

**University of Mumbai**  
**2022-2023**

**M.H. SABOO SIDDIK COLLEGE OF ENGINEERING**  
**8, Saboo Siddik Road, Byculla, Mumbai - 400 008**  
**Department of Computer Engineering**



**Project Report**

**On**

**Profit Booking by Sales Forecasting using Machine Learning**

**By**

**Aman Ahsan Khan [3119017]**  
**Ramiz ur Rehman Khan[3119023]**  
**Sharjeel Ali Momin[3119030]**  
**Aleem Tajir[3119045]**

**Guided By**  
**Dr. Lutful Islam**



M.H. SABOO SIDDIK COLLEGE OF ENGINEERING  
8, Saboo Siddik Road, Byculla, Mumbai -400 008.

This is to certify that,

Roll No	Name
3119017	Aman Khan
3119023	Ramiz Khan
3119030	Sharjeel Momin
3119045	Aleem Tajir

Of Final Year/Second year (B.E./S.E Semester VII/III) degree course in Computer Engineering,  
have completed the specified project synopsis on,

## **Profit Booking by Sales Forecasting using Machine Learning**

As a partial fulfillment of the project work in a satisfactory manner as per the rules of the curriculum laid by the University of Mumbai, during the Academic Year July 2019 — June 2020.

\_\_\_\_\_  
Internal Guide

\_\_\_\_\_  
Internal Examiner

\_\_\_\_\_  
External Examiner

\_\_\_\_\_  
Head of Department

\_\_\_\_\_  
Principal

## **Acknowledgement**

We wish to express our sincere thanks to our director Dr. Mohiuddin Ahmed and our principal Dr. Ganesh Kame,M.H. Saboo Siddik College of Engineering for providing us all the facilities, support and wonderful environment to meet our project requirements.

We would also take the opportunity to express our humble gratitude to our Head of Department of Computer Engineering Dr. Zainab Pirani for supporting us in all aspects and for encouraging with her valuable suggestions to make our project successful.

We are highly thankful to our internal project guide [Type the name of Guide] whose valuable guidance helped us understand the project better, her constant guidance and willingness to share her vast knowledge made us understand this project and its manifestations in great depths and helped us to complete the project successfully.

We would also like to acknowledge with much appreciation the role of the staff of the Computer Department, especially the Laboratory staff, who gave the permission to use the labs when needed and the necessary material to complete the project.

We would like to express our gratitude and appreciate the guidance given by other supervisors and project guides, their comments and tips helped us in improving our presentation skills.

Although there may be many who remain unacknowledged in this humble note of appreciation, there are none who remain unappreciated.

## **Table of Contents**

<b>Sr. No</b>	<b>Topic</b>	<b>Page No.</b>
	<b>Abstract</b>	<b>6</b>
<b>1</b>	<b>Introduction</b>	<b>7</b>
<b>2</b>	<b>Literature Survey</b>	<b>10</b>
<b>2.1</b>	<b>Survey Existing system</b>	<b>12</b>
<b>2.2</b>	<b>Limitation Existing system or research gap</b>	<b>13</b>
<b>2.3</b>	<b>Problem Statement and Objective</b>	<b>13</b>
<b>2.4</b>	<b>Scope</b>	<b>14</b>
<b>3</b>	<b>Proposed System</b>	<b>16</b>
<b>3.1</b>	<b>Analysis/Framework/Algorithm</b>	<b>18</b>
<b>3.2</b>	<b>Details of Hardware &amp; Software</b>	<b>20</b>
<b>3.2.1</b>	<b>Hardware Requirements</b>	<b>20</b>
<b>3.2.2</b>	<b>Software Requirements</b>	<b>20</b>
<b>3.2.3</b>	<b>Technology Used</b>	<b>20</b>
<b>3.3</b>	<b>Design details</b>	<b>21</b>
<b>3.3.1</b>	<b>Flowchart</b>	<b>21</b>
<b>3.3.2</b>	<b>Data Flow Diagram</b>	<b>22</b>
<b>3.3.3</b>	<b>Use Case Diagram</b>	<b>23</b>
<b>3.4</b>	<b>Methodology</b>	<b>24</b>
<b>4</b>	<b>Implementation Plan for Next Semester</b>	<b>26</b>
<b>5</b>	<b>Conclusion</b>	<b>29</b>
	<b>References</b>	<b>31</b>

### **List of Figures**

**[Indicate the figures wherever used]**

<b>Fig. No.</b>	<b>Figures</b>	<b>Page No.</b>
<b>2.1</b>		<b>12</b>
<b>3.1</b>		<b>18</b>
<b>3.2</b>		<b>18</b>
<b>3.3</b>		<b>21</b>
<b>3.4</b>		<b>22</b>
<b>3.5</b>		<b>23</b>
<b>3.6</b>		<b>24</b>
<b>4.1</b>		<b>29</b>
<b>4.2</b>		<b>29</b>
<b>4.3</b>		<b>29</b>

## **Abstract**

Algorithms for machine learning are used by businesses to forecast sales and revenue. This is accomplished by forecasting consumer behavior using information from prior transactions and new market trends. By doing this, businesses may make precise projections and get ready for upcoming events. It involves predicting current or upcoming sales utilizing information such as previous sales, seasonality, holidays, economic situations, etc. So, given a specific set of inputs, this model will forecast sales on a particular day. Past seven day sales were one of the model's eight input factors. The implementation of the model will be done using the latest and advanced machine learning algorithms and classifiers which will provide better Accuracy as well as performance. All inputs are first preprocessed so that the machine can interpret them. Given that this is a supervised learning-based linear regression model, the output will be made available along with the input. The model is then fed the inputs and the intended output. The model will identify a function between the input and output and plot it (learn it). The output for a particular set of inputs is then predicted using this function or relation. In this instance, the amount of sales is designated as the output while input information like the date and past sales are designated as input.

## **1. Introduction**

One of the most crucial things a firm does is sales forecasting. It supports budgeting and personnel across the board for an organisation, as well as sales planning. Despite its significance, many firms still employ outdated methods that lead to inaccurate estimates. It's possible to make a comparison to earlier periods, when farmers relied on indicators like cats cleaning behind their ears or an elderly person's knee pain to predict the weather. This approach to forecasting is useful for gaining insights about how a company should manage its personnel, resources, and cash flows in addition to anticipating sales. This is essential for effectively allocating internal resources to take advantage of forecasts or trends, and it also serves as a critical performance indicator for businesses trying to gather investment funding. However, in big businesses, sales forecasting technologies are still just marginally more accurate than an arthritic knee. Before producing predictions, machine learning analyses millions of data points using a variety of techniques, including as grouping and regression. Demographic details, behavioural patterns, and historical transactions are examples of data points. In order to estimate the amount of income, this approach of sales forecasting assigns a likelihood of successfully closing a deal to each stage of the sales process, then multiplies that probability by the size of a specific opportunity. This forecasting technique is still one of the more well-liked options because of how straightforward it is. The age of the opportunity, or the period of time from when it originally arose to the present, is not taken into account in this method, although being a significant influencing factor. Machine learning can forecast future customer behaviour and the likelihood that a given percentage of consumers will complete a transaction by examining this data. Algorithms for machine learning are used by businesses to forecast sales and revenue. This is accomplished by forecasting consumer behaviour using information from prior transactions. By doing this, businesses may make precise projections and get ready for upcoming events. Good data spanning years is one of the most important requirements for accurate sales forecasting (in the case of an existing business). Newer businesses can instead rely on industry trends and averages since they lack the data necessary to make precise projections. Of course, humans don't always buy and consume in the same ways. However, it is very likely to find trends in sales and consumption when supported by statistics that go back years. This can be done individually or with a particular audience or demographic in mind. Because it can search through years of data to find patterns and repetitive behaviour, machine learning is used in this context to forecast sales and demand. The majority of machine learning algorithms are based on probability; even in instances involving sales forecasting, probability algorithms are essential for projecting certainty far into the future. The fundamental tenant of probability is that an event's probability can only be either 0 or 1, i.e., either impossible or unavoidable. The algorithms consider the likelihood that the sales process will be successful as well as any external and internal influencing factors. Sales leaders can examine how to use these insights as levers to have a larger positive impact on sales by doing this type of "what if" analysis to understand the impact of these factors on sales metrics.

# LITERATURE SURVEY

## **2.Literature Survey**

[Type here the detailed Literature survey of your project. Literature survey means whatever work you have found about your project which is already being implemented. Give the references of the papers.]

### **2.1 Survey Existing system**

### **2.2 Limitation Existing system or research gap**

### **2.3 Problem Statement and Objective**

### **2.4 Scope**

#### **2.4.1 Justification:**

#### **2.4.2 Product scope description:**

#### **2.4.3 Acceptance criteria:**

#### **2.4.4 Deliverables:**

#### **2.4.5 Assumptions:**



### 3. Proposed System

**AIM:** Our main goal is to build an advanced system that can accurately forecast or give the prediction for the profit of the company using previous data and history of transactions of an individual.

**Main Features and goals are given below,**

- Accurately predict the profit
- Predicting profit for specific time span (Time Exponential)
- Building an easily usable and efficient forecasting system

**Objectives of Proposed System:**

- To learn different kinds of ML algorithms that can be used for sales forecasting
- To understand the underlying patterns in product sales and marketing strategies
- To take necessary actions required to make system Accurate, reliable and efficient
- To avoid business loss by proposing advanced sales profit forecasting system

**Overview of Proposed System:**

Businesses can use machine learning to quickly and easily develop more complex forecasting models that make use of a larger data collection. By adding machine learning algorithms to their data, businesses can enhance their goods and services depending on consumer demands. Additionally, they can more correctly predict consumer behaviour, which will improve their planning. Our system will be consisting of different kind of modules which will be working on different Machine Learning Algorithms such as **linear regression and Random Forest Algorithm**. Then we'll use different types of libraries and classifiers such as **XGBoost Classifier** for better performance, execution time and Higher Accuracy.

**System Architecture:**

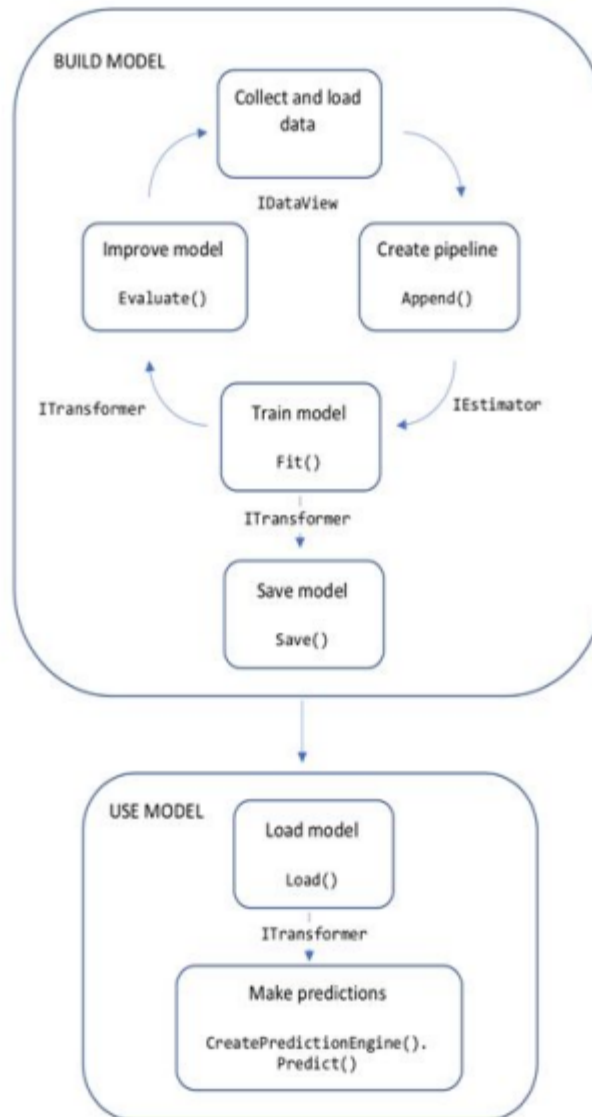


Fig: System Architecture

### 3.1 Analysis/Framework/Algorithm:

We will be using Linear Regression and Random Forest Algorithm in our project. As linear regression is easy to understand and use but it has some limitations in choosing data so we'll also be using Random Forest Algorithm. And we'll use Extreme Gradient boost classifier to provide a higher and enhanced accuracy to our model.

Random Forest Algorithm: Random Forest is supervised Machine Learning algorithm used in classification as well as regression problems. Random forests is an ensemble learning method for classification, regression and other tasks that operate by constructing a multitude of decision trees at training time and giving the output of the class. Random forests correct for decision trees' habit of overfitting to their training set. Also random forests work well for a vast range of data items than a one decision tree does and the scaling of data does not require a random forest classifier as well. This actually forces more diversification and more variation amongst the trees in the model and finally results in lower correlation across trees. So in the random forest, we are left with the trees which are not only trained on various different sets of data but also which use different features which helps them to make decisions.

XGBoost Classifier: A distributed gradient boosting library that has been improved for speed, adaptability, and portability is called XGBoost. Gradient Boosting is used to develop machine learning algorithms.

## **3.2 Details of Hardware & Software**

### **3.2.1 Hardware Requirements**

- System: Pentium 4, Intel Core i3, i5, i7 and 2GHz minimum
- RAM: 4GB or above
- Hard Disk: 10GB or above

### **3.2.2 Software Requirements**

- Operating System: Windows 7, 10 or Higher version
- Programming Language: Python

### **3.2.3 Technology Used**

This system is build keeping different technologies in mind and also the user needs as our first priority. Our main aim is to build a system that is efficient and accurate and eaasy to use. We are using different technologies like classification and regression algorithms such as Linear regression, Random Forest algorithm and we are using different kind of python libraries and the XGBoost Classifier that will provide us the better results and more accurate predictions.

### 3.3 Design details

#### 3.3.1 Flowchart

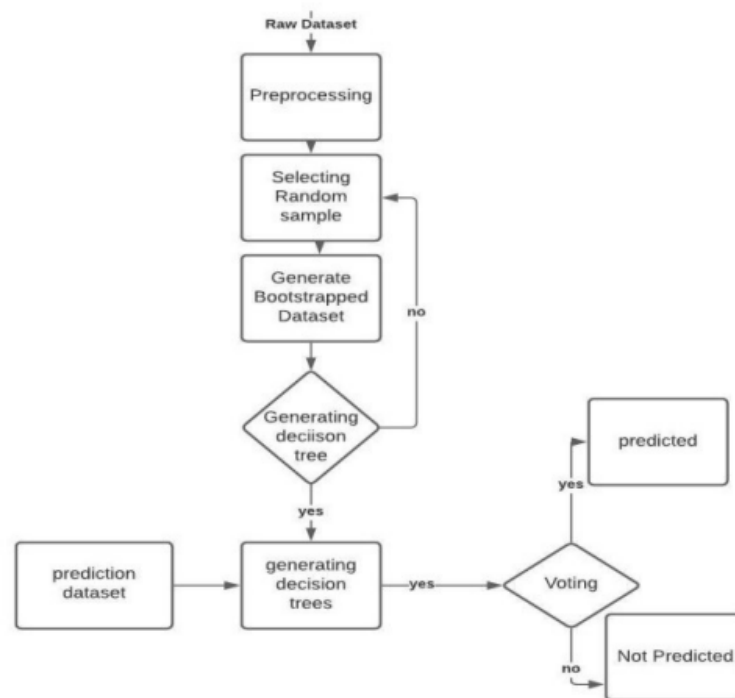


Fig: Flowchart

### **3.4 Methodology**

The main approach to build this model system is by using advanced techniques of machine learning and to make the model more efficient and handy than ever. We are using different techniques such as random forest and XGBoost which are more efficient. Random forest algorithm creates different decision trees and generate more optimal average result out of them which provides more accurate predictions in classification and regression problems.

## **4. Implementation**

Implementation would be done using machine learning algorithms that are described and we'll train our model for better performance and we'll try to reduce execution speed and add time exponential.

## **5. Conclusion**

The profit booking by sales forecasting is a great and efficient platform for all the financial and commercial organizations since they can finally predict their upcoming profits accurately by using the model made with advanced techniques and algorithms of machine learning. We have successfully done all the preprocessing required for future analysis and implementation. The machine learning Algorithms used will provide higher accuracy, so an individual or an organization can now finally sit back and see their blooming profits on their assets daily.

## References

- [1] Research on sales information prediction system of e-commerce enterprises based on time series model Authors: Jian Liu, Chunlin Liu, Lanping Zhang & Yi Xu
- [2] a completely unique Trigger Model for Sales Prediction with data processing Techniques Authors: Wenjie Huang , Qing Zhang, Wei Xu, Hongjiao Fu, Mingming Wang, Xun Liang
- [3] Sales Prediction Model Using Classification Decision Tree Approach For Small Medium Enterprise Based on Indonesian E – Commerce Data Authors: Raden Johannes H. P., Andry Alamsyah
- [4] Sales Demand Forecast in E-commerce Using a Long Short-Term Memory Neural Network Methodology Author: Kasun Bandara, Peibei Shi, Christoph, Bergmeir, Hansika, Hewamanage, Quoc Tran, Brian Seaman
- [5] Sales Prediction System using Machine Learning Authors: Archisha Chandel, Akanksha Dubey, Saurabh Dhawale, Madhuri Ghuge
- [6] A survey of machine learning techniques for food sales prediction. Author: Grigorios Tsoumakas
- [7] Schroeder, G., Klim, A., Heinz, G., et al. (2010) System for predicting sales lift and profit of a product supported historical sales information: U.S. Patent 7,689,456.
- [8] Yuan, H., Xu, W & Wang, M. (2014) Can online user behavior improve the performance of sales prediction in E-commerce? IEEE International Conference on Systems, Man, and Cybernetics, Patent 2377–2382.
- [9] Pavlyuchenko, B.M. Linear, machine learning and probabilistic approaches for statistical analysis. In Proceedings of the IEEE First International Conference on Data Stream Mining & Processing (DSMP), Lviv, Ukraine, 23–27 August 2016; IEEE: Piscataway, NJ, USA, 2016; Patent . 377–381
- [10] Gao, M., Xu, W., Fu, H., Wang, M. & Liang, X. (2014). A novel forecasting method for large-scale sales prediction using an extreme learning machine. The Seventh International Joint Conference on Computational Sciences and Optimization, Patent 602–606 Processing Systems 25. Curran Associates, Inc., Patent. 2951–2959
- [11] Box, G.E.P. & Cox, D.R., 1964. An Analysis of Transformations. Journal of the Royal Statistical Society. Series B, Statistical methodology, 26(2) Patent. 211–252
- [12] Kulkarni, G., Kannan, P.K. & Moe, W., 2012. Using online search data to forecast new product sales. Decision support systems, 52(3), Patent. 604–611
- [13] Trapero, J.R., Kourentzes, N. & Fildes, R., 2015. On the identification of sales forecasting models within the presence of promotions. The Journal of the Operational Research Society, 66(2), Patent. 299–307
- [14] Rokach, L. Ensemble methods for classifiers. Data Mining and Knowledge Discovery Handbook; Springer: Cham, Switzerland, 2005; Patent. 957–980.
- [15] Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; ACM: New York, NY, USA, 2016; Patent 785–794.



