

A
Mini-Project Report on
**SALES PREDICTION USING MACHINE
LEARNING**

Submitted in partial fulfillment of the requirements
for the degree of
BACHELOR OF ENGINEERING
IN
Computer Science & Engineering
Artificial Intelligence & Machine Learning

by

Sarang Sawant (22106080)
Harsh Salunkhe (22106133)

Under the guidance of
Prof. Taruna Sharma



Department of Computer Science & Engineering
(Artificial Intelligence & Machine Learning)
A. P. Shah Institute of Technology
G. B. Road, Kasarvadavali, Thane (W)-400615
University Of Mumbai
2023-2024



A. P. SHAH INSTITUTE OF TECHNOLOGY

CERTIFICATE

This is to certify that the project entitled “**Sales Prediction using Machine Learning**” is a bonafide work of Harsh Salunkhe (22106133), Sarang Sawant (22106080) submitted to the University of Mumbai in partial fulfillment of the requirement for the award of **Bachelor of Engineering in Computer Science & Engineering (Artificial Intelligence & Machine Learning)**.

Prof. Taruna Sharma
Mini Project Guide

Dr. Jaya Gupta
Head of Department



A. P. SHAH INSTITUTE OF TECHNOLOGY

Project Report Approval

This Mini project report entitled “**Sales Prediction using Machine Learning**” by **Harsh Salunkhe, Sarang Sawant** is approved for the degree of *Bachelor of Engineering* in *Computer Science & Engineering*, (AIML) 2022-23.

External Examiner: _____

Internal Examiner: _____

Place: APSIT, Thane

Date:

Declaration

We declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission hasnot been taken when needed.

Sarang Sawant
(22106080)

Harsh Salunkhe
(22106133)

ABSTRACT

Most of the business organizations heavily depend on a knowledge base and demand prediction of sales trends. Intelligent Decision Analytical System requires integration of decision analysis and predictions. The accuracy in sales forecast provides a big impact in business. Data mining techniques are very effective tools in extracting hidden knowledge from an enormous dataset to enhance accuracy and efficiency of forecasting. The detailed study and analysis of comprehensible predictive models to improve future sales predictions are carried out in this research. Traditional forecast systems are difficult to deal with the big data and accuracy of sales forecasting. These issues could be overcome by using various data mining techniques. In this paper, we briefly analyzed the concept of sales data and sales forecast. The various techniques and measures for sales predictions are described in the later part of the research work. On the basis of a performance evaluation, a best suited predictive model is suggested for the sales trend forecast. The results are summarized in terms of reliability and accuracy of efficient techniques taken for prediction and forecasting. The studies found that the best fit model is Gradient Boost Algorithm, which shows maximum accuracy in forecasting and future sales prediction.

INDEX

Index	Page no.
Chapter-1	
Introduction	1
Chapter-2	
Literature Survey	
2.1 History	4
2.1 Review	4
Chapter-3	
Problem Statement	7
Chapter-4	
Experimental Setup	
4.1 Hardware setup	9
4.2 Software Setup	9
Chapter-5	
Proposed system and Implementation	
5.1 Block Diagram of proposed system	11
5.2 Description of Block diagram	11
5.3 Implementation	12
Chapter-6	
Conclusion	15
References	16

Table of figures:

Figure no.	Figure name	Page no.
5.1	Block diagram of Sales Prediction	11
5.3.1	Implementation of Sales Prediction	11
5.3.2	Result of Sales Prediction	12

CHAPTER 1

INTRODUCTION

1. INTRODUCTION

Sales prediction using machine learning is a powerful tool that businesses can leverage to forecast future sales based on historical data and other relevant factors. By applying machine learning algorithms to analyze patterns and trends in sales data, companies can make informed decisions, optimize inventory management, and improve overall business performance. This project aims to develop a predictive model that can accurately forecast sales, helping businesses anticipate demand, allocate resources efficiently, and ultimately drive revenue growth. Through the use of advanced machine learning techniques, this project seeks to provide valuable insights and predictions that can enhance decision-making processes and drive success in the competitive marketplace.

One of the most common methods used to predict sales is **regression analysis**. This method involves using historical sales data to train a model that can predict future sales. The model can take into account factors such as **past sales or marketing campaigns or economic indicators** to make its predictions..

Another approach is using **decision tree-based algorithms** like **Random Forest, Gradient Boosting** etc. These algorithms are particularly useful when there are many factors that can influence sales, such as product features, customer demographics, and market conditions. The algorithm can help identify the most important factors and use them to make predictions..

Machine learning can also be used to predict sales by using **clustering algorithms**, which can help identify groups of similar customers. This information can then be used to create targeted marketing campaigns and improve sales strategies.

.

CHAPTER 2

LITERATURE SURVEY

2.LITERATURE SURVEY

2.1-HISTORY:

The history of sales prediction using machine learning dates back to the early 2000s when businesses started leveraging machine learning algorithms to forecast sales trends more accurately. Over the years, advancements in technology and the availability of vast amounts of data have enabled companies to develop sophisticated predictive models that can analyze historical sales data, customer behavior, market trends, and other relevant factors to make accurate sales predictions.

Machine learning algorithms such as regression, decision trees, neural networks, and ensemble methods have been widely used in sales prediction models to identify patterns and relationships in data that can help businesses optimize their sales strategies, inventory management, and pricing decisions. By continuously refining and updating these models with new data, companies can improve the accuracy of their sales forecasts and make more informed business decisions.

Overall, the history of sales prediction using machine learning showcases the evolution of predictive analytics in the sales domain, enabling businesses to anticipate market changes, identify growth opportunities, and stay ahead of the competition.

2.2-LITERATURE REVIEW

Plenty of studies and researches are being made in sales prediction using machine learning. Machine learning is being chosen for our project because it is such a computer programs which learns by itself and does not need any human interference. Sales forecasts help in predicting the future sales. This is an important precondition for decision-making of enterprises. It helps in predicting the future sales thus helping in increasing the profit.[1] A Forecast for Big Mart Sales Based on Random Forests and Multiple Linear Regression (2018) - Kadam, H., Shevade, R., Ketkar, P. and Rajguru.[2]A Forecast for Big Mart Sales Based on Random Forests and Multiple Linear Regression.XG boost Algorithm is used because it gives more accuracy when compared to random forest and linear regression in case of prediction analysis. [3]Forecasting methods and applications (2008)- Makridakis, S., Wheelwrigth.S.C., Hyndman. R.J.Forecasting methods and applications contains Lack of Data and short life cycles. So some of the historical data if stored can help in accurate prediction of uncertain demands in case of consumer oriented markets. [4]Comparison of Different Machine Learning Algorithms for Multiple Regression onBlack Friday Sales Data (2018) - C. M. Wu, P. Patil and S. Gunaseelan.Comparison of Different Machine Learning Algorithms for Multiple Regression on Black Friday Neural Network was used by sales data for comparing different algorithms.To overcome such complex models simpler algorithms are used for prediction. [5]Prediction of retail sales of footwear using feed forward and recurrent NeuralNetworks (2018) - Das, P., ChaudhuryPrediction of retail sales of footwear using feed forward and recurrent neural networksneural network for predicting of weeklyretail.

CHAPTER 3

PROBLEM STATEMENT

3.PROBLEM STATEMENT

Data Collection and Preprocessing: Collecting and preprocessing historical sales data from retail businesses to build a robust dataset for model training and evaluation.

Model Development and Evaluation: Developing and evaluating various machine learning models, including regression models, time series forecasting models, and ensemble methods, to identify the most suitable approach for accurate sales prediction.

Integration and Deployment: Integrating the developed models into a user-friendly web-based interface or application for easy access and interaction by retail businesses. Deployment will involve testing the system in real-world retail environments to validate its performance and usability.

CHAPTER 4

EXPERIMENTAL SETUP

4.EXPERIMENTAL SETUP

4.1 Hardware Setup

To set up a hardware system for sales prediction using machine learning, one will need a computer with a good amount of processing power and memory. Ideally, a machine with a multi-core processor, at least 8GB of RAM, and a dedicated GPU for faster training of machine learning models would be beneficial. Additionally, having storage space for storing large datasets and model files is important. You may also consider using cloud-based services for scalability and flexibility in your hardware setup. Remember to install necessary software tools and libraries for machine learning, such as Python, TensorFlow, or scikit-learn. It's important to regularly update your hardware and software to ensure optimal performance for your sales prediction tasks.

4.2 Software Setup:

For the software setup for sales prediction using machine learning, one will need to install various tools and libraries to build and train your machine learning models. Here are some key components for your software setup:

1. Visual Studio Code: Visual Studio code is a source-code editor developed by Microsoft for Windows, Linux, macOS and web browsers. Features include support for debugging, syntax highlighting, intelligent code completion, snippets, code refactoring, and embedded version control with Git.
2. Python: Python is a popular programming language for machine learning and data analysis. Make sure to install Python on your system.
3. Pandas: Pandas is a powerful data manipulation library in Python. It provides data structures and functions to work with structured data, such as CSV files or databases.
4. NumPy: NumPy is a fundamental package for scientific computing in Python. It provides support for large, multi-dimensional arrays and matrices, along with a collection of mathematical functions to operate on these arrays.
5. Scikit-learn: Scikit-learn is a machine learning library in Python that provides simple and efficient tools for data mining and data analysis. It includes various algorithms for classification, regression, clustering, and more.
6. Matplotlib or Seaborn: Matplotlib and Seaborn are plotting libraries in Python that help you create visualizations of your data and model results.
7. Jupyter Notebook: Jupyter Notebook is a web-based interactive development environment that allows you to create and share documents that contain live code, equations, visualizations, and narrative text. It's a great tool for experimenting with machine learning models.

CHAPTER 5

PROPOSED SYSTEM & IMPLEMENTATION

5. PROPOSED SYSTEM & IMPLEMENTATION

5.1 Block diagram of proposed system

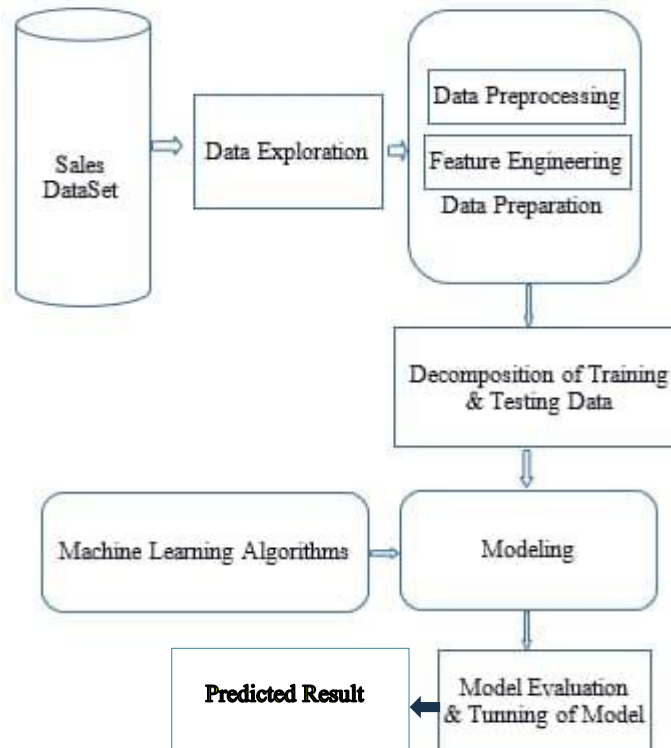


FIG: 5.1 BLOCK DIAGRAM OF SALES PREDICTION

5.2 Description of block diagram

1. **Data Collection:** The first step is to gather relevant data such as historical sales data, customer demographics, market trends, and any other information that may impact sales.
2. **Data Preprocessing:** This stage involves cleaning the data, handling missing values, and transforming the data into a format that can be used by machine learning algorithms.
3. **Feature Engineering:** This step involves selecting and creating relevant features from the data that will help the machine learning model make accurate predictions.
4. **Model Selection:** In this stage, different machine learning algorithms are tested and evaluated to determine which one is best suited for the sales prediction task.
5. **Training:** The selected model is trained on the historical data to learn the patterns and relationships between the features and the sales outcomes.

6. Evaluation: The trained model is then evaluated using a separate set of data to assess its performance.

7. Deployment: Once the model has been trained and evaluated, it can be deployed to make real-time sales predictions based on new data inputs.

Overall, the block diagram of sales prediction using machine learning involves a series of steps to collect, preprocess, model, train, evaluate, and deploy a predictive model to forecast sales accurately.

5.3 Implementation

Implementation of proposed system must be included here. Students can explain implementation using screen shots of output.

The screenshot displays a web browser window with the title 'Big Mart Sales Prediction'. The page contains several input fields and dropdown menus for predicting sales. The inputs are as follows:

- Item Weight: 9.300000
- Item Weight dropdown: Regular
- Item Visibility: 0.016047
- Item Visibility dropdown: Meat
- Item MRP: 249.809204
- Outlet Establishment Year (YYYY): 1999
- Tier dropdown: High
- Tier dropdown: Tier 1
- Outlet type dropdown: outlet_type

At the bottom of the form, there are two buttons: 'Submit' (blue) and 'Reset' (red).

FIG: 5.3.1 IMPLEMENTATION OF SALES PREDICTION



Predicted sales is :
6845 units

FIG:5.3.2 PREDICTED RESULT

5.4 Advantages/Applications:

1. Improved accuracy: Machine learning algorithms can analyze large amounts of data to identify patterns and trends that may not be obvious to humans. This can lead to more accurate sales predictions.
2. Real-time insights: Machine learning models can process data quickly, allowing businesses to make real-time decisions based on the most up-to-date information available.
3. Increased efficiency: By automating the sales prediction process, machine learning can save time and resources for businesses, allowing sales teams to focus on other important tasks.
4. Personalized recommendations: Machine learning can analyze customer data to provide personalized product recommendations, helping businesses tailor their sales strategies to individual customer needs.

Applications:

1. Demand forecasting: Machine learning can be used to predict future demand for products or services, helping businesses optimize their inventory levels and production schedules.
2. Customer segmentation: By analyzing customer data, machine learning algorithms can segment customers into different groups based on their purchasing behavior, allowing businesses to tailor their marketing strategies to specific customer segments.
3. Lead scoring: Machine learning can help sales teams prioritize leads by predicting the likelihood of a lead converting into a customer based on historical data and customer behavior.

CHAPTER 6

Conclusion

6.CONCLUSION

Based on the analysis conducted in the sales prediction using machine learning project, it can be concluded that the model developed shows promising results in accurately forecasting sales. The use of machine learning algorithms has proven to be effective in capturing patterns and trends in the sales data, leading to more accurate predictions. However, it is important to note that the model's performance may vary depending on the quality and quantity of data available. Further optimization and fine-tuning of the model could potentially improve its predictive capabilities. Overall, the project demonstrates the potential of machine learning in enhancing sales forecasting accuracy and efficiency. In conclusion, Machine Learning can be a powerful tool in the hands of businesses to predict sales and make informed decisions. With a combination of various algorithms, historical data and neural networks, businesses can improve their sales and make better decisions for their future

REFERENCES

- [1]Forecasting methods and applications (2008)- Makridakis, S., Wheelwriqh.S.C., Hyndman. R.J
- [2] Black Friday Sales Data (2018) - C. M. Wu, P. Patil and S. Gunaseelan.
- [3] Predictive modeling for retail sales time series data. (2017)-YILDIZ, T., & ACAR, Z.
- [4]] A Forecast for Big Mart Sales Based on Random Forests and Multiple Linear Regression (2018) - C. M. Wu, P. Patil and S. Gunaseelan
- [5]NeuralNetworks (2018) – Das, P., Chaudhury