Tribhuvan University
Faculty of Humanities and Social Science

BigMart Sales Prediction System

A Project Report

Submitted To:

Department of Computer Application

PadmaShree International College

In partial fulfillment of the requirement for the Bachelor in Computer Application

Submitted By:

Aavash Ganesh

Registration No: 6-2-622-25-2018

June 2023

Under the Supervision of

Basanta Chapagain

Tribhuvan University

Faculty of Humanities and Social Science

PadmaShree International College

Supervisor's Recommendation

I hereby recommend that this project prepared under my supervision by **Aavash Ganesh** entitled **"BigMart Sales Prediction System"** in partial fulfillment of the requirements for the degree of Bachelor of Computer Application is recommended for the final evaluation.

....

SIGNATURE

Mr. Basanta Chapagain

Project Supervisor

Department of Computer Application

PadmaShree International College

Tribhuvan University

Faculty of Humanities and Social Sciences

PadmaShree International College

LETTER OF APPROVAL

This is to certify that this project prepared by **Aavash Ganesh** entitled **"BigMart Sales Prediction System"** in partial fulfillment of the requirements for the degree of Bachelor in Computer Application has been evaluated. In our opinion it is satisfactory in the scope and quality as a project for the required degree.

SIGNATURE OF SUPERVISOR Mr. Basanta Chapagain Project Supervisor Department of Computer Application PadmaShree International College	SIGNATURE OF COORDINATOR Mr. Ramesh Kumar Pudasini Program Coordinator Department of Computer Application PadmaShree International College
SIGNATURE OF INTERNAL EXAMINER Mr. Internal Examiner	SIGNATURE OF EXTERNAL EXAMINER Name: External Examiner

ABSTRACT

With the rise in the trend of shopping for their daily needed items from the big mart is increasing, the retailer are in the process of recording vast data related to their sales of items or products of each and every individual with their various dependent and independent factors. However, the large amount of recorded data presents create challenge when it come for manually analysing the sales.

BigMartSales Prediction is a online platform that uses the range of machine learning model and algorithm for predicting the future sales by building a predictive models and find out the future sales. It is the system mainly developed for the analytical solution to forecast sales for a BigMart , a retail chain industries. This system utilize the historical sales data along with the different variable and metrices to predict future sales accurately on the basis of different algorithm model primarily the linear regression algorithm to analyze the relationship between independent and dependent variable. Moreover, to figure out the best fit accuracy result, beside the linear regression other two algorithm i.e. random forest regression and xgbregressor algorithm are also implemented in this system. This system predict the sales on the basis of historical data and pattern mainly developed to help BigMart streamline in its operations, reduce costs, and increase profitability by providing accurate, real-time insights.

Keyword: accurate, best fit, historical data, predective model, regression, retail-chain

ACKNOWLEDGEMENT

TABLE OF CONTENTS

<u>Topics</u>   <u>Pages</u>

LISTS OF FIGURES

LISTS OF TABLE

CHAPTER 1

INTRODUCTION

1.1 Introduction

BigMart Sales Prediction System is one of the web-based platform designed for predicting the future sales of Bigmart, a retail chain store by building a predictive model and finding out the sales and revenue of each product at a particular store using various algorithms and methods such as regression analysis, decision tree, etc. on the base of different factor of dataset such as historical sales data, product sold information, location of the stores.

In the present context, the trend of people shopping for their daily necessary items from the big mart is increasing and in today's modern world, such stores are recording data related to their sales of items or products of each and every individual with their various dependent and independent factors as an important step for predicting future demand, sales and to update the inventory management. [1]

This prediction system can be term as an modern sales prediction tool designed specifically for Big Mart's needs. This platform offer a user-friendly interface that enables predictor to easily input various sales-related data points and metrics, such as sales figures, product prices, and other relevant attributes. The system was developed using powerful machine learning algorithms, such as regression analysis i.e., random forest regression, linear regression, and XGBRegressor, to accurately predict future sales trends based on the input data. In this system, sales predictor is provided with features of loading data and test it to easily train and deploy the models to make accurate sales predictions, helping to optimize inventory management, marketing strategies, and other important business decisions. The platform have been provide users with customizable dashboards and reports that enable them to analyze and visualize the sales data in real-time. With various charts, graphs, and other visualization tools, predictor can easily identify insights pattern that help to inform their decision-making. Overall, this sales prediction system was designed to help BigMart streamline its operations, reduce costs, and increase profitability by providing accurate, real-time insights.

1.2   Problem Statement

The expansion of BigMart's store in a different location has led to a significant increase in the amount of recorded data that the company needs to manage and manipulate. As the number of stores continues to grow, the workload of manually analyzing and forecasting sales data for each outlet becomes increasingly challenging and time-consuming which causes a problem in the effective prediction of revenue for each product, future demand in each outlet, and updating the inventory as per the recorded dataset. Similarly, the competition in such multi-shopping stores for increasing their sales is high, and predicting future sales without the help of a prediction system that consists of features of machine learning models may cause problems in analyzing historical sales data and identifying patterns and trends in future sales which eventually leads to the loss in the company profit. So, to resolve this problem, find out what role certain properties of an item play, and to achieve the goal, a prediction model can be built to find out for every store based on factor dataset that can increase their sales and what changes could be made to the available store product.

1.3   System Objectives

- To forecast future sales for the company and estimate the amount of revenue based on trends and patterns in historical sales data and other relevant factors.

- To predit the foresights sales for BigMart using a range of powerful machine learning algorithms, including linear regression, random forest regression, and XGB regressor.

1.4   Scope and Limitation

1.4.1   Scope

As the trend of shopping through retail chain i.e., bigmart is increasing, the data record related to the individual sales have also reach to the maximum volume and analysing the sales of such vast amount of data through manual process create challenge. Manually analyzing and visualizing the pattern create unscientific problem in decision making for the future goals. To address this problem, bigmart sales prediction came to exist to accurately predict the future sales through different machine learning model and provide a valuable decision making platform.

1.4.2 Limitation

- This system maynot provide the accurate future sales prediction and mean square error value as prediction rely on limited machine learning algorithm model.

- This system is primarily developed for sales prediction of BigMart and limited with its context, as it lack reinforcement model and self learning approach for other domain.

- The sales predictor is responsible for manual input of prediction metrics as this system doesnot have the functionality of automatic prediction features on the basis of file upload.

1.5   Development Methodology

As a part of software development, for the BigMart Sales Prediction System modified waterfall methodology has been use to minimize risk such as bug, cost overruns and changing requirement when add new functionality. The entire software development process was carried ou in a mechanism that allow to return to the previous phase for verification and validation. Further, the software was design under object oriented paradigm considering data and process modelling in

1.6   Report Organization

Chapter 1: Introduction

This chapter encompass the basis introduction to BigMart sales prediction system, problem arises in the current sales forecasting method and the demand of prediction system in BigMart. It also includes the objectives of the system. In additional, this chapter also cover the scope and limitation of the system. Furthermore, the methodology for software development and report organization was also consider in the chapter.

Chapter 2: Background Study and Literature Review

Under this chapter, the information regarding the background of prediction system along with their flourishing period are cover. Likewise, the review of existing system similar to bigmart sales prediction system are includes to find out flaws and required features that need to be include in this type of system.

Chapter 3: System Analysis and Design

Here, all the system analysis and design components are included in this chapter. Different software requirement i.e., functional and non-functional requirement and various feasibility study are encompass in this system analysis part

Chapter 4: System Implementation

Different tools used for implementating the system like case tool, programming language and framework, database platforms are included in this chapter. Similarly, testing of different module and testing as a software are also shown here.Moreover, the result analysis part are also include in this chapter.

Chapter 5: Lesson Learnt / Outcomes

In this chapter, the lesson learnt in the software development process, system outcomes and conclusion are noted here. Likewise, the future recommendations and enhancement part from the developer perspective are also described under this chapter.

CHAPTER 2

BACKGROUND STUDY AND LITERATURE REVIEW

2.1 Literature Review

A literature review is a comprehensive summary of previous research on a topic by authorized scholars and researchers regarding what has been done before on the topic. The purpose of the literature reviews is to express and enumerate, summarize what ideas have been established and what their positive and negative aspects are.

Sales prediction provides insight into how a firm should manage its workforce, cash flow, and means. This is an important precondition for the planning and decision-making of enterprises. Sales prediction and its analysis has been conducted by many authors as summarized.

The paper published titled 'Sales Prediction using Machine Learning' by Divya Choudhary describes the uses of different machine learning techniques such as linear regression, decision trees, and random forests to predict sales for retail stores. As per the authors, these techniques are applied to the BigMart dataset and compare their results. They found that the random forest model performed the best, with an accuracy of 75.3% compared to other algorithms of sales prediction. [2]

"Predicting Sales of BigMart Retail Stores using Data Mining Techniques" by Urmila Shrawankar and Usha Mehta expresses the uses of data mining techniques to predict the sales for BigMart stores. This paper describes the use of a combination of linear regression, decision trees, and neural networks to build their model. The creator achieves an accuracy of 72.8%, with the decision tree model performing the best. [3]

The paper "Big Data Analytics for Sales Prediction in Retail Industry" by Eman AbuKhousa and Ahmad Al-Azab proposes a sales prediction model for the retail industry using big data analytics. Here in this article, the BigMart dataset is used to train their model which includes features such as store size, item visibility, and item type. They use a regression-based approach and achieve an accuracy of 72.1%. [4]

The study by Pradeep Kumar in the paper "Predicting BigMart Store Sales using Machine Learning Techniques" explores the use of machine learning techniques such as decision

5

trees, random forests, and gradient boosting to predict sales for BigMart stores. The writer achieves an accuracy of 79.2% with the random forest model using the BigMart dataset. [5]

"Predicting the Sales for BigMart using Time Series Analysis" by Sridhar Mahadevan and R. Kalaikumaran paper explores the use of time series analysis techniques, specifically the ARIMA i.e., autoregressive integrated moving average model to predict sales for BigMart stores. The writer of this paper also uses the BigMart dataset and achieves an accuracy of 73.8%. [6]

The study of "Sales Forecasting for BigMart: A Comparative Study of Machine Learning Techniques" by M.Akhil and K.S. Eswari compares the performance of several machine learning techniques, including decision trees, random forests, and support vector regression, for sales prediction in BigMart stores. In this paper, it was found that the random forest model outperforms the other techniques, with an accuracy of 74.3%. Moreover, this paper also proposes a hybrid approach that combines multiple regression analysis with fuzzy logic for sales prediction in BigMart stores. The accuracy of 76.8% was obtained using the same BigMart dataset that generated the 74.3% output.

Reviewing the different papers, these studies demonstrate that different techniques and algorithms of machine learning approaches can be used for sales prediction in BigMart stores including time series analysis and hybrid approaches. Random forest models tend to perform well, but other techniques can also be effective depending on the specific dataset and context. Different factors such as store size, location, item availability, and its types are commonly used as a basis of a model train for observing the sales prediction.

CHAPTER 3
SYSTEM ANALYSIS AND DESIGN

3.1 System Analysis

This project is designed for developing prediction system mainly for BigMart providing the future sales considering different So, in the process of developing machine learning model and training as well as testing system for predicting the accuracy result, different steps were taken as consideration. Among the different step, the major step was of system analysis. Different functional and non-functional requirement gathering and feasibility study was carried out as a part of system analysis. Similarly, object modelling, dynamic modelling and process modelling for analyzing as well as component and deployment diagram for system design are drawn and develop it according to the requirement.

3.1.1 Requirement Analysis

Requirement analysis described what the system should perform and what the expectation of the users to the new system. It includes the identification of end users' need and preparation of a corresponding document term as software requirement specification (SRS). Before starting the development of the new system, the following requirement are taken into consideration.

i. Functional Requirement

Table 3.1: Functional Requirement of BigMart Sales Prediction System

S.No.	Requirement	Description
1	Login Section	The sales predictor should be able to log in to this system after matching the credential
2	Load the sales data set	Predictor should be able to load the previous sales data to analyze the trend pattern

3	Preview the data set	Predictor should be able to preview and filter the loaded data set
4	Analyze Previous Sales	Predictor must be able to analyze the previous sales in the form of chart and other graphical representation
5	Prediction Section	The prediction section should contain a value of different varies to predict future sales on the basis of different algorithm
6	Add Prediction Metrices	Predictor should be able to add the prediction metrices like item weight, item quantity etc. on the basis of prediction is performed
7	Visualization Section	The visualization section should contain a future sales, accuracy and mean square of prediction
8	Change Password	Sales Predictor should be able to change the existing password with the new one

ii. Non-Functional Requirement

• Security:

The prediction system will be secure in nature as it used different encryption techniques for password encryption and will be designed with a parameter of at least seven characters for a password.

• User-friendly:

This system interface will be user-friendly and easy to operate as well as provide maximum optimization for smoother performance. Users with basic knowledge of the internet can use this system. The system uses a simplified interface and simple content language to improve user-friendliness.

• Easy Access:

As this system will be deployed over the web and can be accessed through the internet, company administrators with datasets can easily use this system.

- Maintainability:

From the developer's perception, this system will be easy to maintain as the code will be written in a clean and readable manner for future maintenance.

3.1.2 Feasibility Study

For the feasibility analysis, the study of technical, operational, and economic

i. Technical Feasibility:

This system is technically feasible as the technical resources available to implement ideas or propose systems to real software is easily accessible. For this system development, the hardware and software was within the current technology as the resources for software development was satisfied through open-sources and free tools and the hardware requireds for coding and deployment was also present and have not issue with hardware infrastructure.

Operational Feasibility:

This system is feasible for the operation as it is easily accessible due to the users' friendly interface and the basic knowledge of the web will be sufficient to access it. In addition, this system is capable of performing all the operations of a functional requirement that the system should do and includes all the requirements needed.
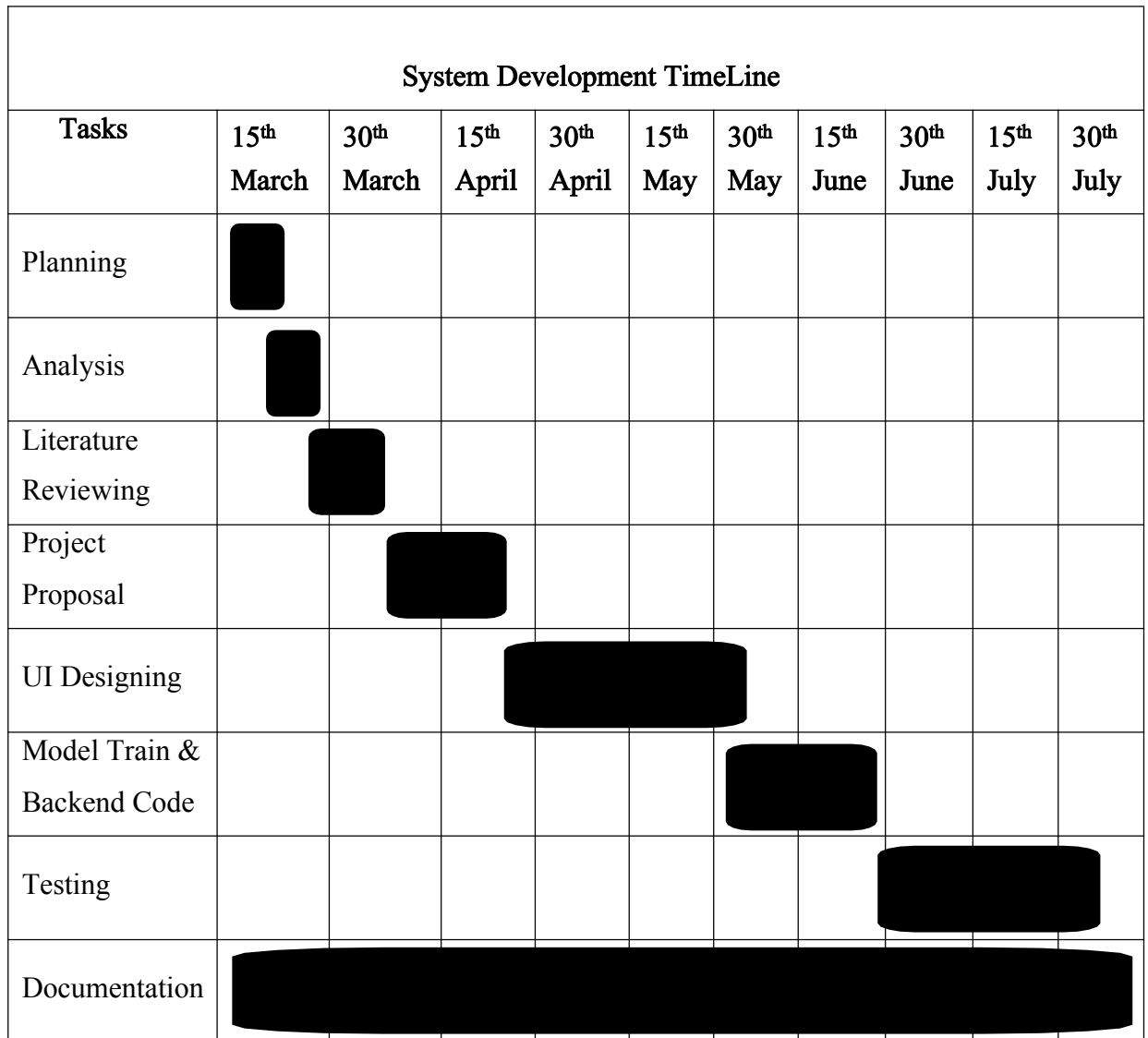
Economic Feasibility:

From the developer's perception, we use open-source development tools and didnot find any economic difficulties in the project. Similarly, the cost of development is within the determined budget as expected. So, the system can be term as economically feasible and cost-effective in nature.
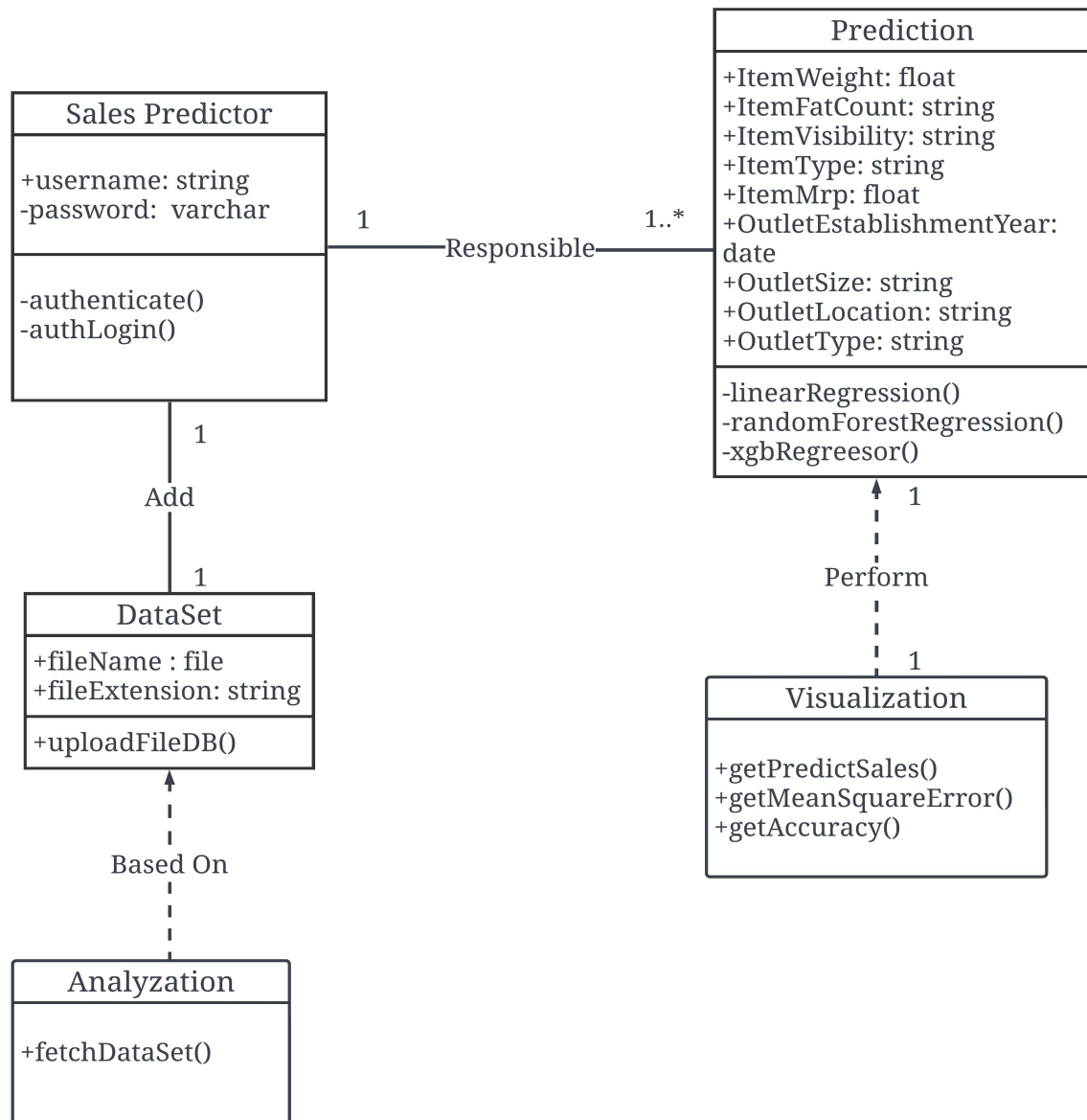
Schedule Feasibility:

On the basis of the prior experience, this project was completed within 4.5 month time phase. This below show the time needed for software development project and concluded as software have no issued and is timely feasibility in essence.

Table 3.2: Gantt Chart for BigMart Sales Prediction System

System Development TimeLine										
Tasks	15th March	30th March	15th April	30th April	15th May	30th May	15th June	30th June	15th July	30th July
Planning	██									
Analysis	██									
Literature Reviewing		██								
Project Proposal			██							
UI Designing				████	████					
Model Train & Backend Code						██				
Testing								████	████	
Documentation	████	████	████	████	████	████	████	████	████	████

3.1.3 Object Modelling

Figure 3.1: Object Modelling for BigMart Sales Prediction

The figure 3.1 shows the object modelling which define the static structure of an object - oriented model using class diagram. The class diagram is represented by a rectangle shape commonly compartment into three section through horizontal line. The first section represent the class name, attribute and methods are defined into second and bottom section respectively. Here, in the above figure Sales Predictor, Prediction, Dataset, Visualization and Analyzation are class. Sales Predictor is responsible for prediction of sales and the visualization is depend upon prediction class.
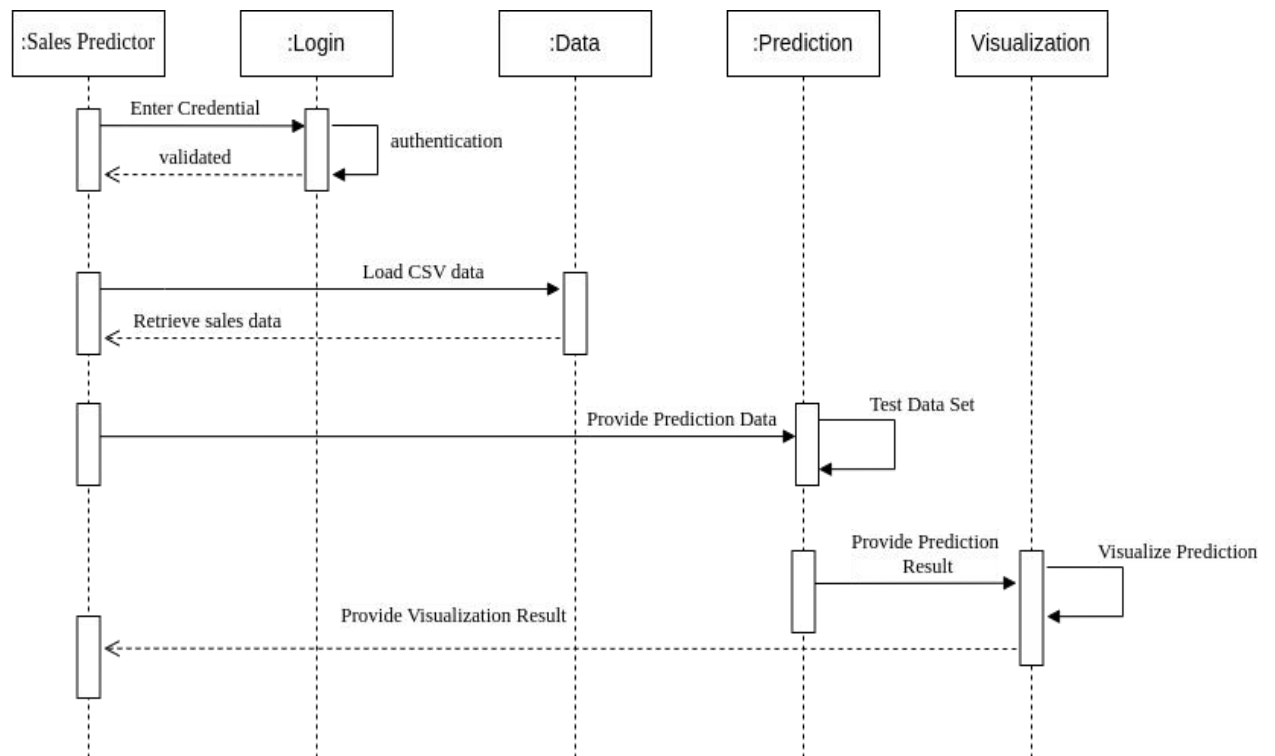
3.1.4 Dynamic Modelling

Figure 3.2: Dynamic Modelling for BigMart Sales Prediction

The figure 3.2 shows the dynamic modelling using sequences diagram between different class. The sequence starts with the initial state of sales predictor.
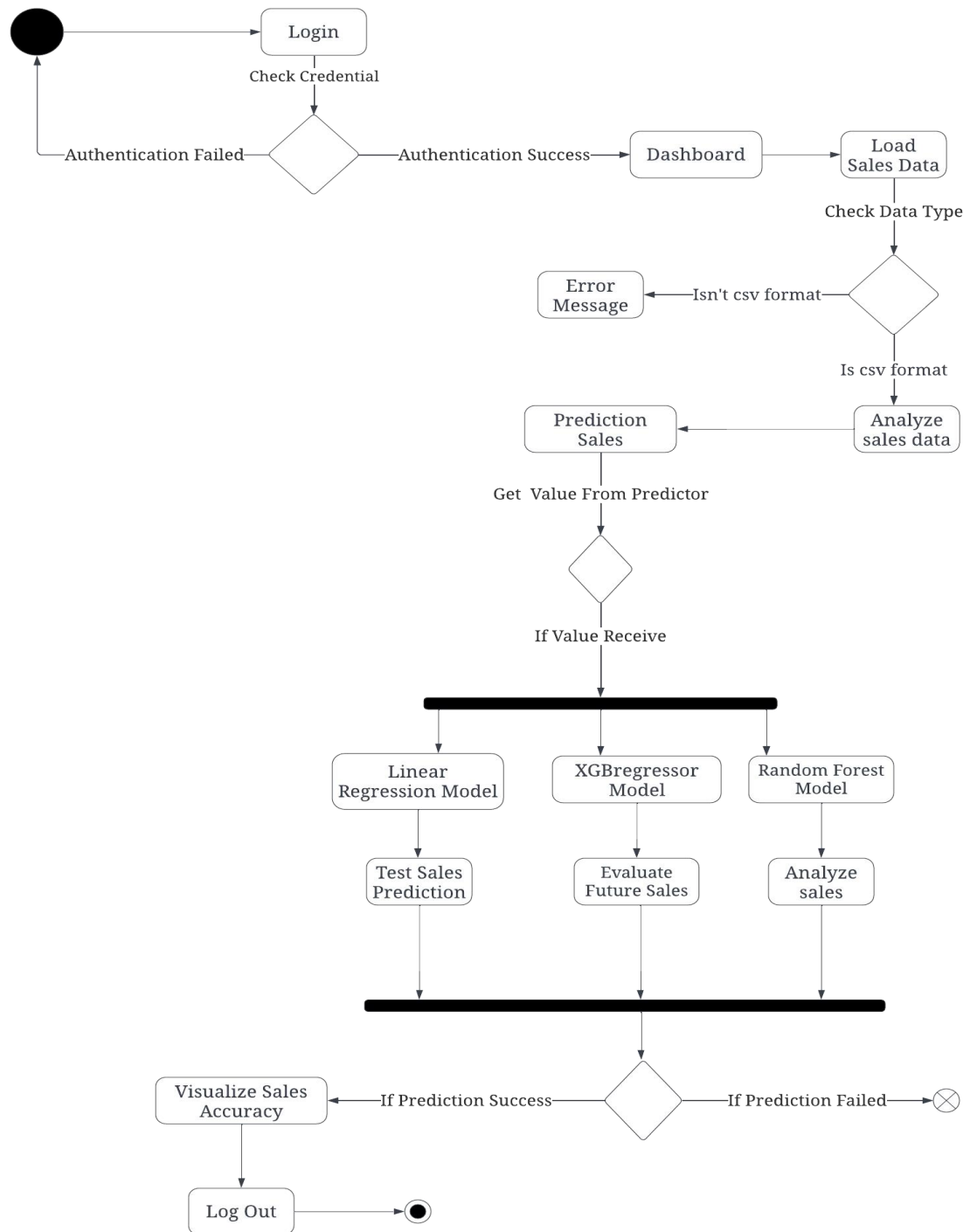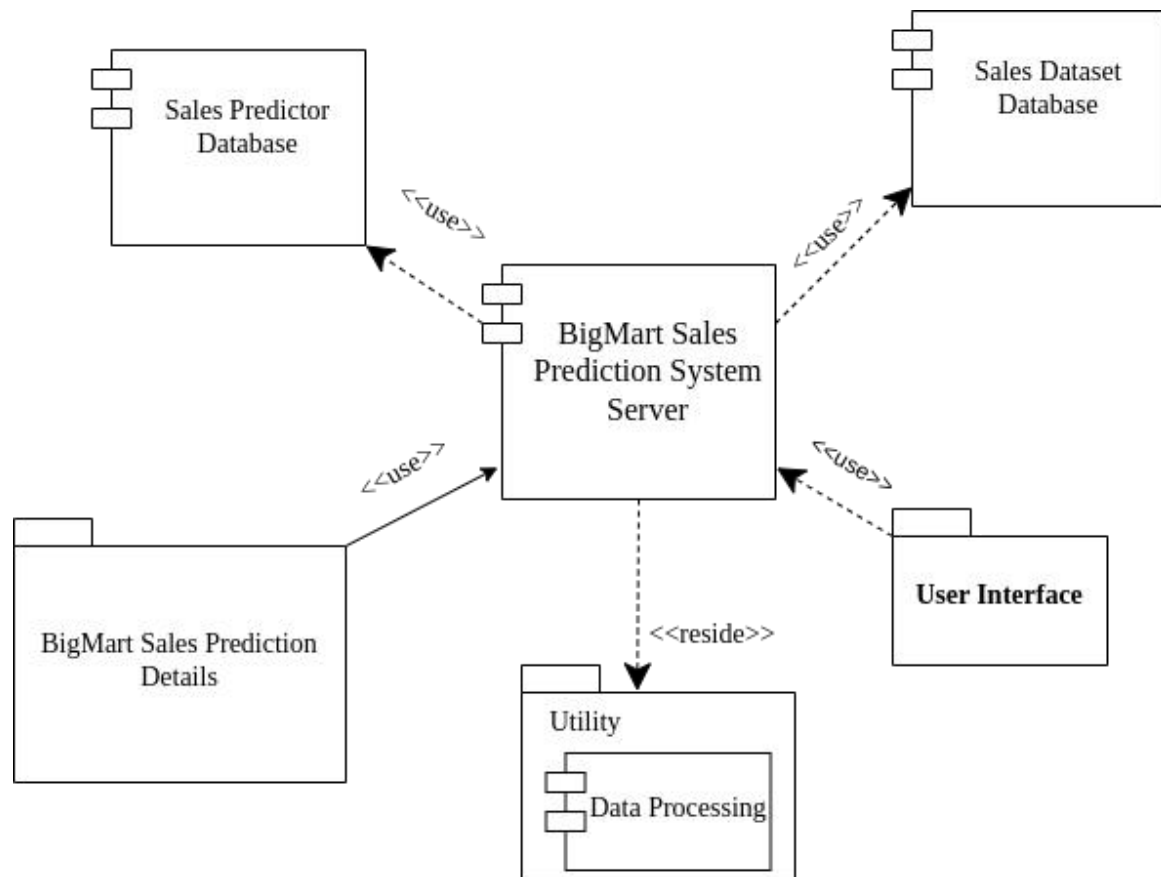
3.1.5 Process Modelling

Figure 3.3: Process Modelling of BigMart Sales Prediction

3.2 System Design

3.2.1 Component Diagram

Figure 3.4: Component Diagram for BigMart Sales Prediction

The figure 3.4 shows the component diagram for BigMart Sales Prediction.Here, for this system, the main component is defined asthe bigmart sales prediction system server. Beside it, data processing, sales predictor database and sales dataset database are also component for this software. In this system, user interface uses the main component to render user interface. Likewise, the bigmart sales prediction details also use the main component i.e., bigmart system server for all the prediction related activities. Moreover, the main component is reside upon the data processing component for all the processing action. Furthermore, the main component use sales prediction database and sales dataset database for their respective actions.

3.2.2 Deployment Diagrams
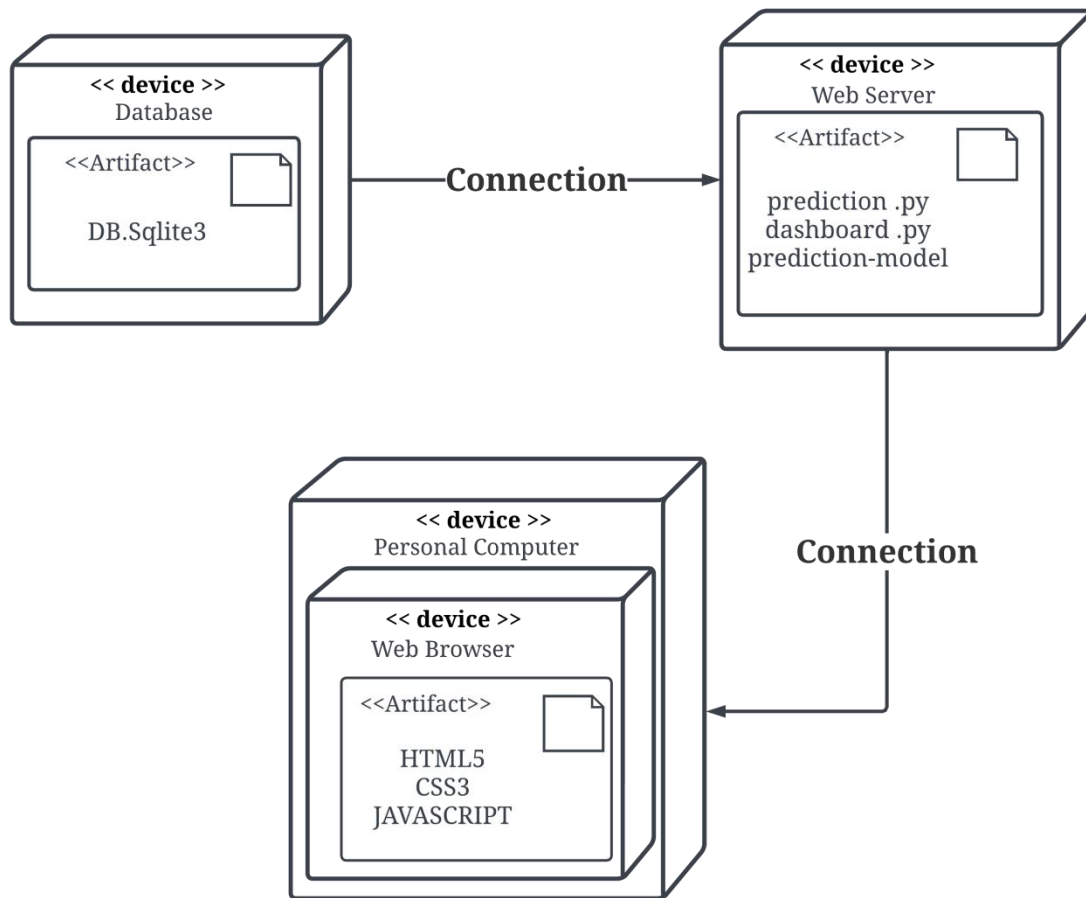
Figure 3.5: Deployment Diagram for BigMart Sales Prediction

The figure 3.5 shows the deployment diagram for bigmart sales prediction. For the deployment of system, mainly the node and artifact are used. So, in the deployment of bigmart sales prediction system, three different nodes have been adopted named as database, webserver and personal computer. Inadditional to this, the personal computer node have one nested node term as web browser. Different artifacts are consist inside the node as per the nature of it. So, for the bigmart sales prediction the artifact db.sqlite3 is locate inside the database, prediction.py and other backend as well as prediction model related artifacts is present in center of web-server node. Moreover, the Html5, CSS and JavaScript artifacts are within the web browser node.

3.3 Algorithm

For this prediction system, the main target was to implement the algorithm that predicts future sales based on variables and metrices that pre-exist in the dataset. So, one of the algorithms that we have been implemented was linear regression which is a statistical method used to train and model the relationship between a dependent variable and one or more independent variables. This algorithm was one of the best-fit line to achieve accuracy result for this type of system as it used to model the relationship between the total sales and the independent variables such as item price, item weight, outlet location, etc. by finding the mean square error

In mathematically form, linear regression can be expressed

as: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p + \varepsilon$

where,

y stands for dependent variables

$x_1, x_2, ..., x_p$ stands for the independent

variables. $\beta_0, \beta_1, ..., \beta_p$ stands for the

coefficient

Lets understand the working principle of linear regression algorithm with the below dataset. Assuming the data under different dependent variables and independent variable from past few year.

Item Weight	Item Quantity	Item Visibility	Item Fat Count	Item Sales Count
0.8	50	2	1	30000
0.11	70	1	0	30000
0.17	30	2	2	30000

Firstly, the algorithm split the variable metrices into independent and dependent variable. Here, item sales count is split as independent variable whereas other variable are term as dependent variable. Now, the variable are divided into train and test module term as X train, X test, Y train, Y test for implementing in the algorithm model

Another algorithm that is implemented in the system is Random Forest Regression which is an ensemble method and a common statistical learning method that uses the bootstrapping method to extract multiple samples from the original sample forecasting and combine the decision tree to perform them. RFR takes the meaning of the test predictions. Random forest regression is one of the additive model that predicted from a sequence of base models by combining decisions.

Mathematically, random forest regression can be

expressed as: $g(x) = f_0(x) + f_1(x) + f_2(x)$

$+ \ldots$

To make it simple, Random Forest Regression builds multiple decisions trees and fuses them together to get a forecast that is more reliable and stable.

REFERENCES

[1] K. S. Nikita Malik, "Sales Prediction Model of Big Mart," *Research Gate,* 2020.

[2] D. Choudhary, "Sales Prediction using Machine Learning," 2017.

[3] U. M. Urmila Shrawankar, "Predicting Sales of BigMart Retail Stores using Data Mining," 2018.

[4] E. A. a. A. Al-Azab, "Big Data Analytics for Sales Prediction in Retail Industry," 2015.

[5] P. Kumar, "Predicting BigMart Store Sales using Machine Learning Techniques," 2017.

[6] S. M. a. R. Kalaikumaran, ""Predicting the Sales for BigMart using Time Series Analysis"," 2020.