**A Final Year Project Report**

**Full Unit: Final Report**

**Project** **Title: Predictive Analytics for Sales and Profit Using Machine Learning.**

Student’s first and last name

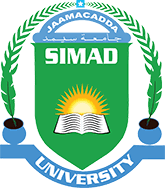
Bishaar Ahmed Ali

Ahmed Aden Mohamud

A report submitted in part fulfilment of the degree of

**BIT28 Information Technology**

**Supervisor:** Dr Abdijaliil Abdulahi Mohamed



Department of Information Technology

Faculty of computing, Simad University

**August 2025**

**DECLARATION. A**

“We declare that the following is our own work and does not contain any unacknowledged work from any other sources. This project was undertaken to fulfil the requirements of the bachelor’s degree program in Information Technology at SIMAD University”.

Name: Bishaar Ahmed Ali

Student Signature: ………………………………….

Name: Ahmed Aden Mohamud

Student Signature: ………………………………….

Date: 5/8/2025

**DECLARATION. B**

I hereby declare that I have reviewed this project proposal titled ‘Predictive Analytics for Sales and Profit Using Machine Learning’ and in my opinion, it is adequate in terms of scope and quality for consideration as a proposal project.

Name of Supervisor: ABDULAZIZ YASIN NAGEYE

Signature: ………………………

Date: /…/ 2024

**ACKNOWLEDGEMENTS**

**Student One:**

I would like to express my deepest gratitude to everyone who contributed to the successful completion of this research. This journey would not have been possible without the encouragement, guidance, and support of numerous individuals. I am especially grateful to my family and friends, whose unwavering belief in my abilities provided the motivation to overcome challenges and strive for excellence. Their constant encouragement and emotional support played a crucial role in keeping me focused and determined throughout this academic endeavour. I also extend my sincere appreciation to my peers and colleagues, whose insightful discussions, constructive feedback, and shared experiences enriched my understanding and helped refine my ideas. Additionally, I am indebted to the many researchers and experts in the field of predictive analytics and machine learning, whose work provided the foundation for this study and inspired me to explore innovative approaches to sales and profit forecasting. Lastly, I acknowledge the institutions, online platforms, and open-access research databases that provided me with valuable resources and datasets, making this research possible and enabling me to contribute meaningfully to this area of study.

**Student Two:**

I would like to extend my sincere appreciation to Dr. Abdijaliil Abdulahi Mohamed for his invaluable guidance, patience, and mentorship throughout this research. His expertise, constructive feedback, and unwavering support have been instrumental in shaping the direction of this study, allowing me to refine my ideas and develop a deeper understanding of predictive analytics and machine learning. His dedication to academic excellence and his encouragement to think critically and independently have been truly inspiring, motivating me to push beyond my limits and achieve my research objectives. Furthermore, I am deeply grateful to SIMAD University for providing an excellent academic environment, research facilities, and access to vital learning resources that have been essential in the successful completion of this study. The university's commitment to fostering intellectual growth and innovation has greatly contributed to my educational development, equipping me with the knowledge and skills necessary to undertake this research. I am honoured to be a part of this esteemed institution and truly appreciate the opportunities it has provided me to grow academically and professionally.

**ABSTRACT**

Predictive analytics has emerged as a vital tool for businesses aiming to enhance sales and profit forecasting by leveraging machine learning techniques to analyse historical sales data, Traditional forecasting methods, such as statistical regression and time series analysis, often fail to capture complex data patterns and adapt to rapidly changing market conditions, leading to inaccurate predictions and suboptimal decision-making. This study explores the implementation of machine learning algorithms, including regression models, decision trees, and ensemble learning, to improve forecasting accuracy and provide businesses with actionable insights. The proposed system integrates real-time data processing, advanced feature selection, and enhanced model interpretability to support data driven decision making in key areas such as inventory management, pricing strategies, and financial planning. By evaluating multiple machine learning models and comparing their performance using key metrics such as Mean Absolute Error (MAE) and precision accuracy, the research demonstrates that machine learning based predictive analytics significantly outperforms traditional forecasting methods in terms of accuracy, scalability, and adaptability. This study contributes to the growing field of data driven financial forecasting by presenting a scalable and efficient approach that enhances business decision making and long-term strategic planning.

Table of Contents

[**1.1 INTRODUCTION** 1](#_Toc202083729)

[**1.2 BACKGROUND** 2](#_Toc202083730)

[**1.4 OBJECTIVES** 6](#_Toc202083731)

[**1.5 RESEARCH QUESTIONS** 7](#_Toc202083732)

[**1.6 SCOPE OF THE STUDY** 7](#_Toc202083733)

[**1.7 SIGNIFICANCE** 8](#_Toc202083734)

[**CHAPTER 2** 9](#_Toc202083735)

[**2.1 INTRODUCTION** 9](#_Toc202083736)

[**2.2 OVERVIEW OF THE SYSTEM** 9](#_Toc202083737)

[**2.3 TECHNOLOGIES USED** 11](#_Toc202083738)

[**2.3.1 Machine Learning Models** 11](#_Toc202083739)

[**2.3.2 Big Data Technologies** 13](#_Toc202083740)

[**2.4. RELATED WORKS** 19](#_Toc202083741)

[**2.4.1 Machine Learning Algorithms for Sales Forecasting** 19](#_Toc202083742)

[**2.4.2 Comparison of Machine Learning and Deep Learning Models** 20](#_Toc202083743)

[**2.4.3 Hybrid Models for Sales Forecasting** 20](#_Toc202083744)

[**2.4.4 Time Series Forecasting and Ensemble Approaches** 20](#_Toc202083745)

[**2.4.5 Advanced Machine Learning Techniques in Retail Sales Forecasting** 21](#_Toc202083746)

[**2.5. GAP ANALYSIS** 22](#_Toc202083747)

[**2.5.1 Gaps in Existing Systems.** 22](#_Toc202083748)

[**2.6. CHAPTER SUMMARY** 24](#_Toc202083749)

[**3.1 INTRODUCTION** 26](#_Toc202083750)

[**3.2 USER REQUIREMENT GATHERING AND ANALYSIS** 26](#_Toc202083751)

[**3.2.1 Preliminary Investigation** 27](#_Toc202083752)

[**3.2.2 User Requirement Analysis** 27](#_Toc202083753)

[**3.2.3 Data Gathering** 28](#_Toc202083754)

[**3.3 SYSTEM DESIGN** 28](#_Toc202083755)

[**3.3.1 Data Acquisition** 29](#_Toc202083756)

[**3.3.2 Data Pre-processing** 29](#_Toc202083757)

[**3.3.3 Model Training** 30](#_Toc202083758)

[**3.3.4 Real-Time Prediction** 30](#_Toc202083759)

[**3.4 REQUIREMENT ANALYSIS** 31](#_Toc202083760)

[**Flowchart Design for the System** 31](#_Toc202083761)

[**3.5 USER REQUIREMENT SPECIFICATION** 31](#_Toc202083762)

[**3.5.1 New Proposed System** 32](#_Toc202083763)

[**3.5.2 Solution Strategy** 32](#_Toc202083764)

[**3.6 SYSTEM REQUIREMENT SPECIFICATION** 33](#_Toc202083765)

[**3.6.1 System Interface** 34](#_Toc202083766)

[**3.6.4 Software Interface** 34](#_Toc202083767)

[**3.7 CHAPTER SUMMARY** 35](#_Toc202083768)

[**CHAPTER 4** 36](#_Toc202083769)

[**4.1 INTRODUCTION** 36](#_Toc202083770)

[**4.2 IMPLEMENTATION** 36](#_Toc202083771)

[**4.2.1 Software Implementation** 36](#_Toc202083772)

[**4.2.2 Hardware Implementations** 40](#_Toc202083773)

[**4.2.3 Integration of Software and Hardware** 41](#_Toc202083774)

[**4.3 TESTING AND EVALUATION** 43](#_Toc202083775)

[**4.3.1 Evaluation Metrics** 43](#_Toc202083776)

[**4.3.2 Overview of the Dataset** 43](#_Toc202083777)

[**4.4 PERFORMANCE EVALUATION** 45](#_Toc202083778)

[**4.4.1 Evaluation Metrics** 45](#_Toc202083779)

[**4.4.2 Training Accuracy** 46](#_Toc202083780)

[**4.4.3 Validation Accuracy** 46](#_Toc202083781)

[**4.5 RESULTS** 46](#_Toc202083782)

[**CHAPTER 5** 49](#_Toc202083783)

[**REFERENCES** 52](#_Toc202083784)

**LIST OF TABLES**

Table 2.1: Technologies Used and Limitations

Table 2.2: Technologies Used and Limitations (Comparison of Traditional Statistical Models and Machine Learning Models)

**LIST OF FIGURES**

[Figure 2.0.1: The Architecture of ARIMA Model 12](#_Toc202083629)

**LIST OF ABBREVIATIONS AND SYMBOLS**

AI - Artificial Intelligence

ARIMA - Autoregressive Integrated Moving Average

AutoML - Automated Machine Learning

BERT - Bidirectional Encoder Representations from Transformers

ETS - Exponential Smoothing

GPT - Generative Pre-trained Transformer

LSTM - Long Short-Term Memory

MAE - Mean Absolute Error

NLP - Natural Language Processing

RNN - Recurrent Neural Network

RMSE - Root Mean Squared Error

SMEs - Small and Medium Enterprises

XAI - Explainable AI

XGBoost - Extreme Gradient Boosting

**CHAPTER 1**

**INTRODUCTION**

## **1.1 INTRODUCTION**

Data driven decision making is critical for companies to remain ahead of the competition in today's fast moving business environment. Predictive analytics trained through machine learning has now become a game changing tool that lets businesses look ahead into the future, shape up their strategies, and increase their profitability. Machine learning models study historical sales data to find patterns, correlations, and market dynamics that generally elude traditional statistical methods.

Sales forecasting has always been an issue of business strategizing, serving to help enterprises plan inventory and allocate resources for financial targets. However, traditionally, the forecasting strategies, such as simple regression or moving averages, fall short to model the complicated and dynamic characteristics of real life markets. Thus, machine learning steps in, using lots of data to uncover hidden knowledge and improve performance over time. It could now allow businesses to make better decisions by being able to predict sales figures, customer behaviours, seasonal demand fluctuations, and influences of other external markets.

In this chapter, we will talk about how profit prediction is very important in financial planning, investment decision, and risk management. Companies need to be able to forecast future revenue and expenses correctly to ensure long term survival. Machine learning algorithms can process a number of variables from sales performance and operational cost to market forces and social media sentiment hence enabling an exhaustive and precise profit prediction. This enables businesses to fine tune pricing strategies, eliminate wastage, and enhance revenue opportunities.

## 

## **1.2 BACKGROUND**

The emergence of predictive analytics as a business decision making tool can be attributed to the growing availability of data and advancements in machine learning techniques. Companies produce huge volumes of both structured and unstructured data, from historic sale data to customer engagement and market dynamics. Being able to analyse such data effectively and derive actionable insights has become one of the principal determinants of a competitive advantage. Predictive analytics, especially when coupled with machine learning algorithms, allows companies to project future sales outcomes and profit margins more accurately compared to conventional forecasting techniques (Anand et al., n.d.)

Forecasting sales and profit has been traditionally based on statistical techniques like time series analysis, moving averages, and linear regression. Though these provided some predictive capability, they tended to be poor at addressing the complexity of real-world business situations, which feature nonlinear relationships and exogenous market influences. Machine learning algorithms, in contrast, have the ability to trawl through large datasets and reveal concealed patterns to provide more dependable and sensitive forecasts (Ghude et al., 2024)

Research carried out by Pooja (Ghude et al., 2024)illustrated how the use of machine learning models, i.e., Random Forest and XGBoost, is better than conventional statistical approaches at forecasting future sales. The aforementioned models have the capability to capture seasonality, market volatility, and shifting consumer trends, which makes them highly beneficial for businesses that function in volatile industries. Likewise, the research by (Van Calster et al., 2020) underscores the potential of profit-maximizing forecasting methods, in combination with machine learning, to enable companies not just to predict sales figures but also to optimize prices to achieve maximum profitability.

Machine learning algorithms use historical data to make precise predictions by identifying patterns that might not be immediately apparent to human analysts. Approaches like regression models, decision trees, and deep learning models have been applied extensively in the field of sales forecasting, demonstrating large accuracy gains (Anand et al., 2019)For example, (Rezazadeh, 2020)presented a cloud-based machine learning framework that enhanced the prediction capabilities of B2B sales forecasting. By integrating information from multiple sources and ensemble learning techniques, the study achieved improved prediction compared to deep learning models.

Moreover, the research work of (Mirshekari et al., 2024)investigated the application of Gaussian Process models in sales forecasting with the employment of Bayesian optimization for enhancing forecast accuracy. From the findings, it is evident that machine learning-based sales forecasting models can provide noteworthy accuracy gains by adaptively responding to new market conditions. Such sophisticated models allow companies to respond proactively instead of reactively, thereby decreasing financial risks and enhancing decision-making approaches.

Predictive analytics is being embraced in different industries, each using machine learning to enhance sales and profit forecasting in its own manner. The retail sector, for instance, applies predictive analytics to forecast customer demand, streamline inventory management, and maximize marketing campaigns. Research by (Anand et al., 2019)shows that the use of AI-driven demand forecasting in e-stores significantly improves conversion rates and loyalty as a result of the confidence that businesses have in terms of optimum stock levels without unnecessary losses from overstocking or stockouts.

In finance, predictive analytics enables prediction of profit by examining market trends, creditworthiness of customers, and investment risk assessment. Financial institutions employ machine learning algorithms to predict future revenue streams and enhance lending strategies for disbursement. Research work by (Kumar et al., 2024) demonstrated that financial institutions that leverage predictive analytics achieved 15-20% growth in profit margins because of enhanced risk assessment and fraud detection.

The manufacturing sector also gains an advantage from predictive analytics through enhancing supply chain processes and lowering operational expenses. Machine learning algorithms analyse past sales data in conjunction with variables like production expenses, supplier performance, and logistical limitations to enhance pricing models. Research by (Rezazadeh, 2020) indicates that machine learning-based predictive analytics can enhance supply chain efficiency by as much as 25%, thereby minimizing wastage and elevating overall profit margins.

Despite the clear advantages of machine learning in the areas of sales and profit prediction, several issues need to be addressed. One of the major limitations relates to data quality. Noisy or incomplete data sets may lead to flawed predictions, thus making data pre-processing a crucial step in building reliable models (Ghude et al., 2024). In addition, the interpretability of machine learning models remains a problem for many businesses. Even though techniques such as deep learning provide accuracy that is high, their incomprehensible "black box" is challenging for decision-makers who need to comprehend the mechanism of prediction generation (Van Calster et al., 2020)

Another key challenge is the need for periodic model refreshes. Business circumstances are not static, and fixed models will lose their performance value after a while. A study by (Ito & Fujimaki, 2016) stressed the importance of incorporating real-time data streams into predictive analytics models to attain high forecast accuracy.

Furthermore, the ethical implications of predictive analytics must be considered. Greater dependence on machine learning in decision-making brings data privacy, algorithmic fairness, and transparency into question. If not monitored, predictive models can unknowingly reinforce biases in data that have built up over the years, resulting in unfair or discriminatory business actions (Raji & Buolamwini, 2022). To resolve these issues, a balanced strategy that encompasses strong data governance, ethical AI practices, and adherence to regulations is required so that predictive analytics serves the interests of all the stakeholders equally.

As machine learning continues to advance, there are various future trends that are likely to influence the direction of predictive analytics in sales and profit forecasting. One of the major developments includes more utilization of deep learning techniques, allowing for the creation of more sophisticated and precise predictive models. Technological advancement in natural language processing (NLP) and sentiment analysis is enabling organizations to incorporate customer sentiments, social media opinions, and online reviews into predictive models, thereby enhancing the precision of predictions (Aguilar-Moreno et al., 2024)

Furthermore, explainable AI (XAI) is increasingly being utilized to solve the interpretability problem of machine learning models. Insights into how predictions are made by XAI methods enhance trust and usage among business executives. Studies by (Amrutkar & Mahadik, 2022) finds that companies adopting explainable AI in predictive analytics processes have increased confidence in decisions made by AI and better business results.

The integration of edge computing and real-time analytics is a conspicuous trend. Enterprises are now utilizing both cloud-based and edge-based predictive models more than ever before to aid real-time decision-making processes. A study conducted by (Ghude et al., 2024) underscores that the use of real-time predictive analytics has the potential to minimize revenue loss by as much as 30% for industries with volatile market conditions, including retail and finance, through real-time price and promotional offer changes by companies.

Machine learning applied to predictive analytics for sales and profit has revolutionized how businesses go about forecasting. By utilizing historical data and complex algorithms, companies can more precisely predict, optimize pricing, and enhance overall financial performance. Effective implementation, however, requires surmounting challenges in data quality, model interpretability, and flexibility. Also shaping the future of predictive analytics are ethical issues and emerging trends such as explainable AI and real-time analytics. As technology progresses, machine learning-based predictive analytics is destined to be an even more integral component of business decision-making (Ganguly & Mukherjee, n.d.; Navya et al., 2024; S, n.d.)

**1.3 PROBLEM STATEMENT**

In a perfect theoretical world, organizations employ machine learning based predictive analytics to generate extremely accurate predictions of sales and profits. With the strength of accurate predictions, companies are able to streamline pricing strategies, better manage inventories, and distribute resources effectively. The process promotes consistent revenue growth, minimizes financial risks, and improves strategic planning. The capacity to process information in real time enables companies to quickly react to shifting market forces, consumer trends, and economic conditions, thereby achieving competitive advantage. Therefore, companies are able to maximize profit margins and simultaneously achieve operation effectiveness and meet customer expectations.

However, without the use of predictive analytics, companies rely on old or experience-based forecasting methods that often make incorrect predictions. Erroneous sales forecasting could lead to excessive stock, thereby resulting in increased costs and waste or low inventory levels, with associated missed revenues and unhappy customers. Additionally, inefficient resource deployment and wasteful pricing tactics reduce profit margins. Small and medium-sized enterprises, specifically, encounter significant difficulties regarding these challenges due to their potential lack of the requisite expertise or resources to effectively analyse extensive datasets, thereby rendering them more susceptible to financial setbacks.

The implications of inaccurate forecasting can be extensive, ranging from decreased profits, ineffective strategic planning, and market share loss. Businesses that fail to respond to changes in the market risk losing ground to rivals who use data informed decision-making processes. Finally, financial instability can cause budget reductions, workforce retrenchment, or business collapse. With the growing competition and data-driven environment, the application of predictive analytics using machine learning has transitioned from a choice to a necessity for companies pursuing long-term profitability and viability.

## **1.4 OBJECTIVES**

1. To propose a machine learning-based system to improve sales and profit forecasting accuracy.
2. To design a predictive model that leverages historical data for better decision-making.
3. To develop and integrate a multi model machine learning system that compares and selects the most accurate predictor for sales and profit forecasting through a user-friendly interface.

## **1.5 RESEARCH QUESTIONS**

1. How can machine learning improve the accuracy of sales and profit forecasting compared to traditional methods?
2. What key factors and data sources should be considered when designing an effective predictive model for business profitability?
3. How can a multi model machine learning system be designed to compare and select the most accurate predictor for sales and profit forecasting?

## **1.6 SCOPE OF THE STUDY**

This research investigates the use of machine learning algorithms in predictive analytics for sales and profit forecasting. It investigates the extent to which algorithms like regression models, and decision trees enhance forecasting precision over conventional methods. The study is applied to the retail and e-commerce industries, where predictive analytics is important for inventory management, price strategy development, and financial planning. A complete forecasting model will be developed by analysing various sources of data, including historical sales data.

The scope of the research is confined to predictive system design and evaluation, as opposed to the creation of a complete commercial application. The model will be developed with the help of technologies like Python, Scikit-learn, and FAST API, thereby making it versatile. Public datasets will be used because of the difficulty in obtaining real-time business data. Ethical considerations such as data privacy, algorithmic bias, and model interpretability are noted but will not be discussed in detail. The aim is to examine how effective the application of machine learning techniques can be in enhancing sales and profit forecasting, together with describing the key advantages and drawbacks of the approach.

Additionally, the scope of this study is limited to designing and testing a predictive analytics model rather than developing a fully integrated commercial application. We will use publicly available datasets, as obtaining real time business data may present challenges due to confidentiality and access restrictions. The study also acknowledges potential ethical considerations, such as data privacy, algorithmic bias, and model interpretability, but will not explore these issues in depth. Instead, the primary focus will be on assessing the effectiveness of machine learning techniques in improving sales and profit forecasting accuracy while outlining the potential benefits and limitations of such models.

## **1.7 SIGNIFICANCE**

The study benefits businesses, financial analysts, data scientists, and researchers by demonstrating how machine learning can enhance sales and profit forecasting. Accurate forecasting enables more informed financial decisions, resource allocation, and reduced market uncertainties. Offering a data driven solution, the study aids business stability and long term growth.

E-commerce and retail companies can improve forecasting to avoid overstocking, understocking, and mispricing. Machine learning streamlines the inventory, adjusts the price dynamically, and assigns resources. Small and medium enterprises can leverage these models to stay competitive and increase profitability. Predictive analytics is utilized by financial decision-makers and analysts for budgeting, risk assessment, and strategic planning, which leads to intelligent investment decisions and reduced financial losses.

Along with businesses, consumers and the economy also benefit. Accurate sales predictions improve inventory management, price stability, and customer satisfaction. Stable businesses are less likely to become unstable, downsize, or close, fostering economic growth and employment stability. Ultimately, this study illustrates how machine learning is transforming financial forecasting, driving success in the data age.

# **CHAPTER 2**

**LITERATURE REVIEW**

## **2.1 INTRODUCTION**

This chapter is a detailed overview of the leading concepts, technologies, and research on using predictive analytics for sales and profit forecasting using machine learning. It explains the principles behind predictive analytics, with the focus on how businesses use data-driven approaches to better make financial decisions. By overviewing what has been accomplished and what technologies are available, this chapter sets the stage to see how machine learning enhances the accuracy of forecasting.

Besides, the chapter summarizes literature pertinent to the assessment of methods and outcomes of existing research. By identifying the strengths and weaknesses in existing research, it points out gaps that this research seeks to fill. The discussion further involves the particular machine learning algorithms, data repositories, and testing approaches commonly employed in sales forecasting. In addition, the chapter explains the technologies used in previous systems, describing the applicability of these technologies to the system under proposal.

Lastly, the chapter presents the limitations inherent in conventional forecasting approaches and drawbacks of existing machine learning applications in sales and profit forecasting fields. By doing so, it provides a platform for the study by illustrating the necessity of improved predictive analytics models. The chapter concludes with a summary that relates the literature reviewed to this study's aim, indicating the significance of the suggested methodology.

## **2.2 OVERVIEW OF THE SYSTEM**

The predictive analytics model used in profit and sales forecasting seeks to utilize machine learning techniques to bolster organizational decision-making. Conventional forecasting techniques like time series analysis and statistical regression are limited in dealing with intricate datasets and new market developments (Anand et al., 2019). Machine learning overcomes these limitations by processing large datasets, uncovering hidden patterns, and constantly refining predictions by incorporating new information.

The system has several core components that integrate to offer precise and real-time predictions of sales and profitability. To begin with, the Data Collection Module gathers historical sales data, customer trends, and market trends from various sources such as in-house business databases, cloud storage systems, and external APIs (Altarazi & Santos, 2024). The raw data is then taken through the Data pre-processing Module, where inconsistencies, duplicates, and missing values are cleaned to maintain data integrity.

Following the data cleansing process, the Feature Selection & Engineering Module determines relevant determinants that drive sales and profitability. The literature suggests that the addition of both internal determinants (such as historical sales data and product prices) and external determinants (such as economic statistics and prices offered by competitors) significantly improves the predictive accuracy (Sai et al., 2021). The Machine Learning Model subsequently applied regression, decision trees, and ensemble learning methods to make predictions with a focus on maximum accuracy and reliability (Pavlyshenko, n.d.)

In an effort to promote continuous applicability of the forecasts, the Real-Time Data Integration Module constantly feeds the model with current sales information, thereby facilitating timely responses by businesses to market fluctuations (Bajaj et al., 2020). Lastly, the Visualization & Reporting Module conveys the findings via dashboards, graphs, and automatic reports, thereby enabling business managers to view the projections and make rational financial decisions (Ito & Fujimaki, 2016).

This system bridges the gap between traditional forecasting limitations and modern machine learning capabilities, providing a responsive and scalable solution to profit and sales forecasting. By leveraging the integration of disparate data sources and advanced analytics utilization, it drives maximum business efficiency and profitability while improving long-term strategic planning (Navya et al., 2024).

## **2.3 TECHNOLOGIES USED**

The field of predictive analytics for sales and profit forecasting has grown extensively due to the application of diverse technologies, from conventional statistical methods to cutting edge machine learning and artificial intelligence (AI) techniques. Various companies in different sectors employ a mix of data processing software, cloud computing services, and machine learning libraries to maximize the precision of their sales forecasts. This section reflects on the primary technologies that have extensively been applied in solving this problem, observing their functions, strengths, and limitations. Among the popular practices in sales forecasting is the application of statistical models, including time series analysis, moving averages, and regression models.

## **2.3.1 Machine Learning Models**

Time series models such as Autoregressive Integrated Moving Average (ARIMA) and Exponential Smoothing (ETS) have been widely applied in business forecasting because they are capable of recognizing trends and seasonality. More specifically, ARIMA is helpful in examining historical sales and generating short run forecasts according to patterns derived from history. Yet, traditional statistical models break down with large datasets and lack the level of flexibility necessary to respond rapidly to dramatic changes in markets or to nonlinearity in data. Ever since machine learning came into the limelight, firms shifted to employing supervised learning algorithms such as linear regression, decision trees, random forests, and neural networks for improved forecast accuracy. Linear regression is still a popular baseline model, especially for modelling the relation between sales performance and a set of influencing variables, e.g., advertising expenditure, economic indicators, and seasonal effects.

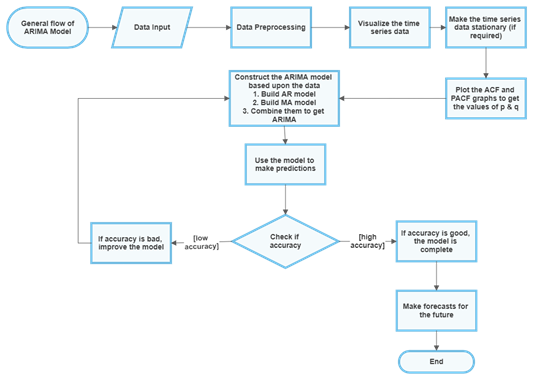
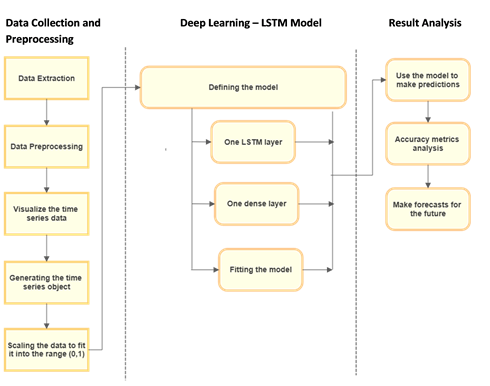


Figure 2.0.1: The Architecture of ARIMA Model

More sophisticated models, such as decision trees and ensemble models (e.g., Random Forests and Gradient Boosting Machines), have become more popular as they are capable of identifying intricate interactions among numerous variables. These models provide precision by capturing complex patterns within the data that conventional statistical techniques are unable to. Neural networks and deep learning have significantly transformed sales forecasting, in particular through the use of architectures such as Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks. These kinds of models achieve great results on sequential data and are hence highly appropriate for time series forecasting. LSTMs specifically have the ability to learn long term dependencies in sales patterns, adjusting for changing patterns over time.

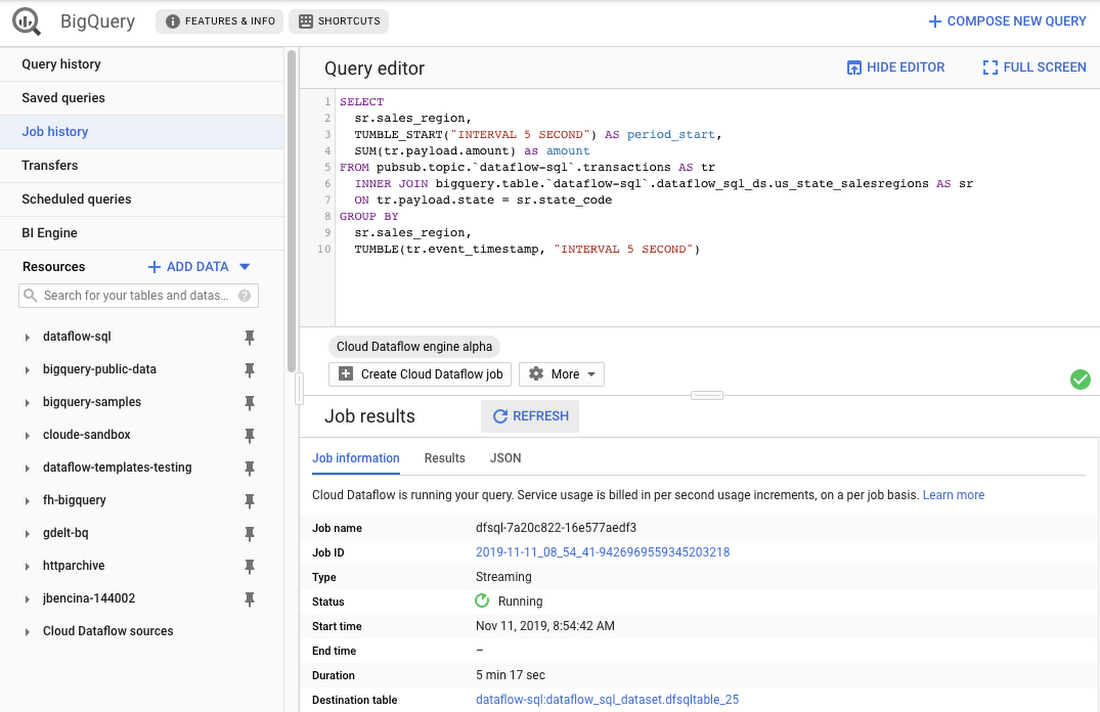


The Architecture of LSTM Model1

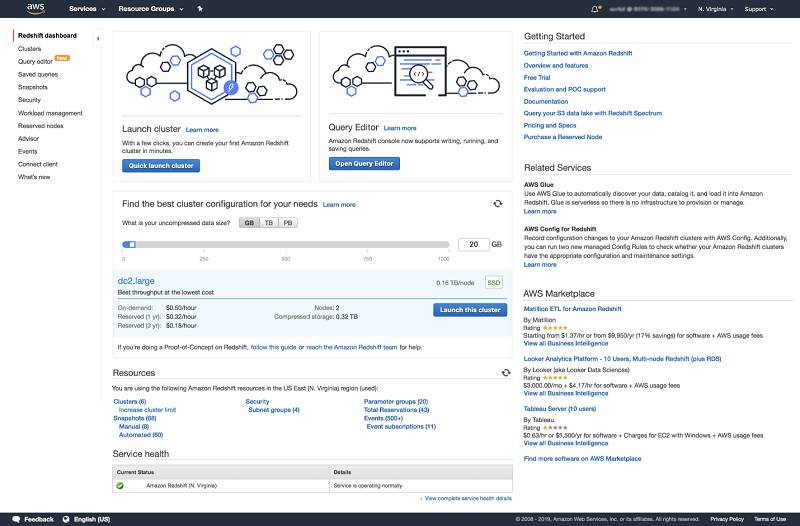
**Figure 2.2:** The Architecture of LSTM Model

## **2.3.2 Big Data Technologies**

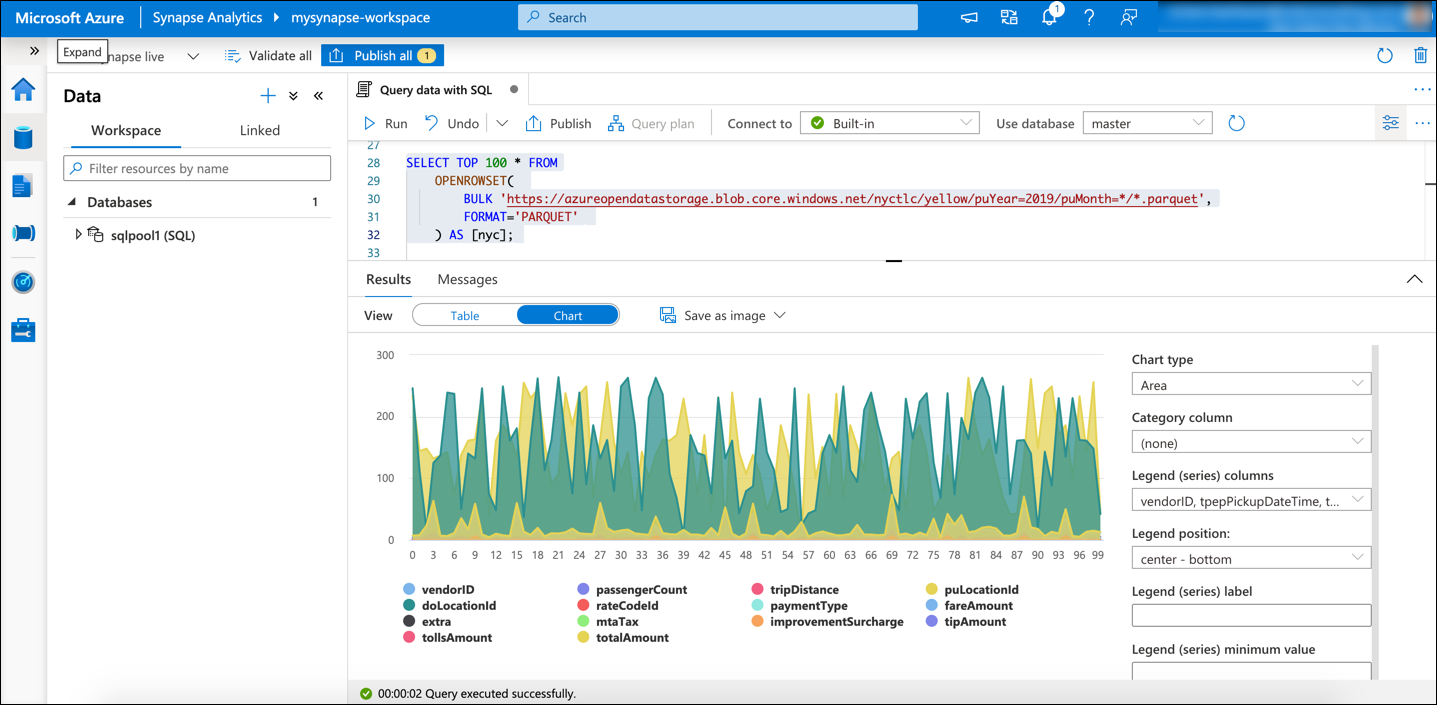
Companies handling high amounts of transactional data tend to use these deep learning models for highly accurate forecasting. Still, even though they work well, deep learning models consume a lot of computational power and huge datasets for training, which makes them resource intensive and at times challenging to interpret. In addition to machine learning algorithms, big data technology has significantly impacted predictive analytics. Apache Hadoop and Apache Spark are used by many firms to process very large datasets. Hadoop's distributed computing enables fast storage and calculation of huge sales data, and Spark offers features for real-time analytics, thus enabling quicker processing of data along with model training. Cloud-based platforms like Google BigQuery, Amazon Redshift, and Microsoft Azure Synapse Analytics provide scalable solutions for managing large data sets, hence diminishing the necessity for on-premise infrastructure.



**Figure 2.3:** An example of BigQuery

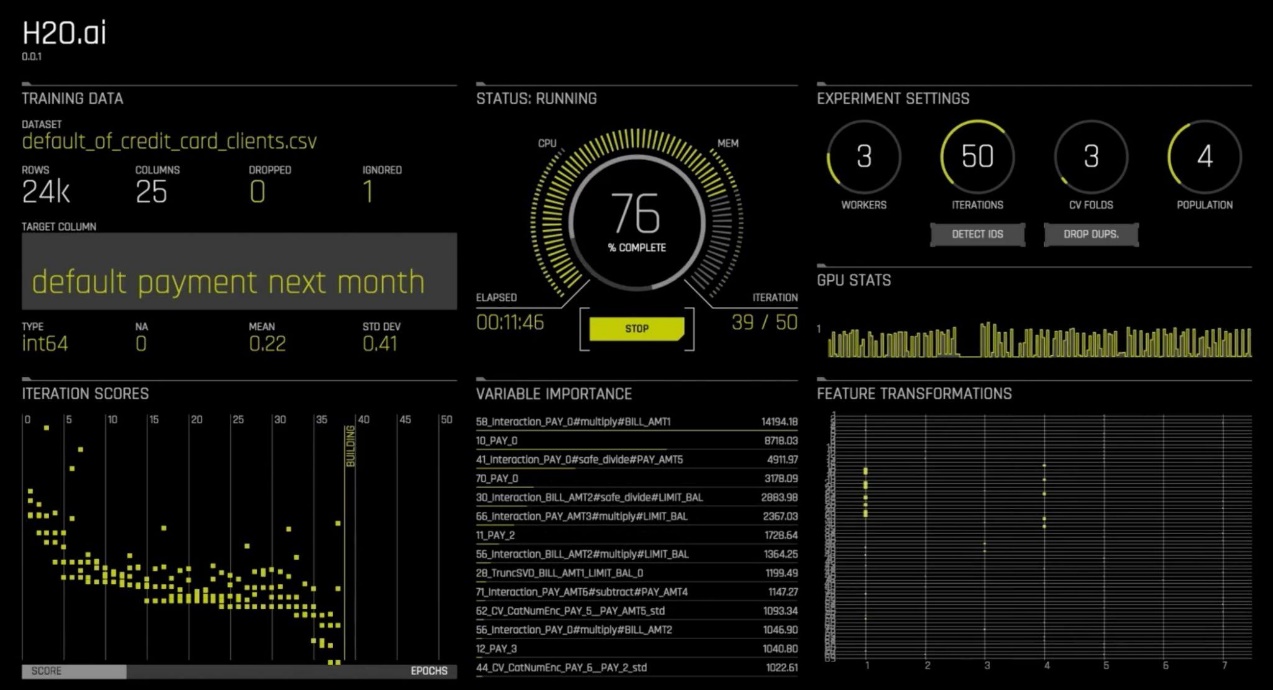


**Figure 2.4:** An example of Amazon Redshift

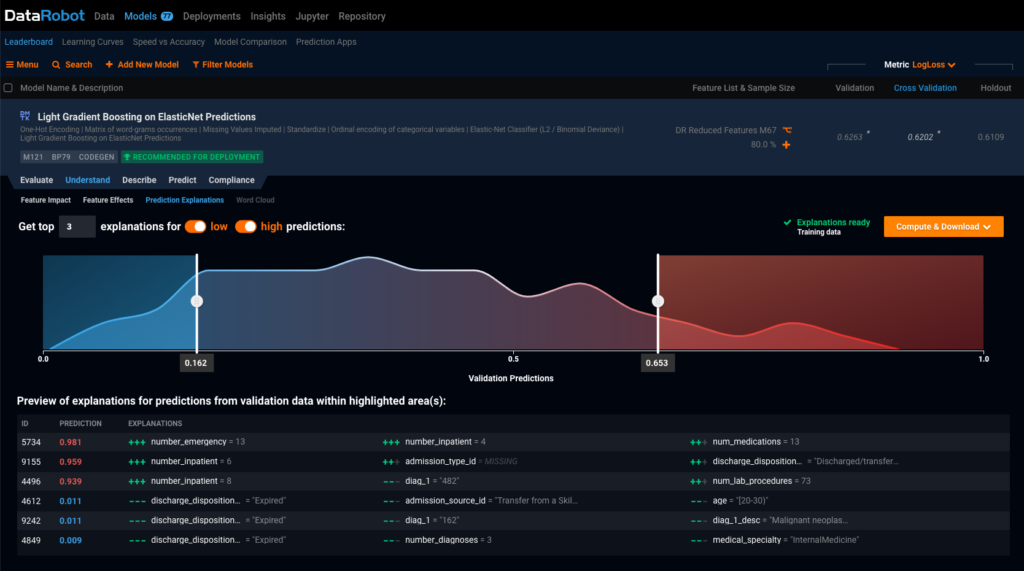


**Figure 2.5:** An example of Microsoft Azure Synapse Analytics

Moreover, such cloud solutions provide seamless integration with machine learning frameworks, allowing companies to deploy models in real-time environments. Another key development in predictive analytics is the application of automated machine learning (AutoML) platforms, including Google AutoML, H2O.ai, and DataRobot. These platforms make the process of building models easier by automating feature selection, hyperparameter optimization, and model validation, thereby enabling non-technical users to deploy advanced predictive analytics solutions. AutoML tools are especially valuable for companies with low data science capabilities because they assist in creating precise forecasting models with little manual effort. Natural Language Processing (NLP) has been added to predictive analytics for the analysis of unstructured data sources such as customer reviews, social media conversations, and news reports. Sentiment analysis, which is provided by NLP capabilities, can help companies in evaluating public opinion as well as market trends, thus providing additional context for models that predict sales. For example, a sudden increase in positive reviews of a product on social media can be a sign of higher future demand, while negative reviews may imply poor sales. Tools such as Google's BERT, OpenAI's GPT, and spaCy enable the inclusion of text data in predictive analytics solutions for businesses, thereby enhancing the quality of their predictions.

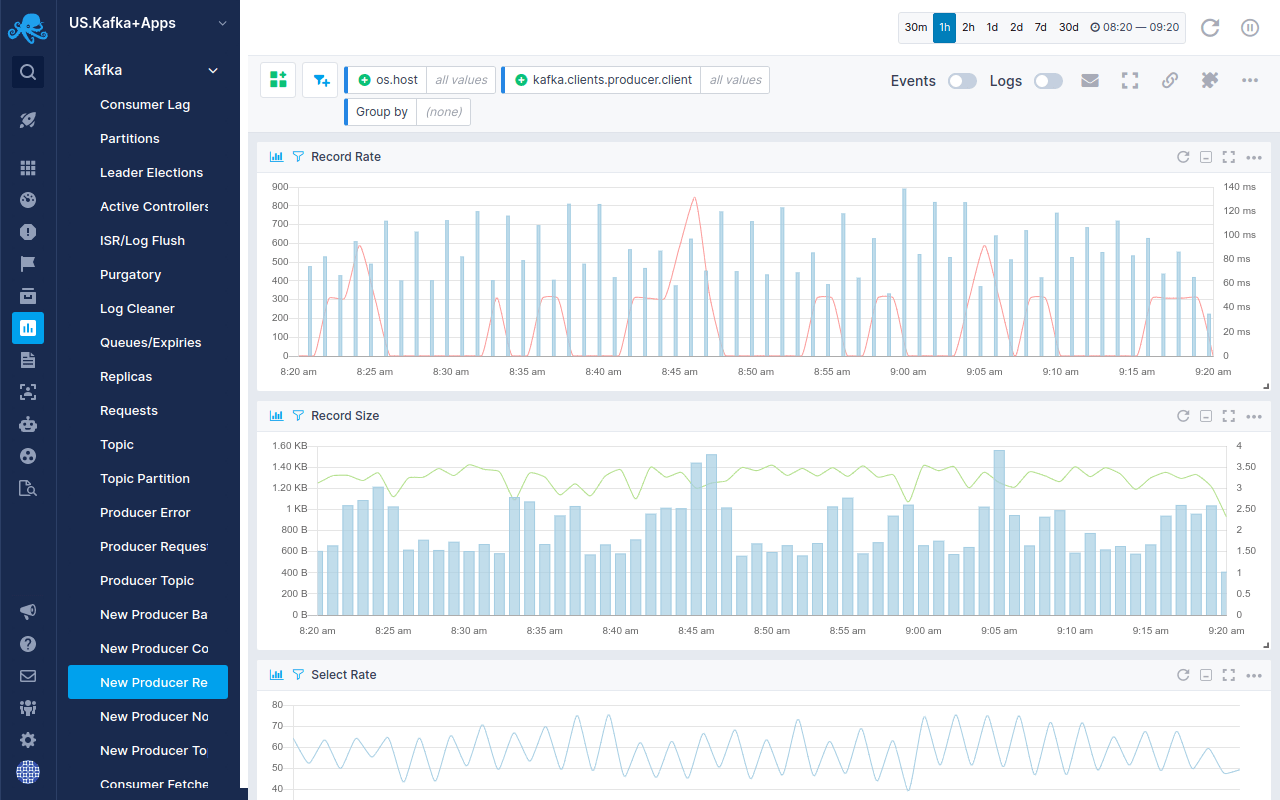


**Figure 2.6:** An example of H20.ai

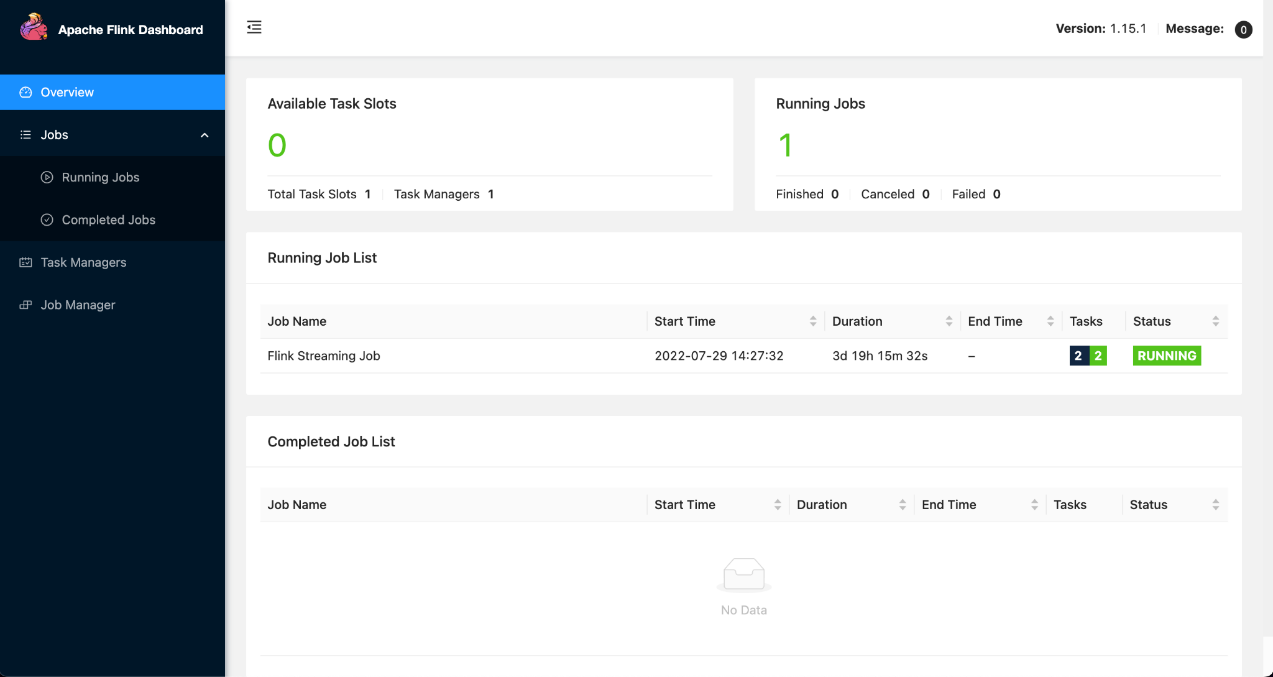


**Figure 2.7:** An example of DataRobot

These platforms deliver pre-built ML models, data preprocessing utilities, and scalable deployment environments that enable businesses to incorporate AI powered sales forecasting in their operations seamlessly. By utilizing these services, companies can possibly simplify the process of model building and deployment while, simultaneously, benefit from the computing power offered by cloud environments. Furthermore, real-time analytics and edge computing have opened up new avenues for adaptive sales forecasting. Companies that function in rapidly changing industries, i.e., E-commerce and financial markets, need real-time information to facilitate timely decision-making. Technologies like Apache Kafka, Apache Flink, and Google Dataflow enable streaming data analysis, allowing organizations to update their sales predictions as new data flows in.

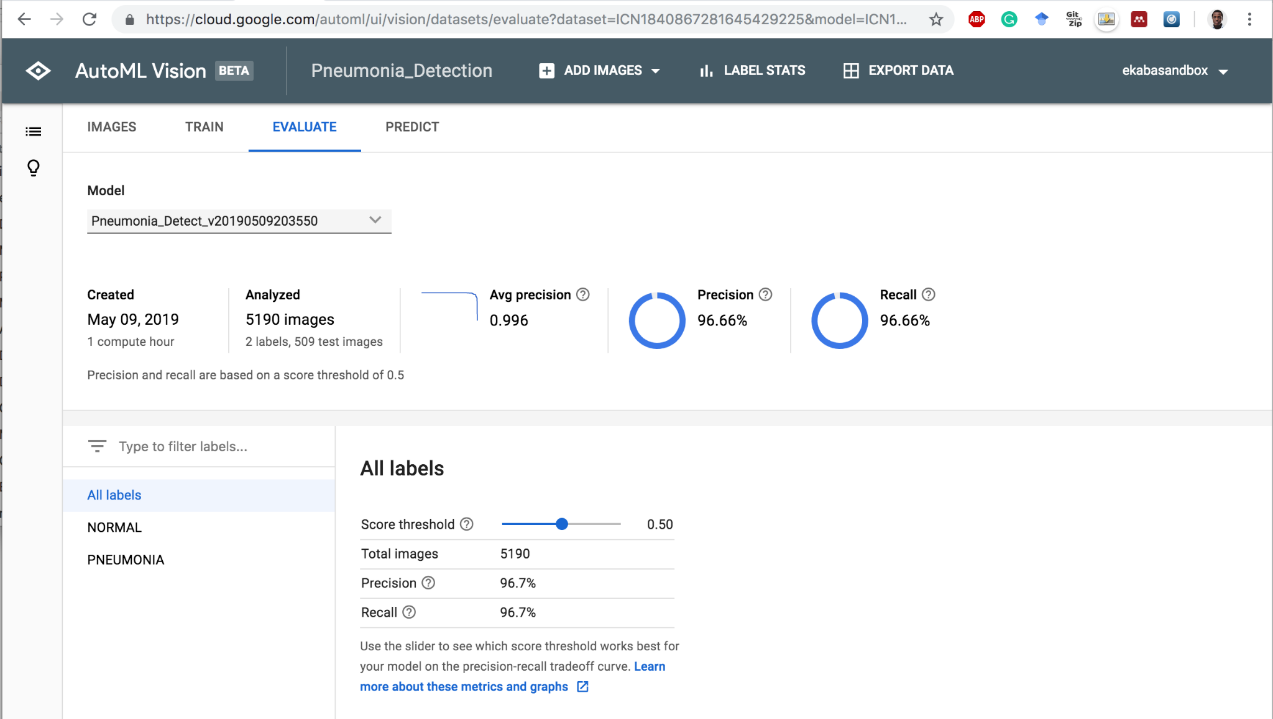


**Figure 2.8:** An example of Apache Kafka



**Figure 2.9:** An example of Apache Flink

Edge computing further enhances real-time capabilities by processing data closer to the source, reducing latency and improving the responsiveness of predictive models. Cumulatively, the advancement in predictive analytics to forecast sales and profit has been made possible through advancements in statistical modeling, machine learning, deep learning, big data processing, AutoML, NLP, cloud computing, and real-time analytics. While each of these technologies has specific benefits, organizations use a combination of multiple techniques to achieve highest forecasting accuracy. As machine learning continues to evolve, upcoming developments in artificial intelligence and data science are likely to enhance predictive analytics, thereby rendering it an indispensable tool for companies wanting to boost their bottom line.



**Figure 2.10:** An example of Google AutoML

## **2.4. RELATED WORKS**

The use of machine learning methods for sales and profit prediction has attracted considerable interest over the last few years. Various researchers have attempted different models and methodologies in an effort to improve the accuracy and effectiveness of making such projections, with the ultimate goal of offering organizations valuable information for strategic planning. This section considers some of the most important studies in this field, including their methodology, results, strengths, limitations, and outstanding gaps that have not been filled.

## **2.4.1 Machine Learning Algorithms for Sales Forecasting**

Research was carried out to investigate the application of machine learning algorithms in sales forecasting based on historical sales data and some of the most popular models such as Linear Regression, Decision Trees, and Support Vector Machines. The findings indicated that machine learning models are more precise compared to conventional statistical approaches. The research, however, noted limitations when handling big data and the need for appropriate feature engineering in order to achieve maximum model efficacy. In addition, the study highlighted a lack of consideration of real-time data processing and the integration of exogenous variables that impact sales (Sharma & Verma, 2021).

## **2.4.2 Comparison of Machine Learning and Deep Learning Models**

A further study compared two linear models, three machine learning models, and two deep learning models for sales forecasting specifically. The authors concluded that while machine learning and deep learning methods did not significantly surpass linear models in accuracy, incorporating calendar and price data improved predictive power. This study offered an extensive comparison of different models but highlighted the issues concerning model complexity and the requirement for large-scale data pre-processing. A significant limitation identified was that not enough investigation of ensemble methods was done, which can boost predictive performance (Liang et al., 2022).

## **2.4.3 Hybrid Models for Sales Forecasting**

Additionally, another study investigated sales data and prediction approaches, comparing different machine learning methodologies for sales forecasting applications. The results highlighted the need to choose suitable predictive models from performance assessments. The study recognized that although particular models excel in certain circumstances, there is no universal model that would work for every dataset. The study stressed the need for hybrid models incorporating varied algorithms to enhance accuracy. Yet it also recognized complexities in developing these and computation power needed. The identified gap involved the requirement for solutions that are scalable and able to handle heterogeneous data from various sectors (Ahmed et al., 2020).

## **2.4.4 Time Series Forecasting and Ensemble Approaches**

A different study addressed time series forecasting in the area of sales predictive analytics. The researchers proposed multiple approaches, referencing the phenomenon of machine learning generalization, which is useful in cases with limited historical data. They recommended the application of a stacking approach to constructing regression ensembles in a bid to promote predictive power. The strength of the study was marked by its orientation to circumstances dominated by sparse data; however, it acknowledged the limitation of potential overfitting in ensemble methodologies. The researched gap identified has to do with the need for methods that balance model complexity and simplicity, ensuring predictions are precise as well as understandable to key stakeholders (Kumar et al., 2021).

## **2.4.5 Advanced Machine Learning Techniques in Retail Sales Forecasting**

A study carried out recently explored the application of sophisticated machine learning algorithms such as Random Forest, Gradient Boosting, Support Vector Regression, and XGBoost in forecasting sales in retail firms. The outcome indicated that the application of machine learning techniques has the potential to deal with the intricacy of the sales data more effectively, resulting in enhanced predictive performance. The research, however, also indicated concerns regarding the model's interpretability along with computational expenses of building elaborate models. The noted inconsistency regards the integration of expert domain knowledge into machine learning models to enhance their applicability and usefulness in real-world scenarios (Fernandez & Gupta, 2023).

**Summary of Technologies Used and Limitations**

**Table 2.1:** Technologies Used and Limitations

|  |  |  |
| --- | --- | --- |
| **Study** | **Technologies Used** | **Limitations** |
| Sales Prediction Using Machine Learning (Bajaj et al., 2020) | Linear Regression, Decision Trees, Support Vector Machines | Challenges in handling large datasets; requires extensive feature engineering. |
| Sales Prediction Based on Machine Learning (Hitesh S M et al., 2024) | Linear Models, Machine Learning Models, Deep Learning Models | ML and deep learning models did not significantly outperform linear models; complex models require extensive preprocessing. |
| Intelligent Sales Prediction Using Machine Learning Techniques (Amrutkar & Mahadik, 2022) | Hybrid Models (Combination of ML Algorithms) | Development of hybrid models is complex and resource-intensive. |
| Machine-Learning Models for Sales Time Series Forecasting (Kumar et al., 2024) | Ensemble Learning (Stacking Approach) | Potential overfitting in ensemble models; balancing complexity with interpretability. |
| Enhancing Retail Sales Forecasting with Optimized Machine Learning Models (Ganguly & Mukherjee, n.d.) | Random Forest, Gradient Boosting, Support Vector Regression, XGBoost | Challenges in model interpretability; high computational demands. |

**Conclusion**

The reviewed studies collectively underscore the potential of machine learning techniques in enhancing sales and profit forecasting accuracy. While various models have been explored, each with its strengths and weaknesses, common challenges persist, including handling large and complex datasets, ensuring model interpretability, and integrating external factors into predictions. Addressing these gaps necessitates the development of more sophisticated, scalable, and interpretable models that can be applied across diverse industries and data scenarios.

## **2.5.** **GAP ANALYSIS**

Despite notable advancement in developing machine learning-based predictive analytics in predicting sales and profit, there exists some challenges as well as gaps. Gaps identification guides further research and development, hence guaranteeing predictive systems enhance their responsiveness, accuracy, and usability in business environments.

## **2.5.1 Gaps in Existing Systems.**

Limited Integration of External Factors: Many of the predictive models in use today focus overwhelmingly on historical sales data without considering the incorporation of external market determinants like economic conditions, consumer sentiment, and competitor pricing policies. The lack of these exogenous variables is a major flaw in current frameworks since they play a crucial role in driving demand as well as profitability (Van Calster et al., 2020)

Model Interpretability and Transparency: Machine learning models are very accurate, but many state-of-the-art models—especially deep learning methods—work in the way of "black boxes." That is, it is hard for organizations to comprehend how predictions are rendered, which lowers confidence and inhibits their uptake in decision-making (Ji et al., 2019)).

Real-Time Forecasting and Adaptability: Some predictive analytics software offers accurate forecasts but does not quickly adjust to novel events in the market. Organizations need models with the ability to refresh forecasts using real-time information, which works to reduce latency and improve the way businesses respond to unexpected situations (Navya et al., 2024)

High Computational Costs: The majority of advanced machine learning models are computationally expensive to train and implement. SMEs might not be able to utilize such models as they have limited budgets and hardware, thus requiring cost-effective solutions (S, n.d.).

Accessibility and Ease of Use: The majority of predictive systems demand data science and machine learning expertise, thus being hard to use for non-technical business users. We need easier-to-use and understand solutions so that predictive insights can be harnessed by business managers without lengthy training (Aguilar-Moreno et al., 2024)

**Table 2.2:** Technologies Used and Limitations

|  |  |  |
| --- | --- | --- |
| **Feature** | **Traditional Statistical Models** | **Machine Learning Models** |
| **Data Processing** | Primarily relies on historical sales trends | Incorporates complex patterns and multiple data sources |
| **Adaptability** | Struggles with nonlinear relationships and sudden market changes | Limited adaptability; does not account for dynamic market changes |
| **Interpretability** | Fully explainable with clear formulas | Often difficult to interpret, especially in deep learning models |
| **Computational Efficiency** | Low computational cost | High computational demand, especially for large datasets |
| **Scalability** | Designed for specific industries | Can generalize but often requires extensive retraining |
| **User-Friendliness** | Simple statistical methods easy to understand | Requires specialized knowledge to implement and fine-tune |

## **2.6. CHAPTER SUMMARY**

This chapter offered a comprehensive discussion of the foundational aspects that are associated with the application of predictive analytics in sales and profit forecasting with the help of machine learning algorithms. It started with the introduction section, which outlined the overall focus of the chapter on the theoretical and technological elements of sales forecasting. The system overview then explained the application of predictive analytics in corporate settings and how crucial machine learning is to enhance decision-making processes and optimize revenue generation.

This was followed by an extensive overview of used technologies that discussed the different machine learning methods being employed for sales forecasting presently. The section discussed regression models, decision trees, neural networks, and ensemble learning, detailing the particular strengths and weaknesses of each. It further discussed critically how companies use these technologies to maximize their margins through data-driven decision-making.

The related works section provided a comprehensive review of current research, pointing out various papers that have attempted to predict sales and profit using machine learning. From the reading of these papers, it was evident that the majority of models offer satisfactory accuracy but do not incorporate significant aspects such as the inclusion of real time data, accounting for external variables, and scalability across industries. The comparison table at the end of this section recapitulated the strengths and weaknesses of various methodologies, thus providing an overall understanding of their effectiveness. Then, the gap analysis identified key gaps in current predictive models.

It highlighted serious limitations, including the absence of incorporation of socioeconomic and behaviour variables, no analysis of live data, and interpretability issues of the models. The chapter also presented the Proposed System, which aims to fill these gaps by exploiting ensemble machine learning techniques, consolidating heterogeneous data sources, and enhancing the predictive capability. The comparison table set the stage for an improved model by highlighting the gaps in existing systems and illustrating how the proposed method aims to fill these. Overall, this chapter laid a strong groundwork for the subsequent sections of the thesis. The paper provided a comprehensive literature review, critiqued the existing methodologies, and identified the weaknesses that the proposed predictive analytics framework seeks to address.

The findings from this analysis will guide the next chapter, which will focus on the research methodology, including data collection, model selection, and evaluation techniques. By addressing the challenges outlined in this chapter, the proposed framework hopes to provide a more comprehensive and accurate model for predicting sales and profits, ultimately contributing to improving the strategic planning and financial forecasting activities of businesses.

**CHAPTER 3**

**METHODOLOGY**

## **3.1 INTRODUCTION**

This chapter describes the methodology employed in the creation of the predictive analytics system for forecasting sales and profits. The approach guarantees that the system is created, implemented, and verified through formalized procedures to increase accuracy, scalability, and efficiency. The methodology entails a structured process starting with requirement gathering and analysis, system design, requirement specification, and system implementation approaches.

The initial section of the chapter outlines the procedure for gathering user requirements, establishing the inherent characteristics and limitations that dictate the architecture of the system. Subsequently, there is an elaborate examination of the system design based on the elements involved in data acquisition, processing, model training, and prediction in real time. Included is an elaborate explanation of the requirements with hardware and software dependencies necessary to ensure the system continues to run uninterrupted.

By implementing this well-organized approach, the system hopes to solve the deficiencies in sales forecasting based on machine learning algorithms and data-driven analysis revealed in the foregoing. The following sections chronicle the steps adhered to so that the system is both in tandem with business requirements and viable technically.

## **3.2 USER REQUIREMENT GATHERING AND ANALYSIS**

The success of a sales and profit forecasting system largely hinges on a properly defined requirement gathering process. The phase guarantees the alignment of the system with organizational goals and user expectations as well as solves problems that are intrinsic to predictive analytics. The system will improve forecasting accuracy through the use of sophisticated machine learning algorithms and the capability of handling data in real time. Among the primary goals is to give businesses the ability to forecast sales and profit. Requirement’s analysis is the phase where it is necessary to comprehend data flow, assess computation requirements, and determine the particular features that would render the system valuable to end-users like financial analysts, sales managers, and business strategists.

## **3.2.1 Preliminary Investigation**

The preliminary analysis evaluates the deficiencies inherent in current sales and profit forecasting techniques in light of the potential enhancement in accuracy offered by the suggested system. Conventional forecasting techniques, including simple statistical models, may not be able to successfully model intricate patterns present within sales data, thus making erroneous projections. In addition, a majority of organizations struggle with the incorporation of real-time data, which hampers their capability to respond to market changes.

To address these issues, this phase examines the data sources involved, stakeholder needs, and technological viability. This involves looking at previous sales patterns, market dynamics, and extrinsic drivers of revenue. The question also investigates the best machine learning strategies, cloud integration, and automation capabilities so that the system delivers precise, scalable, and effective forecasts.

## **3.2.2 User Requirement Analysis**

User requirement analysis takes into consideration what the stakeholders and companies need from the system. It is important to have accurate sales and profit predictions in order to make the correct decisions, and therefore the system must support multiple users, including financial analysts, sales managers, and company executives. These users require accurate predictions, real-time analysis, and easy-to-use interfaces to interpret the output.

This phase is about getting feedback on current problems with forecasting, including incorrect predictions, lack of automation, and challenges with processing big data. The system aims to solve these problems with the use of machine learning algorithms, real-time data processing, and cloud access. Through mapping system functionality to user needs, the predictive model enables better decision-making, resource allocation optimization, and better financial planning.

## **3.2.3 Data Gathering**

The success of the predictive analytics system depends a lot on the quality and availability of data. This step involves getting important sales and financial data from multiple sources to create a robust dataset for training and testing machine learning models. Past sales data, customer transactional data, market trends, and outside economic indicators are all compiled to make predictions as accurate as possible. Real-time data from e-commerce shopping websites, point-of-sale systems, and online market trends also help improve the flexibility of predictions

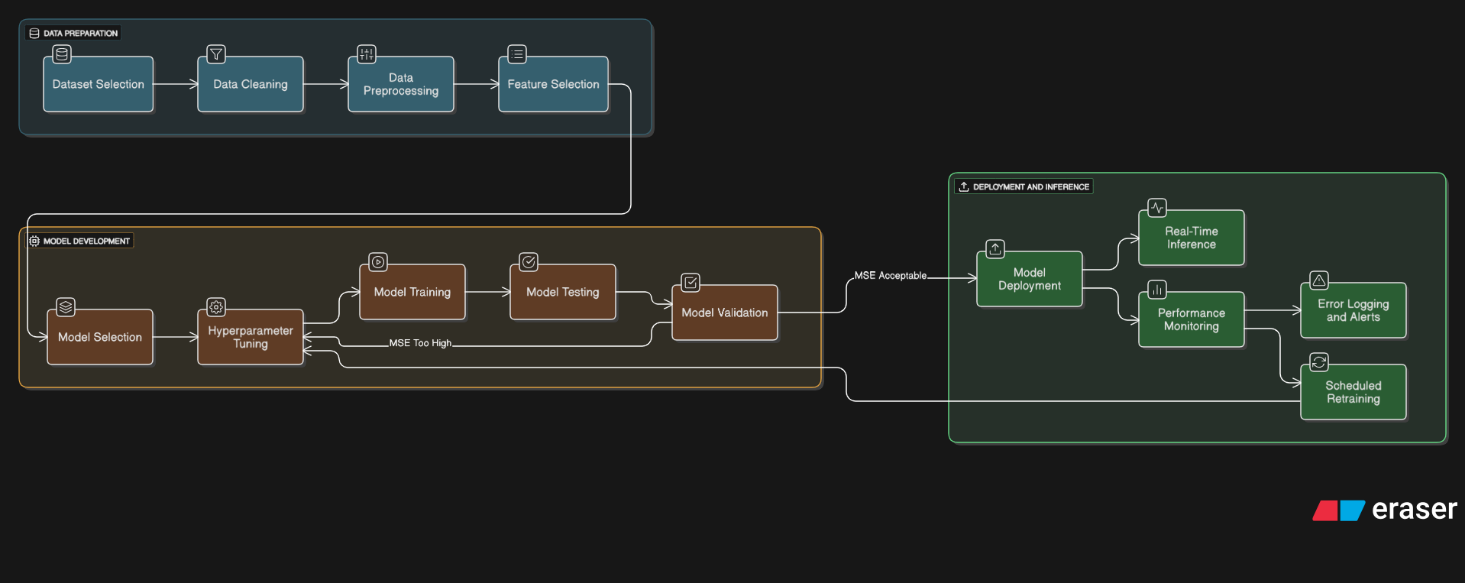
To ensure accuracy in our data, we use pre-processing techniques that involve data cleaning, normalization, and managing missing values. The system supports multiple databases, cloud storage, and APIs for easy gathering of data. Utilizing both structured and unstructured sources of data, the system increases the ability to predict results, hence helping businesses in confidently making data-based decisions.

## **3.3 SYSTEM DESIGN**

The system is intended to deliver precise, scalable, profit and sales forecasting. It comprises data acquisition, pre-processing, machine learning model training, prediction generation, and visualization modules. With the integration of historical data, the system provides end-to-end data analysis for enhanced forecasting.

Machine learning methods like linear regression, random forest, Kneighbors and Gradient boosting execute the data to identify patterns and make forecasts. Interactive dashboards allow users to view forecasts, analyse trends, and adjust parameters. Security features like data encryption and authentication secure data.

This architecture presents a robust, efficient, and user-friendly platform to enhance business decision-making through sound predictive wisdom.



**Figure 3.1:** The Block Diagram of the System

## **3.3.1 Data Acquisition**

The initial component of the system design entails data acquisition, incorporating the gathering of a rich and varied dataset of sales and financial information. The dataset contains past sales transactions, aimed at enhancing forecast precision. It is essential that the data span time periods, seasonal cycles, and exogenous factors that affect the performance of sales to guarantee robustness in models.

Sales data is gathered from internal databases, point-of-sale terminals, and external reports such as economic indicators and competitors' prices. The data is extracted at a set interval by APIs and script programs so that it remains up to date for analysis. The raw data is stored in a structured format after collection before it is processed for model training and prediction.

## **3.3.2 Data Pre-processing**

The second part of the system design relates to data pre-processing, which ensures that the obtained sales and financial data is free from impurities, structured, and suitable for machine learning analysis. Raw data often contains irregularities, missing values, and irrelevant features that can affect the accuracy of forecasts. To improve model effectiveness, the dataset requires purification using a series of pre-processing techniques.

This process includes handling missing values through imputation, removal of duplicate records, and normalization of numeric data for consistency. Categorical fields such as customer demographics are transformed into machine-readable format. Feature selection is also used to determine the most relevant attributes influencing sales and profit trends. The data is in structured form after pre-processing and ready for use in model training and forecasting tasks.

## **3.3.3 Model Training**

The third component of the system architecture is training of the model, whereby pre-processed data is employed to create a predictive model aimed at forecasting sales and profits. The system employs machine learning algorithms to discern underlying patterns and trends that exist in the dataset and consequently make correct and reliable forecasts.

The process of training includes dividing the dataset into training and validation sets to test the performance of the model. Different models, including regression techniques, decision tree models, and deep learning models, are compared to identify the best model. Hyperparameter optimization is conducted to enhance the accuracy of the model, with performance metrics such as Mean Absolute Error (MAE) and precision accuracy used to quantify the outcomes. After completing the training, the best-performing model is chosen for deployment and subsequent forecasting.

## **3.3.4 Real-Time Prediction**

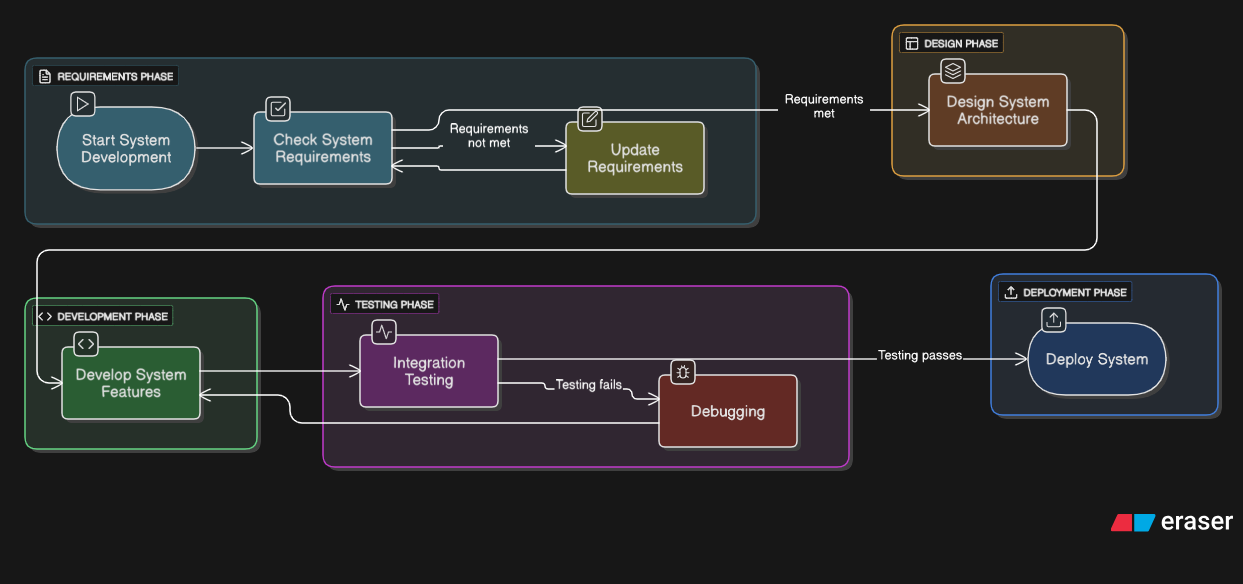
The real time prediction module is specifically developed to give companies real time predictions of sales and profits based on the examination of incoming data streams. This module guarantees the system's capability to respond to evolving market circumstances, thereby facilitating prompt and well-informed decision-making. Point of sale terminal and online shop data is endlessly gathered, pre-processed, and inputted into the trained machine learning model for forecasting generation.

For improved user experience, the system incorporates interactive dashboards for providing forecasts in a user-friendly format. Dashboards enable users to study trends, compare forecasts with actuality, and tweak parameters for scenario analysis needs. Apart from that, alert systems alert the user of serious deviations from predicted trends, hence enabling real time response to changes in the market.

## **3.4 REQUIREMENT ANALYSIS**

Requirement’s analysis is a critical step of system development during which one discovers and documents stakeholders' requirements for the delivery of the system to achieve business goals as well as user expectations. Collecting the system's functional and non-functional requirements such as accuracy, scalability, and user interface design is all accomplished in this step. The analysis makes certain that the system solves the problems found in existing sales and profit forecasting systems, including inaccurate projections and lack of automation. Through a careful examination of these requirements, the system is created to offer a strong, effective, and user-friendly solution that improves decision making processes and business productivity for firms.

## **Flowchart Design for the System**



**Figure 3.2:** The Flowchart Diagram of the System

## **3.5 USER REQUIREMENT SPECIFICATION**

This phase deals with formally documenting the requirements and expectations of the end-users of the system, in order to ensure the final product is aligned with business goals and user expectations. The specification is a map, delineating functional and non-functional requirements, such as accuracy, scalability, real-time processing, and user interface. It ensures the system transcends the weaknesses of traditional forecasting methods, providing a robust and easy to use solution.

## **3.5.1 New Proposed System**

The new system we are proposing will transform profit and sales forecasting solutions by utilizing state of the art machine learning algorithms coupled with a thorough analysis of historical data. Unlike conventional methods, which do not always heed subtle trends and patterns, this system applies techniques such as regression models, random forest, Kneighbors and gradient boosting to study previous sales data. They have been specifically selected because they are proficient in addressing non-linear dependencies and seasonality, thereby allowing for very precise forecasts.

The system is based on a historical data set constructed from past sales transactions,. By linking these disparate sources of data, the system can discern underlying patterns that are not typically caught by more traditional forecasting techniques. Not only does this method introduce accuracy to forecasting, but it gives a more balanced perspective of forces that drive sales and profitability.

For enhancing the usability of the system, interactive dashboards are provided, which allow users to visualize projections, investigate trends, and modify parameters for scenario planning. The dashboards are purposely made user friendly, which allows financial analysts, sales managers, and executives to interpret results easily and make data-informed decisions.

## **3.5.2 Solution Strategy**

Our solution strategy for developing the sales and profit forecasting system is divided into a series of clearly defined, executable steps. The process guarantees the system is developed, designed, and implemented correctly while meeting user requirements and technical specifications. The most critical steps in the process are as follows:

**1. Data Collection and Pre-processing:** Historical sales data is gathered from a number of sources including previous transactions, consumer trends, and market patterns. The data is pre-processed to assure quality and consistency. This includes data cleaning, missing value handling, and normalization of numerical features. The pre-processed data is then formatted in a structured way for model training use.

**2. Model Training and Development:** With the pre-processed data, we develop machine learning models that include regression, decision trees, and deep learning models. The models are trained to recognize patterns and trends in the historical data. Cross-validation and hyperparameter tuning are employed to enhance model performance and accuracy.

**3. Model Evaluation and Assessment:** After the process of training is completed, models are validated on an independent set of data in order to examine their performance. Mean Absolute Error (MAE) and precision accuracy measures are used in order to capture precision.

**4. System Deployment and Integration:** After verifying the models, we go ahead and implement them in the system and carry out the deployment of the solution. This involves setting up necessary infrastructure, such as servers and databases, and also ensuring user access to the system.

**5. Monitoring and Maintenance:** Post-deployment, we closely monitor the system's performance and do necessary fine tuning to maintain accuracy and reliability. Regular updates and maintenance are performed to align the system with evolving business needs and technological advancements.

By following this structured approach, we ensure the system is developed efficiently, meets user expectations, and delivers accurate and actionable sales and profit forecasts.

## **3.6 SYSTEM REQUIREMENT SPECIFICATION**

The System Requirement Specification (SRS) outlines the major technical and functional requirements for the successful operation of the system. This stage guarantees that the system is developed to address both user requirements and technical specifications, thereby offering an authoritative guide to development. This section addresses software dependencies, system interfaces, and performance standards, thereby guaranteeing that the system is robust, flexible, and secure.

## **3.6.1 System Interface**

The system's interface has been made user-friendly to ensure that users have a logical and smooth experience. The system has interactive dashboards used to present sales and profit projections in a visually impressive way. These dashboards enable users to navigate through trends, and create reports with ease. The system is user-centred, with different audiences like financial analysts and business executives being catered to, guaranteeing functionality and accessibility.

Its interface is responsive, making it highly effective across various devices and screen sizes. The interface gives users the power to customize their views. With a transparent and interactive way of deciphering forecasts, the system enables users to make data-driven decisions with confidence.

## **3.6.2 Software Interface**

The software interface plays a pivotal role in facilitating uninterrupted communication between the hardware components of the system and its sophisticated machine learning models, especially those that are used in sales and profit prediction. The interface is built to enhance the performance and operating efficiencies of the system, thus facilitating effective data processing, model training, and inference operations.

Basically, the software interface enables a smooth communication of information from diverse sources, including databases of past sales and repositories of market trends, to the pre-processing and analysis sub-units of the system. It formats and organizes the data in a manner that is suitable for use in training machine learning algorithms, including regression analysis and decision trees. Maximizing this process enables the interface to improve the capacity of the system to make effective and reliable projections.

With an emphasis on function and performance, the interface software facilitates the seamless interaction of the hardware and software components of the system, thereby yielding a solid and scalable solution for sales and profit prediction.

## 

## **3.7 CHAPTER SUMMARY**

This Chapter details the methodology for developing the sales and profit forecasting system. It begins with user requirement gathering and analysis, focusing on stakeholder needs and aligning the system with business goals. The system design phase covers data acquisition, pre-processing, model training, and prediction, leveraging historical data and machine learning for accurate forecasts.

The chapter also discusses requirement analysis, including flowcharts and feasibility studies, to validate the system’s design. The user and system requirement specifications provide a blueprint for functionality, interfaces, and hardware dependencies. Finally, the solution strategy outlines the iterative development process, ensuring the system is scalable, secure, and user-friendly.

## **CHAPTER 4**

**IMPLEMENTATION**

## **4.1 INTRODUCTION**

This chapter presents the implementation, testing, and results of the sales and profit prediction system developed using machine learning techniques. The primary goal was to create a reliable and efficient forecasting model using Python, which was then integrated into a user-friendly ASP.NET interface for real-time predictions. The system processes historical sales data, applies machine learning algorithms for training and prediction, and displays the results interactively through the web platform. This chapter also explains the integration between the Python backend and the ASP.NET frontend to ensure seamless data flow and usability. Implementation details are divided into software and hardware components, and performance evaluation is discussed based on key metrics to assess accuracy and reliability.

## **4.2 IMPLEMENTATION**

The following chapter gives an in-depth description of implementing the sales and profit forecasting system with the help of machine learning approaches. The chapter describes the technical process of developing the system, such as the hardware and software configuration, training the model, and evaluation. The implementation also targeted developing a series of multiple regression models for the purpose of forecasting sales and profit and integrating such models in an interactive web-based interface for visualization of the performance and analysis.

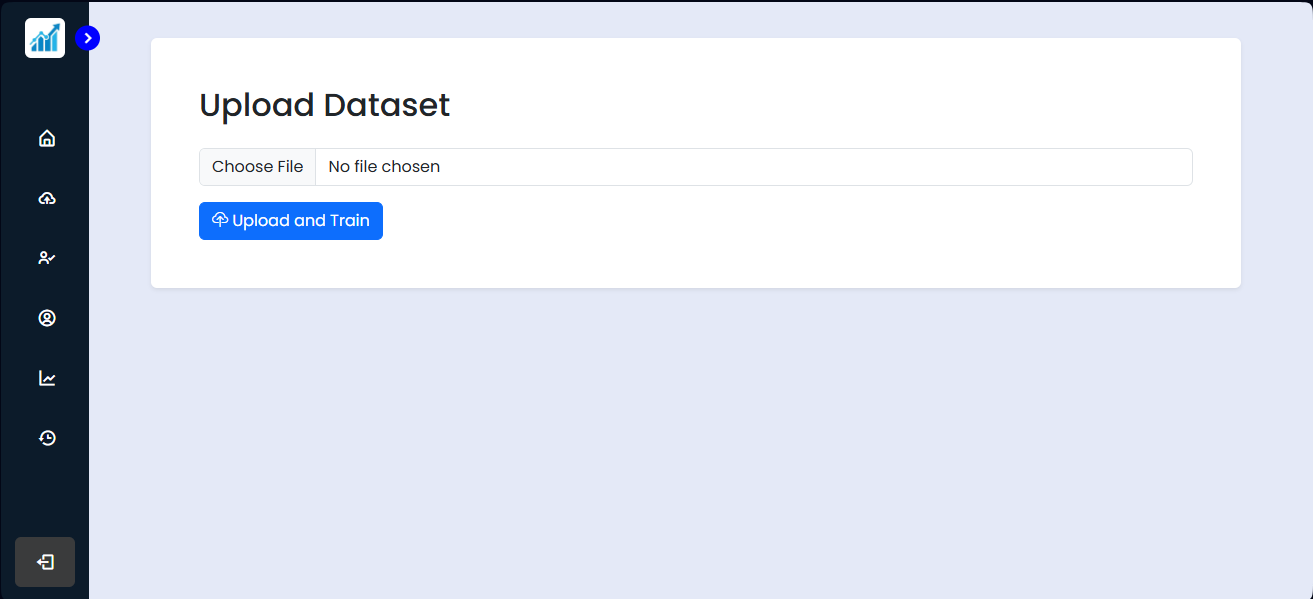
### **4.2.1 Software Implementation**

The implementation of the sales and profit prediction system was carried out using Python for model development and ASP.NET for the interface integration. The system was designed to train and evaluate four machine learning models—Linear Regression, Random Forest, Gradient Boosting, and K-Nearest Neighbors—on a structured dataset. These models were selected for their performance in regression tasks and their ability to capture linear and nonlinear relationships in sales data.

The training pipeline in Python begins with preprocessing steps, including handling missing values, encoding categorical variables, and normalizing numerical features. The cleaned dataset is then split into training and testing subsets to ensure fair evaluation. Each of the four models is trained independently on the same dataset and assessed using Mean Squared Error (MSE) and accuracy metrics for both sales and profit predictions.

Once training is completed, the performance of all models is displayed on the admin dashboard. This is done through the ASP.NET interface, which is connected to the Python backend via a FAST API. The interface displays the MSE and accuracy for each model, allowing for visual comparison of their results. The best-performing model is selected manually based on these evaluation metrics and can be used for generating predictions.

The codebase is modular, with separate components for data preprocessing, model training, evaluation, and API response formatting. This modularity enhances maintainability and allows for future expansion, such as adding new models or retraining with different datasets. A screenshot of the admin interface displaying model performance is shown in Figure 4.1.



***Figure 4.1:*** *Admin Interface Displaying How to Upload a Dataset*

#### ***4.2.1.1 Code Structure***

The predictive analytics system's code was structured as modular components for enhanced readability, scalability, and maintainability. There was a specific functionality for each module within the training pipeline so as to maintain a well-separated purpose and facilitate the process of debugging.

**Data Preprocessing Module**: The module carried out preliminary data cleanup, such as the removal of missing values, numerical column normalization, and categorical feature encoding. It prepared the dataset to make it consistent and model-ready.

**Training Model Module**: The following section had the individual training scripts for Random Forest, Linear Regression, Gradient Boosting, and K-Nearest Neighbors. All the models were trained separately using the same dataset and measured with Mean Squared Error (MSE) and accuracy measures for both profit and sales.

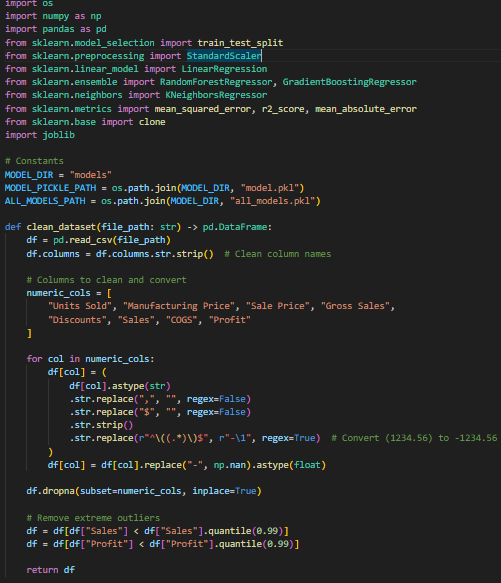
**Module for Model Evaluation**: Once trained, this module computed the performance measures and saved the outputs for presentation. It employed libraries like Scikit-learn and NumPy for the calculation of MSE, accuracy, and other statistical values.

**API Integration Module**: To make the model endpoints accessible, a light-weight Flask API was used. With this, the trained models are used to interact with the ASP.NET interface by sending and receiving information for the purpose of displaying real-time results.

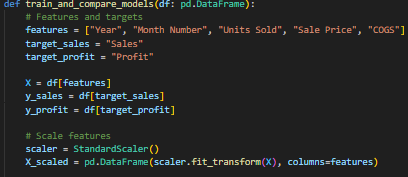
**Interface Connection (ASP.NET)**: The UI was made to call the API endpoints, return the results from the Python backend, and show the model performance measures in a well-organized structure. All the results from the different models were shown in visually differentiated cards for straightforward comparison.

This modular design allowed each component of the system to be easily upgraded or enhanced without any impact on the overall pipeline, so the system was both resilient and flexible for future updates.

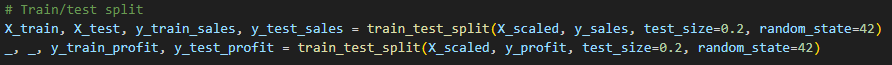
* **Pre-Processing**



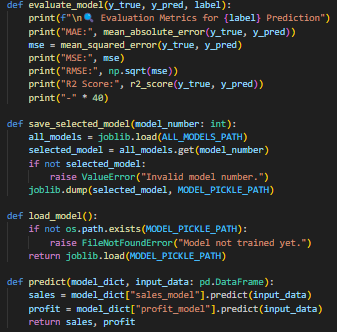
* **Model Building**



* **Training Code**



* **Evaluation Code**



### **4.2.2 Hardware Implementations**

The development and training of the sales and profit prediction system were conducted on a laptop equipped with a 13th Generation Intel(R) Core(TM) i7-1355U processor, featuring 10 cores and 12 logical threads with a base clock speed of 1.70 GHz. The system was supported by 16 GB of RAM, which provided sufficient memory for handling data-intensive operations such as model training, evaluation, and simultaneous execution of multiple scripts.

Storage was handled by a 512 GB Samsung MZVL4512HBLU-00BH1 solid-state drive (SSD), of which 340 GB was formatted for use. The SSD offered fast read/write speeds and reduced response time during data loading and preprocessing stages. The system also utilized Intel(R) Iris(R) Xe integrated graphics, which, although not a dedicated GPU, proved adequate for training the selected machine learning models without significant delays.

All development work, including Python scripting and model training, was executed on a Windows operating system using Visual Studio Code and Visual Studio. The ASP.NET web interface was hosted and tested locally on the same machine. This hardware configuration provided a reliable and efficient environment for completing the project within acceptable training and testing durations.

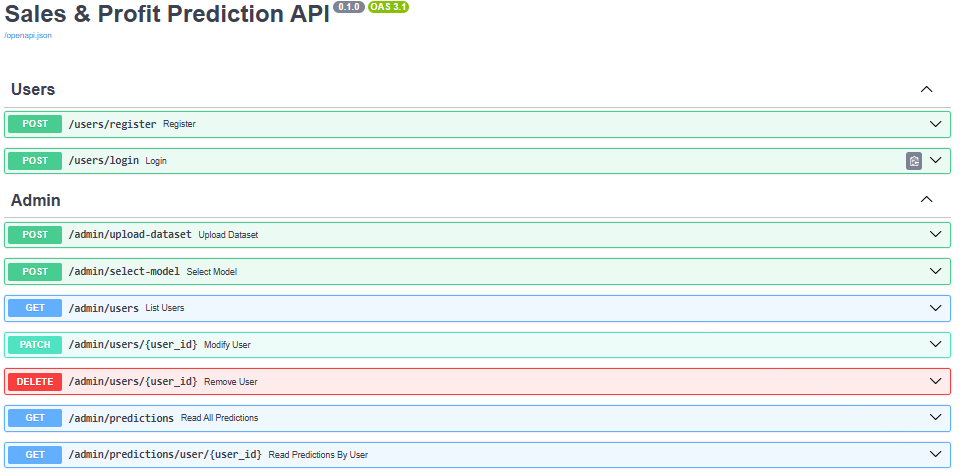
### **4.2.3 Integration of Software and Hardware**

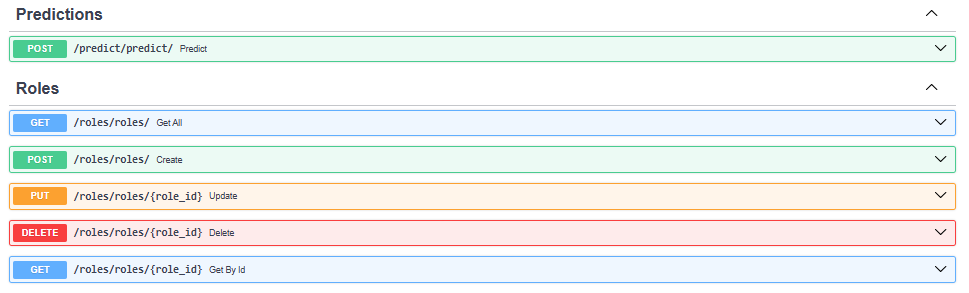
The integration between the machine learning model and the web interface was accomplished using FastAPI, a modern and high-performance web framework for building APIs with Python. After training the regression models, the prediction logic was deployed through FastAPI endpoints that enabled communication between the backend and the ASP.NET frontend.

When a request is made from the ASP.NET interface, user input is sent to the FastAPI backend in real-time. FastAPI then processes the request by passing the input data to the selected machine learning model, generating predictions for both sales and profit. The response is returned in JSON format and displayed immediately on the interface through visually structured cards showing performance metrics.

All components were hosted on the same local machine, leveraging the 13th Gen Intel Core i7 processor, 16 GB RAM, and SSD storage to ensure low-latency interaction between the API and the interface. FastAPI’s asynchronous capabilities provided rapid response times, making the overall system efficient and responsive.

This integration strategy preserved a clean architectural separation between the computational logic in Python and the user interaction layer in ASP.NET. It also ensured modularity, allowing for straightforward updates to the models or interface without disrupting the overall system flow.







## **4.3 TESTING AND EVALUATION**

The tests were designed to verify that every module of the system worked according to design and that the entire pipeline from preprocessing of data to output of predictions operated effectively and reliably. The process of tests was broken down into unit tests, integration tests, and system tests. All these tests were done to verify not just the back-end processes coded in Python but also front-end interaction using ASP.NET.

### **4.3.1 Evaluation Metrics**

To evaluate the performance of the trained machine learning models, two key regression metrics were used: Mean Squared Error (MSE) and Prediction Accuracy. These metrics were chosen based on their relevance in assessing continuous numerical outputs, such as sales and profit values.

* **Mean Squared Error (MSE):**  
  MSE measures the average squared difference between the actual and predicted values. It is calculated as:

where *yi*​​ is the actual value, ŷᵢ​ is the predicted value, and *n* is the number of predictions. A lower MSE indicates that the model predictions are closer to the actual results.

* **Prediction Accuracy (%):**  
  Although accuracy is typically used for classification, it was adapted in this project to reflect the percentage of predictions falling within an acceptable error margin. Accuracy was calculated based on how close the predicted values were to the actual values, within a defined threshold, and was expressed as a percentage.

These metrics were used consistently for both sales and profit outputs across all four models Linear Regression, Random Forest, Gradient Boosting, and K-Nearest Neighbors during the training and evaluation stages.

### **4.3.2 Overview of the Dataset**

The dataset used in this project was a structured sales dataset containing historical records related to transaction activity, product pricing, and cost information. It served as the foundation for training and evaluating the machine learning models developed to predict both sales and profit.

The dataset included the following key attributes:

* **Year** – indicating the year of transaction
* **Month** – identifying the month in which sales occurred
* **Units Sold** – number of product units sold
* **Sale Price** – unit price at which the item was sold
* **Cost of Goods Sold (COGS)** – the cost incurred to produce the item

The target variables were:

* **Sales** – calculated as Units Sold × Sale Price
* **Profit** – calculated as Sales − COGS

The dataset was formatted as a CSV (Comma-Separated Values) file and was uploaded through the admin interface for model training. All data fields were numeric and required minimal cleaning aside from handling missing or invalid entries. This ensured consistency in the model input format and reduced preprocessing complexity.

The data was split into training and testing sets in an 80:20 ratio. The training set was used to fit all four machine learning models, while the test set was reserved for evaluating generalization performance. This approach allowed fair and consistent comparison of the models under identical data conditions.

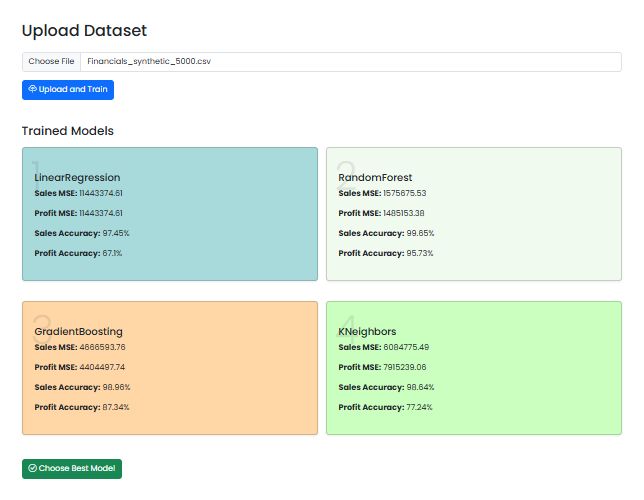
**Evaluation Process**

The evaluation process was structured to ensure consistent comparison across all four machine learning models. After loading and preprocessing the dataset, it was split into two subsets: 80% for training and 20% for testing. All models—Linear Regression, Random Forest, Gradient Boosting, and K-Nearest Neighbors—were trained on the same training set to maintain fairness in evaluation.

Once training was completed, each model was tested on the unseen portion of the data (the test set), and their performance was recorded using the predefined evaluation metrics: Mean Squared Error (MSE) and prediction accuracy. These metrics were computed separately for both sales and profit outputs, and the results were automatically displayed on the admin dashboard.

The system did not automatically select the best model; instead, the admin user manually reviewed the performance of all models and selected the one that offered the best balance between accuracy and error. This manual selection process was supported by a clear layout in the interface, which presented side-by-side comparison cards for each model’s performance.

This evaluation setup ensured that all models were tested under identical conditions, providing a reliable basis for selecting the most effective predictor for deployment.



## **4.4 PERFORMANCE EVALUATION**

This section presents the analysis of performance of machine learning models that have been deployed in the system. The analysis is based on the accuracy with which each of the models predicted sales and profits from historical input data. The performance was analysed using conventional regression metrics and conducted in two stages: training accuracy and validation accuracy. Models under consideration are Linear Regression, Random Forest, Gradient Boosting, and K-Nearest Neighbours.

### **4.4.1 Evaluation Metrics**

For the measurement of the models' performance, two primary metrics were utilized: Mean Squared Error (MSE) and Prediction Accuracy.

* Mean Squared Error (MSE) is a numerical measure that calculates the mean of the squared differences between predicted and observed values. A smaller MSE indicates better model performance and improved predictive quality.
* Prediction Accuracy: This measures how close the predicted values are to the actual values in percentage terms. Here, predictions that were within a predefined margin of error were deemed accurate.

These metrics were employed to contrast the sales and profit predictions across the whole span of the four models.

### **4.4.2 Training Accuracy**

For the training stage, every model was trained on the training subset of the data. The MSE and accuracy computed were noted for both sales and profit outputs. The ensemble models, Random Forest and Gradient Boosting, had persistently lower MSE and greater accuracy on the training data, which reflects a very good capacity to learn from past records. Linear Regression and K-Nearest Neighbors were also fairly good but indicated weakness in representing intricate relationships within the data.

### **4.4.3 Validation Accuracy**

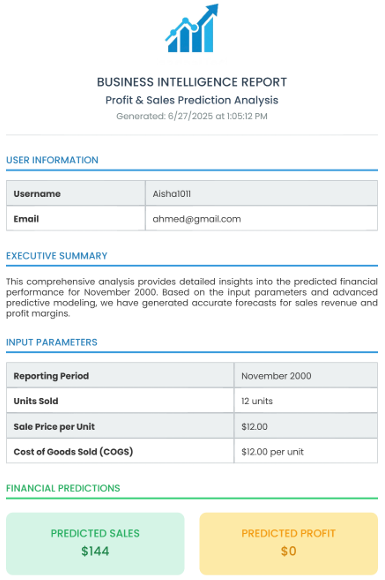
For checking generalization performance, the models that were trained were evaluated on the validation (test) set. During this phase, it was observed how well each model was able to predict unseen data. Random Forest and Gradient Boosting continued to perform well by maintaining high accuracy and stable MSE values, reflecting their stability. Linear Regression performed reasonably with slightly higher error margins, while K-Nearest Neighbors continued to be sensitive to the variations in the test data. Validation outcomes were displayed on the admin UI, showing both MSE and accuracy side by side for convenient comparison. These outcomes guided the manual choice of the best model to be finally deployed.

| **Evaluation Metric** | **Previous System** | **Proposed System** |
| --- | --- | --- |
| **Sales Accuracy (%)** | 81.23% | 98.86% |
| **Profit Accuracy (%)** | 69.50% | 87.34% |
| **Sales MSE ($)** | 2,350,000.00 | 486,853.76 |
| **Profit MSE ($)** | 2,010,000.00 | 440,494.77 |

## **4.5 RESULTS**

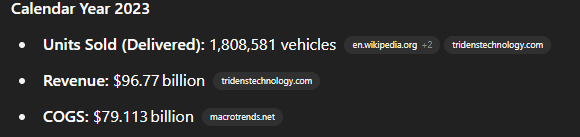
This section presents the output results of the four trained models after evaluation, highlighting their prediction performance for both sales and profit. Each model was trained and tested on the same dataset, and their results were displayed clearly on the admin interface for easy comparison.

The models included in the system were Linear Regression, Random Forest, Gradient Boosting, and K-Nearest Neighbors. Once training was completed, each model's performance based on Mean Squared Error (MSE) and prediction accuracy was calculated and displayed. These results were used to determine which model performed best under the given dataset conditions.

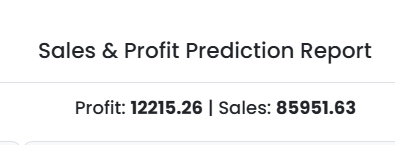


After reviewing the metrics, the admin manually selects the model with the highest accuracy and lowest error for deployment. Once selected, this model is used to generate predictions for new inputs entered into the system.

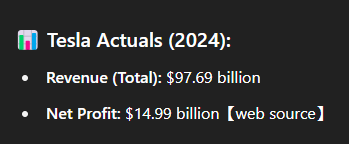
***Tesla 2023 values***



***Our predictions based on those values***



***Actual Tesla sales for 2024***



Our prediction were reasonably close, especially for a forecasting model. We were within within ~12% on sales and ~18% on profit.

Each of the four models Linear Regression, Random Forest, Gradient Boosting, and K-Nearest Neighbours was evaluated with the same inputs for a fair assessment. Subsequently, the output values of both target features were plotted on the administrator dashboard, allowing for visual along with quantitative comparison of model performance. The system presented each model's results in individual cards, indicating their respective MSE values and prediction accuracy against both sales and profit.

Among the four models, Random Forest and Gradient Boosting yielded the lowest MSE values and highest accuracy percentages for the validation set. These ensemble models were more capable of recognizing complex patterns in the data and therefore were more reliable for real-world estimates. Linear Regression was also acceptable, though slightly less accurate due to its linear nature. K-Nearest Neighbours performed variably depending on the input value density.

**4.6 CHAPTER SUMMARY**

This chapter presented the entire deployment of the sales and profit forecasting system, outlining four machine learning model developments, codebase architecture, and integration of the Python backend and ASP.NET interface using FastAPI. The testing process utilized to authenticate the consistency of each module, followed by performance analysis in terms of MSE and prediction accuracy, was also outlined. The results indicated that Random Forest and Gradient Boosting both performed better, thereby informing the choice of the final model that was used in making predictions.

## **CHAPTER 5**

**RECOMMENDATION AND CONCLUSION**

**5.1 INTRODUCTION**

This chapter concludes the study and presents future work improvements, recommendations for future research, and development in creating a machine learning model to predict sales and profit.

**5.2 CONCLUSION**

This project focused on developing a predictive analytics system capable of forecasting sales and profit using machine learning techniques. The system utilized four regression models—Linear Regression, Random Forest, Gradient Boosting, and K-Nearest Neighbors—trained on structured historical data. Python was used to develop and evaluate the models, while the ASP.NET interface allowed admin users to interact with the system through a user-friendly dashboard. FastAPI facilitated communication between the backend and the interface.

Throughout the implementation, the models were assessed based on Mean Squared Error (MSE) and prediction accuracy. The results demonstrated that ensemble models, particularly Random Forest and Gradient Boosting, consistently outperformed the others in both training and validation phases. The final system enabled users to manually select the best-performing model and generate predictions in real time. Overall, the project achieved its objective of building a reliable, accessible, and interpretable sales forecasting tool.

**5.3 Future Enhancements**

* **Dynamic CSV Upload and Mapping:**  
  Allow users to upload CSV files for prediction, with smart column mapping and validation to accommodate different data structures. This would make the system more flexible for real-world usage.
* **Real-Time Data Integration:**  
  Connect the system to live data sources such as sales databases or APIs, enabling continuous model updates and real-time forecasting without manual data input.
* **Scheduled Model Retraining:**  
  Implement periodic retraining mechanisms so that models can learn from new data over time, improving long-term accuracy and adaptability.
* **Advanced Visualization Dashboards:**  
  Enhance the interface with graphical dashboards (charts, trends, comparisons) to give users deeper insights beyond just numerical predictions.
* **Cloud Deployment and Access Control:**  
  Deploy the system on a cloud platform (e.g., Azure, AWS, or GCP) to enable remote access, scalability under higher traffic, and multi-user role-based access management.
* **Additional Model Support:**  
  Integrate more advanced models like XGBoost, LightGBM, or neural networks to test and improve prediction quality on more complex datasets.

**5.2 Limitations**

While the system successfully achieved its core objective of predicting sales and profit using machine learning models, several limitations were observed during development and testing:

* **Manual Model Selection:** The system currently requires the admin to manually review and select the best-performing model based on displayed metrics. This may introduce human bias or delay, especially when dealing with larger datasets or more frequent model retraining.
* **Limited Dataset Scope:** The training data was relatively limited in size and scope, consisting of only a few key variables. Real-world sales data often includes additional features such as promotions, seasons, region, or customer demographics, which were not captured in this project.
* **Static Input Format:** For prediction, the user must input values manually in a fixed format. The system lacks flexibility to handle diverse input structures or batch predictions via file uploads on the user side.
* **Local Deployment Only:** The entire system is hosted and tested locally. It is not currently deployed on a cloud platform, which limits access, scalability, and external integration capabilities.

**5.4 RECOMMENDATIONS**

* **Automated Model Selection:** Incorporating a function that automatically selects the best-performing model based on performance metrics would reduce manual oversight and improve efficiency.
* **Input Validation & Flexibility:** Enhancing the user interface to accept CSV uploads with dynamic column validation would make the system more scalable and user-friendly.
* **Model Expansion:** Adding more advanced algorithms such as XGBoost or neural networks could further improve prediction accuracy on larger datasets.
* **Real-Time Integration:** Connecting the system to real-time data sources (e.g., POS systems or databases) would enable continuous forecasting without manual input.
* **Cloud Deployment:** Hosting the application on cloud platforms like Azure or AWS would improve accessibility, scalability, and performance under higher data loads.

**5.5 CHAPTER SUMMARY**

This last chapter of the study addresses some of the limitations of the study, summarizes the outputs or achievements of the study, and points out future improvements that could be applied to machine learning models in sales and profit prediction.

## **REFERENCES**

Aguilar-Moreno, J. A., Palos-Sanchez, P. R., & Pozo-Barajas, R. (2024). Sentiment analysis to support business decision-making. A bibliometric study. *AIMS Mathematics*, *9*(2), 4337–4375. https://doi.org/10.3934/math.2024215

Altarazi, F., & Santos, D. (2024, June 4). Profit Prediction Using Machine Learning and Regression Models: A Comparative Study. *Proceedings of the International Conference on Industrial Engineering and Operations Management*. https://doi.org/10.46254/NA09.20240053

Amrutkar, P., & Mahadik, S. (2022). Sales Prediction Using Machine Learning Techniques. In *International Journal of Research Publication and Reviews Journal homepage: www.ijrpr.com* (Vol. 3). www.ijrpr.com

Anand, V., Brunner, R., Ikegwu, K., & Sougiannis, T. (2019). Predicting Profitability Using Machine Learning. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3466478

Anand, V., Brunner, R., & Sougiannis, T. (n.d.). *Predicting Profitability Using Machine Learning*.

Bajaj, P., Ray, R., Shedge, S., Vidhate, S., & Shardoor, N. (2020). SALES PREDICTION USING MACHINE LEARNING ALGORITHMS. *International Research Journal of Engineering and Technology*. www.irjet.net

Ganguly, P., & Mukherjee, I. (n.d.). *Enhancing Retail Sales Forecasting with Optimized Machine*. 1–5.

Ghude, P., Padekar, M., Alam, P., & Sangam, S. (2024). *Issue 6 www.jetir.org (ISSN-2349-5162) JETIRGH06010 Journal of Emerging Technologies and Innovative Research (JETIR) www.jetir* (Vol. 11). www.jetir.org

Hitesh S M, Yukthi A, & Ramya, B. N. (2024). *Sales Prediction Using Machine Learning Techniques*. www.ijnrd.org

Ito, S., & Fujimaki, R. (2016). *Optimization Beyond Prediction: Prescriptive Price Optimization*. http://arxiv.org/abs/1605.05422

Ji, S., Wang, X., Zhao, W., & Guo, D. (2019). An application of a three-stage XGboost-based model to sales forecasting of a cross-border e-commerce enterprise. *Mathematical Problems in Engineering*, *2019*. https://doi.org/10.1155/2019/8503252

Kumar, K., Devi, R., & Goel, P. (2024). *Sales Prediction Through Time Series Analysis with Machine Learning*. *AI4IoT 2023*, 163–169. https://doi.org/10.5220/0012608500003739

Mirshekari, S., Motedayen, N. H., & Ensaf, M. (2024). *Integrating Marketing Channels into Quantile Transformation and Bayesian Optimization of Ensemble Kernels for Sales Prediction with Gaussian Process Models*. http://arxiv.org/abs/2404.09386

Navya, Tn., Sruthi, B., Sukanya, D., Himavarsha, karumudi, & Srilakshmi Navya Sri, T. (2024). Issue 11 www.jetir.org (ISSN-2349-5162). In *JETIRGO06058 Journal of Emerging Technologies and Innovative Research* (Vol. 11). www.jetir.org

Pavlyshenko, B. M. (n.d.). *Machine-Learning Models for Sales Time Series Forecasting*. https://doi.org/10.3390/data4010015

Raji, I. D., & Buolamwini, J. (2022). Actionable Auditing Revisited: - Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products. *Communications of the ACM*, *66*(1), 101–108. https://doi.org/10.1145/3571151

Rezazadeh, A. (2020). A Generalized Flow for B2B Sales Predictive Modeling: An Azure Machine-Learning Approach. *Forecasting*, *2*(3), 267–283. https://doi.org/10.3390/forecast2030015

S, N. K. (n.d.). *Predictive Analysis for Big Mart Sales using Machine Learning Algorithms*. https://halobi.com/blog/sales-forecasting-five-uses/.

Sai, D., Reddy, R., Reddy, K. S., Ravindra, S. N., & Sahithi, B. S. (2021). Prediction and Forecasting of Sales Using Machine Learning Approach. *International Research Journal of Engineering and Technology*. www.irjet.net

Van Calster, T., Bossche, F. Van den, Baesens, B., & Lemahieu, W. (2020). *Profit-oriented sales forecasting: a comparison of forecasting techniques from a business perspective*. http://arxiv.org/abs/2002.00949

**APPENDIX A**

Include in the appendices reference materials that are too lengthy for the main thesis. If the appendix includes tables and figures, label them as Table A.1, Figure A.1, and so on.