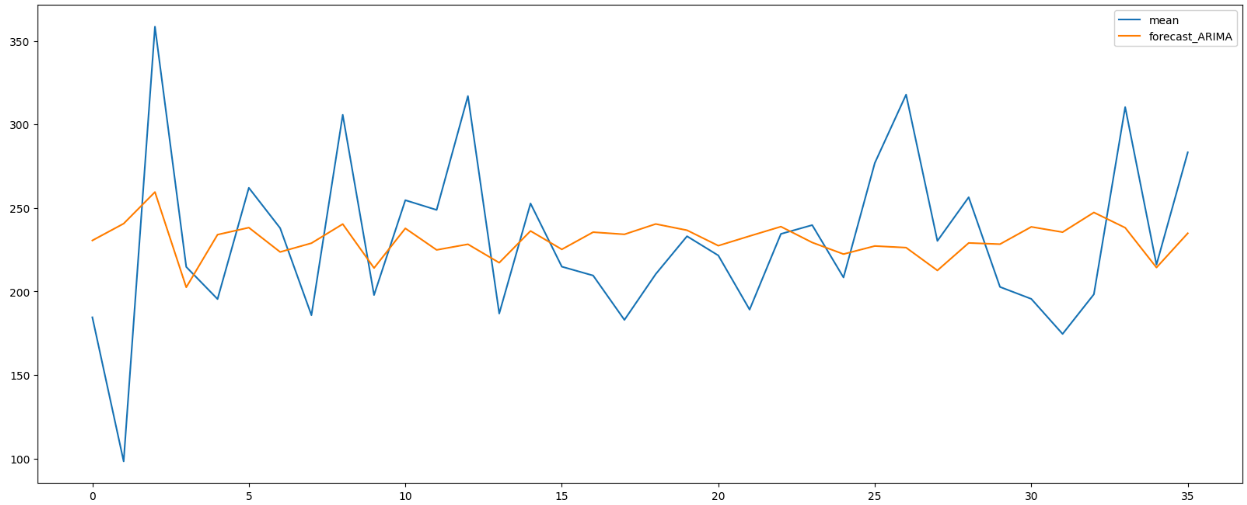

Figure 4.2.1: Mean of Training Data vs Prediction of ARIMA on Training Data

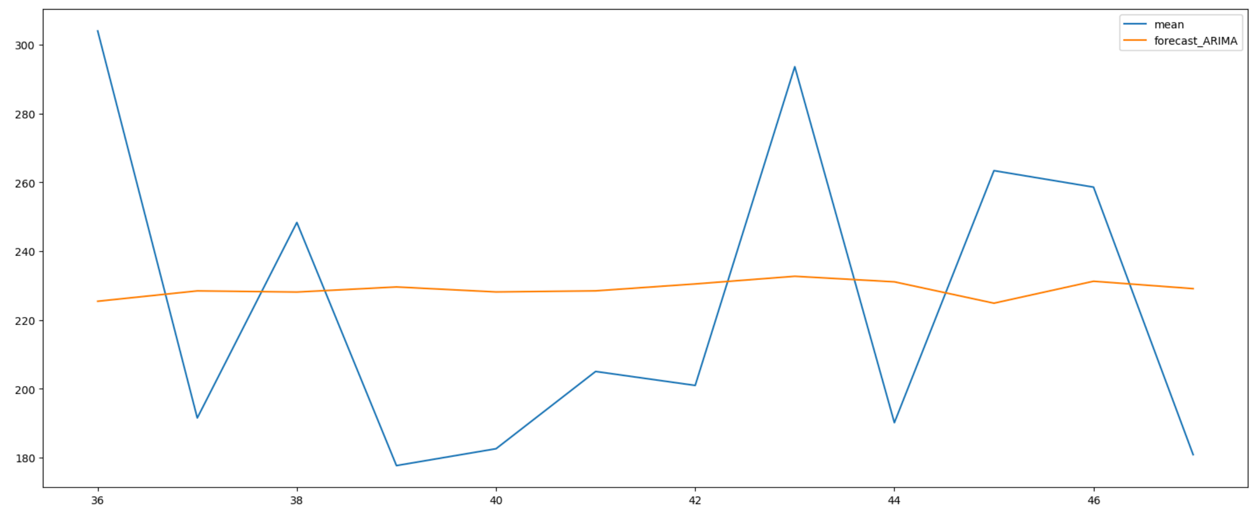

Figure 4.2.1: Mean of Testing Data vs Prediction of ARIMA on Testing Data

4.3 Discussion

The Mean Squared Error (MSE) and Mean Absolute Error (MAE) of several regression models and an ARIMA model are compared in the Table 1 for a specific dataset. The research shows that there are considerable temporal relationships in the data, with the ARIMA model performing much better than all other models. With the lowest MSE (44.796) and MAE (41.840), the ARIMA model performs much better than all other models, demonstrating its superior accuracy in this situation. With the lowest MSE (265424.179) and MAE (183.198) of all the regression models, Random Forest Regression appears to be the most successful non-time series model for the given data. Support Vector Regression's MAE (204.165) is comparable as well, but its MSE (330885.301) is greater. The MSE and MAE values of Lasso, Elastic Net, and Linear regressions are comparable, with Lasso regression marginally outperforming the others. Among the regression models, Decision Tree Regression performs the worst, as seen by its largest errors. Among the regression models, ARIMA is the most accurate overall for this dataset, followed by Random Forest.

CHAPTER 5

IMPACT ON SOCIETY, ENVIRONMENT, AND SUSTAINABILITY

5.1 Impact on Society

Economic Efficiency and Growth

Machine learning-driven sales forecasting has the potential to significantly boost economic efficiency across various industries. By enabling more accurate predictions of demand, businesses can optimize their supply chains, reduce waste, and improve resource allocation. This not only enhances profitability but also contributes to broader economic stability by aligning production closely with consumer needs, thereby supporting sustained business growth and innovation.

Consumer Benefits

Consumers stand to benefit from the application of machine learning in sales forecasting through better product availability and more competitive pricing. Accurate forecasts prevent overproduction and underproduction scenarios, which often lead to price volatility. Stable pricing and consistent product availability enhance consumer satisfaction and trust in brands, fostering a more reliable shopping experience.

Environmental Sustainability

Efficient inventory management powered by accurate sales forecasts can significantly reduce the environmental impact of overproduction. By producing only what is expected to sell, companies can minimize waste and the carbon footprint associated with the production, storage, and disposal of unsold goods. This contributes to more sustainable business practices and helps companies move towards their environmental responsibility goals.

Labor Market Adaptations

As machine learning transforms sales forecasting, there is a potential shift in the labor market within sectors that rely heavily on predictive analytics. While there may be a reduced need for traditional roles focused on manual analysis, there will be an increased demand for data scientists, machine learning specialists, and analytics consultants. This shift encourages the development of a more tech-savvy workforce and may lead to higher productivity and job creation in high-tech roles.

Social Inequalities

The integration of advanced technologies such as machine learning in sales forecasting could widen the gap between large corporations and small businesses. Large companies typically have more resources to invest in advanced analytics systems, potentially gaining a competitive advantage over smaller entities that cannot afford such technologies. Addressing this disparity requires policy interventions, such as providing access to affordable technologies and supporting digital education to level the playing field.

Adoption and Ethical Considerations

The widespread adoption of machine learning in sales forecasting must be managed ethically to avoid misuse of consumer data and ensure privacy. Companies must navigate the balance between leveraging data for business insights and respecting consumer rights, which calls for robust data governance policies and compliance with regulations such as GDPR.

Overall, the impact of machine learning on sales forecasting touches various aspects of society, from economic activities and environmental sustainability to labor markets and ethical considerations. While it presents numerous benefits, it also brings challenges that require thoughtful management and regulatory oversight to ensure that its potential is harnessed responsibly and inclusively.

5.2 Impact on Environment

Resource Optimization and Waste Reduction

Machine learning in sales forecasting can significantly contribute to environmental sustainability by optimizing resource allocation. By accurately predicting sales, businesses can tailor production schedules and inventory levels more precisely, reducing overproduction and minimizing waste. This not only helps in managing resources more efficiently but also in reducing the carbon footprint associated with the production and disposal of excess products.

Energy Efficiency

Machine learning models, especially when run on optimized, modern computing infrastructure, can enhance energy efficiency. Traditional sales forecasting methods might require extensive manual processes and repetitive tasks that consume more energy and resources. In contrast, machine learning can automate these processes, potentially reducing the energy consumption of data centers by ensuring that computational resources are used only when necessary and are scaled appropriately based on demand.

Supply Chain Streamlining

A precise sales forecast allows companies to streamline their supply chain operations. This efficiency minimizes the transportation needs by planning more direct routes and consolidating shipments, which in turn reduces fuel consumption and greenhouse gas emissions. Furthermore, better forecasting can help in reducing the reliance on expedited shipping methods, which are often more carbon-intensive.

Reduction in Chemical Usage

In industries like agriculture or manufacturing where products are perishable or have specific storage requirements, accurate sales forecasting can lead to a decrease in the use of chemicals used for preservation or speedy production. Reducing the production of surplus goods decreases the need for preservatives and other chemicals that may be harmful to the environment.

Support for Sustainable Practices

Accurate forecasts make it easier for businesses to commit to sustainable practices. With better visibility into future sales, companies can afford to prioritize environmentally friendly materials and processes without the risk of significant unsold inventory. This can encourage a shift toward more sustainable production practices industry-wide.

Economic Resilience

Enhanced forecasting contributes to economic resilience by enabling businesses to adjust quickly to changing market conditions. This adaptability can reduce the likelihood of economic shocks that can lead to abrupt changes in resource use, helping maintain a more consistent and predictable demand on natural resources and energy.

The integration of machine learning into sales forecasting not only optimizes business operations but also offers significant environmental benefits. By improving the accuracy of sales predictions, machine learning helps in reducing waste, saving energy, streamlining supply chains, and supporting sustainable business practices. These environmental impacts are crucial for building sustainable business models that can thrive without excessively straining our natural resources.

5.3 Ethical Aspects

The integration of machine learning (ML) into sales forecasting brings about a number of ethical considerations that must be addressed to ensure responsible use and maintain public trust. Below are key ethical aspects to consider when implementing ML for sales forecasting:

1. Data Privacy and Security: Sales data often includes sensitive information about customers and their buying habits. It's crucial to ensure that this data is handled with the highest level of confidentiality and security to protect customer privacy. Businesses must obtain explicit consent from individuals before using their data for forecasting or other analytical purposes, especially when data collection includes personal identifiers.

2. Bias and Fairness: Machine learning models can inadvertently perpetuate or amplify biases present in the training data. Rigorous testing and validation are required to identify and mitigate biases that could lead to unfair outcomes for certain customer groups or products. Ensuring that the training data represents all relevant demographics and market segments fairly is essential to prevent skewed or unjust predictive outcomes.

3. Transparency and Explainability: There should be clarity and openness about how forecasting models make predictions. This includes sharing information on the data inputs, model design, and decision processes. Sales forecasting models should be explainable to a wide range of stakeholders, including those without technical expertise, to ensure understanding and trust in how decisions are made.

4. Accountability: Organizations must take responsibility for the decisions made based on ML forecasting models, including addressing any negative consequences that arise. Continuous oversight is needed to monitor ML models for ongoing compliance with ethical standards and legal regulations. Establishing a governance framework helps in maintaining accountability and addressing issues proactively.

5. Impact on Employment: The automation of forecasting tasks might lead to job displacement. Organizations should consider the social impact of integrating these technologies and explore ways to retrain or repurpose affected employees. Investing in training for staff to adapt to new technologies and roles that emerge from ML deployment is crucial for maintaining an ethical approach to workforce management.

6. Consumer Impacts: There is a risk that predictive models might manipulate consumer behavior or limit consumer choices through targeted marketing. Businesses need to be aware of these impacts and strive to maintain ethical marketing practices. Ensuring that the benefits of improved forecasting, such as optimized pricing or better product availability, are shared equitably among all consumers is important.

Adhering to these ethical principles is essential not only for legal compliance but also for building and maintaining trust with customers and the public. Companies that prioritize these aspects are likely to benefit from greater customer loyalty and a stronger reputation, which are invaluable assets in today’s competitive business environment.

5.4 Sustainability Plan

To ensure the long-term viability and continuous improvement of the machine learning (ML) models used for sales forecasting, it is essential to develop a comprehensive sustainability plan. This plan focuses on maintaining and enhancing the accuracy, reliability, and operational efficiency of the ML models over time. Key components of the sustainability plan include:

1. Continuous Data Integration

Continuously feed new sales data into the machine learning models to keep them current and reflective of the latest market conditions. Incorporate a variety of data sources, including market trends, consumer behavior, and economic indicators, to enhance the models' robustness and accuracy.

2. Model Monitoring and Maintenance

Implement monitoring systems to track the performance of the ML models in real-time, identifying any deviations or declines in accuracy. Schedule periodic reviews and updates of the models to incorporate new data, adjust to changing market dynamics, and integrate advanced modeling techniques.

3. Training and Development

Provide regular training for data science and analytics teams on the latest ML techniques and tools to ensure they remain at the forefront of ML technology. Educate business stakeholders on the capabilities and limitations of ML forecasts to set realistic expectations and foster informed decision-making.

4. Technological Advancements

Stay abreast of advancements in artificial intelligence and machine learning, integrating new tools and algorithms that can enhance forecasting accuracy and efficiency. Ensure the infrastructure supporting the ML models can scale up or down based on business growth and data volume changes.

5. Ethical and Responsible AI Use

Regularly review and address potential biases in data and model outputs to ensure fair and equitable sales forecasting. Develop models that are not only accurate but also transparent and explainable, allowing users to understand how predictions are made.

6. Feedback Loop Integration

Establish mechanisms to capture feedback from users of the forecasting models to continuously improve their functionality and user experience. Use feedback and performance data to make iterative improvements to the models, ensuring they remain effective and relevant.

7. Financial Planning

Monitor and manage the costs associated with the deployment and maintenance of ML models, ensuring they deliver value for money. Allocate budgets for ongoing research and development in ML applications to maintain competitive advantage.

By implementing these strategies, the sales forecasting ML models can be maintained sustainably, ensuring they continue to provide high-quality, accurate predictions that support business objectives and growth over the long term. This sustainability plan is designed to adapt to changes in technology and market conditions, thereby securing the ongoing effectiveness and relevance of the sales forecasting initiative.

CHAPTER 6

CONCLUSION AND FUTURE WORK

6.1 Summary of the study

This thesis explored the application of machine learning (ML) techniques to enhance the accuracy and reliability of sales forecasting. Through the comparative analysis of various ML models, including ensemble methods, neural networks, and regression analyses, it was determined that these advanced computational methods significantly outperform traditional forecasting approaches. The ARIMA model, in particular, demonstrated superior performance, providing highly accurate forecasts as evidenced by its low Mean Squared Error (MSE) and Mean Absolute Error (MAE).

Key findings of the research indicated that machine learning models could effectively utilize large datasets encompassing diverse variables—from historical sales data to real-time market trends—to generate nuanced and dynamically updated sales predictions. These models not only offer greater accuracy but also adapt to new data, allowing for continuous improvement in forecast quality.

Furthermore, the study addressed the broader implications of using ML in sales forecasting, highlighting both the potential benefits in terms of operational efficiency and strategic planning, as well as the ethical considerations important for responsible use. The integration of machine learning into sales forecasting processes promises significant enhancements in decision-making capabilities, offering businesses a critical edge in the competitive market landscape.

6.2 Conclusions

This thesis has successfully demonstrated the efficacy of machine learning (ML) techniques in enhancing the accuracy and reliability of sales forecasting. Throughout this research, it has become evident that ML provides a substantial improvement over traditional forecasting methods, which often struggle to adapt to the dynamic and complex nature of modern market environments. By leveraging a diverse array of algorithms—including ensemble methods, neural networks, and regression analysis—the study has shown that machine learning can effectively process and analyze vast datasets from multiple sources to produce more precise forecasts.

The findings highlight that the ARIMA model, in particular, outperforms other models in forecasting accuracy as evidenced by its lower Mean Squared Error (MSE) and Mean Absolute Error (MAE). This superiority underscores the potential of time-series analysis in capturing the intricate patterns of sales data, which are crucial for making informed business decisions.

Moreover, the thesis has addressed the broader impacts of implementing machine learning in sales forecasting, discussing both the potential benefits and the ethical considerations. Machine learning's ability to enhance decision-making in inventory management, resource allocation, and strategic planning can lead to significant economic efficiencies and growth. However, this technological integration must be managed with a keen awareness of privacy, bias, and the need for transparency to ensure ethical usage and sustain public trust.

In conclusion, the research presented in this thesis provides compelling evidence that machine learning is not merely a technological advancement in forecasting but a transformative tool that can significantly enhance business operations. The continued exploration and improvement of these models are recommended, with a focus on integrating real-time data analysis and expanding the adaptability of the models to further enhance their predictive power. This ongoing development will pave the way for smarter, more efficient business practices that are responsive to the ever-evolving market demands.

6.3 Implication for further study

The promising results obtained from the integration of machine learning techniques into sales forecasting lay the groundwork for numerous avenues of further research:

Advanced Machine Learning Models: Future studies could explore the use of more advanced machine learning models such as deep learning and reinforcement learning. These models may offer enhanced capabilities in handling non-linear relationships and high-dimensional data, potentially leading to even more accurate forecasts.

Real-time Data Integration: Investigating the impact of incorporating real-time data streams, such as instant social media feedback and live market trends, could provide insights into the dynamic nature of consumer behavior and market conditions. This could help in developing models that are more responsive and adaptive to sudden market changes.

Cross-industry Validation: Expanding the research to include multiple industries could validate the versatility and applicability of machine learning models across different market contexts. Each industry comes with its unique challenges and dynamics, and a broad-based study could help in fine-tuning the models to specific needs.

Global Market Forecasting: Extending the models to forecast sales in diverse geographical markets could help multinational corporations make more informed global strategies. Different regions may exhibit unique consumer behaviors and economic conditions, and machine learning models need to be adapted accordingly.

Impact of External Shocks: Researching how machine learning models handle external shocks such as economic crises or pandemics could provide valuable insights. Understanding the resilience and adaptability of these models in extreme conditions can help in designing systems that are robust under a variety of scenarios.

Cost-benefit Analysis: An in-depth cost-benefit analysis of implementing machine learning solutions in sales forecasting would help businesses in decision-making processes. This includes assessing the infrastructure costs, the required expertise, and the potential return on investment.

References

[1] V. Sohrabpour, P. Oghazi, R. Toorajipour, and A. Nazarpour, "Export sales forecasting using artificial intelligence," *Technol. Forecast. Soc. Change*, vol. 163, no. 120480, p. 120480, 2021.

[2] F. Haselbeck, J. Killinger, K. Menrad, T. Hannus, and D. G. Grimm, "Machine learning outperforms classical forecasting on horticultural sales predictions," *Mach. Learn. Appl.*, vol. 7, no. 100239, p. 100239, 2022.

[3] S. T. Londhe and S. Palwe, "Customer-centric sales forecasting model: RFM-ARIMA approach," *Bus. Syst. Res. J.*, vol. 13, no. 1, pp. 35–45, 2022.

[4] A. Krishna, Akhilesh, A. Aich, and C. Hegde, "Sales-forecasting of retail stores using machine learning techniques," in *2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS)*, 2018.

[5] V. Wineka Nirmala, D. Harjadi, and R. Awaluddin, "Sales forecasting by using exponential smoothing method and trend method to optimize product sales in PT. Zamrud Bumi Indonesia during the covid-19 pandemic," *Int. J. Eng. Scie. and Inform. Technology.*, vol. 1, no. 4, pp. 59–64, 2021.

[6] M. A. I. Arif, S. I. Sany, F. I. Nahin, and A. S. A. Rabby, "Comparison study: Product demand forecasting with machine learning for shop," in *2019 8th International Conference System Modeling and Advancement in Research Trends (SMART)*, 2019.

[7] I.-F. Chen and C.-J. Lu, "Sales forecasting by combining clustering and machine-learning techniques for computer retailing," *Neural Comput. Appl.*, vol. 28, no. 9, pp. 2633–2647, 2017.

[8] M. Gurnani, Y. Korke, P. Shah, S. Udmale, V. Sambhe, and S. Bhirud, "Forecasting of sales by using fusion of machine learning techniques," in *2017 International Conference on Data Management, Analytics and Innovation (ICDMAI)*, 2017.

[9] W. K. Wong and Z. X. Guo, "Intelligent sales forecasting for fashion retailing using harmony search algorithms and extreme learning machines," in *Optimizing Decision Making in the Apparel Supply Chain Using Artificial Intelligence (AI)*, Elsevier, 2013, pp. 170–195.

[10] N. Harz, S. Hohenberg, and C. Homburg, "Virtual reality in new product development: Insights from prelaunch sales forecasting for durables," *J. Mark.*, vol. 86, no. 3, pp. 157–179, 2022.

[11] D. S. Rajpoot, B. Mittal, H. Dudani, and U. Singhal, "Sales analysis and forecasting using machine learning approach," in *Proceedings of the 2023 Fifteenth International Conference on Contemporary Computing*, 2023.

[12] D. Rout, B. Roy, and P. Kapse, "A subtle design of prediction models using machine learning algorithms for advocating selection and forecasting sales of garments: A case study," in *Advances in Data-Driven Computing and Intelligent Systems*, Singapore: Springer Nature Singapore, 2024, pp. 387–397.

[13] J. S. Kiran, P. S. V. S. Rao, P. V. R. D. P. Rao, B. S. Babu, and N. Divya, "Analysis on the prediction of sales using various machine learning testing algorithms," in *2022 International Conference on Computer Communication and Informatics (ICCCI)*, 2022.

[14] B. Kumar Jha and S. Pande, "Time series forecasting model for supermarket sales using FB-prophet," in *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, 2021.

[15] Y. Li, Y. Yang, K. Zhu, and J. Zhang, "Clothing sale forecasting by a composite GRU–prophet model with an attention mechanism," *IEEE Trans. Industr. Inform.*, vol. 17, no. 12, pp. 8335–8344, 2021.

[16] G. Tsoumakas, "A survey of machine learning techniques for food sales prediction," *Artif. Intell. Rev.*, vol. 52, no. 1, pp. 441–447, 2019.

[17] R. S. Sengar and D. F. Ahmed, "Review on trends in machine learning applied to demand & sales forecasting," *SMART MOVES JOURNAL IJOSCIENCE*, vol. 5, no. 6, p. 4, 2019.

[18] Y. K. Elalem, S. Maier, and R. W. Seifert, "A machine learning-based framework for forecasting sales of new products with short life cycles using deep neural networks," *Int. J. Forecast.*, vol. 39, no. 4, pp. 1874–1894, 2023.

[19] R. Puspita and L. A. Wulandhari, "Hardware sales forecasting using clustering and machine learning approach," *IAES Int. J. Artif. Intell. (IJ-AI)*, vol. 11, no. 3, p. 1074, 2022.

[20] M. O. Anwer and S. Akyuz, "Sales forecasting of a hypermarket: Case study in Baghdad using machine learning," in *2022 30th Signal Processing and Communications Applications Conference (SIU)*, 2022.

[21] Y. Ensafi, S. H. Amin, G. Zhang, and B. Shah, "Time-series forecasting of seasonal items sales using machine learning – A comparative analysis," *International Journal of Information Management Data Insights*, vol. 2, no. 1, p. 100058, 2022.

[22] D. M. U. Ashraf, "A predictive analysis of retail sales forecasting using machine learning techniques," *Research Journal Of Computer Science And Information Technology*, vol. 6, no. 04, pp. 23–33, 2022.

[23] R. S. Mallik, R. Abhiram, S. R. Reddy, and R. M. Jagadish, "A comprehensive survey on sales forecasting models using machine learning algorithms," in *2022 Fourth International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT)*, 2022.

[24] O. M. E. Team, "10 popular regression algorithms in Machine Learning," *Online Manipal*, Oct. 11, 2023. https://www.onlinemanipal.com/blogs/popular-regression-algorithms-in-machine-learning.

Plagiarism Report:

Sales Forecasting

ORIGINALITY REPORT

11%	8%	4%	7%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	3%
2	Submitted to Daffodil International University Student Paper	2%
3	Submitted to National College of Ireland Student Paper	1%
4	Submitted to University of Tampa Student Paper	1%
5	Submitted to University of Central Lancashire Student Paper	1%
6	Submitted to Higher Education Commission Pakistan Student Paper	1%
7	Submitted to Delaware Military Academy Student Paper	1%
8	Submitted to National University of Ireland, Galway Student Paper	1%
9	ijmaberjournal.org	

	Internet Source	1%
10	curve.carleton.ca Internet Source	1%
11	Submitted to ebsu Student Paper	1%
12	ijrpr.com Internet Source	1%