A Project Report on

E-commerce Based Sales Prediction Framework

Submitted in partial fulfillment of the requirements
of the degree of

Bachelor of Engineering

in

Information Technology

by

Tejal Tandel(16104047)
Sayali Wagal(17204016)
Nisha Singh(17204011)

Under the Guidance of

Name of Guide:Ms.Rujata Chaudhari
Name of Co-Guide:Mr.Vishal Badgujar

Department of Information Technology
A.P. Shah Institute of Technology
G.B.Road,Kasarvadavli, Thane(W), Mumbai-400615
UNIVERSITY OF MUMBAI
Academic Year 2019-2020

Approval Sheet

This Project Report entitled ***"E-commerce Based Sales Prediction Framework"*** Submitted by ***"Tejal Tandel" (16104047),"Sayali Wagal" (17204016),"Nisha Singh" (17204011)***is approved for the partial fulfillment of the requirenment for the award of the degree of ***Bachelor of Engineering*** in ***Information Technology*** from ***University of Mumbai***.

Mr. Vishal Badgujar
Co-Guide

Mrs. Rujata Chaudhari
Guide

Mr. Kiran Deshpande
Head Department of Information Technology

Place:A.P.Shah Institute of Technology, Thane
Date:

CERTIFICATE

This is to certify that the project entitled ***"E-commerce Based Sales Prediction Framework"*** submitted by ***"Ms. Tejal Tandel" (16104047),"Ms. Sayali Wagal" (17204016),"Ms. Nisha Singh" (17204011)*** for the partial fulfillment of the requirement for award of a degree ***Bachelor of Engineering*** in ***Information Technology***,to the University of Mumbai,is a bonafide work carried out during academic year 2019-2020 .

Mr. Vishal Badgujar
Co-Guide

Mrs. Rujata Chaudhari
Guide

Mr. Kiran Deshpande
Head Department of Information Technology

Dr. Uttam D.Kolekar
Principal

External Examiner(s)

1.

2.

Place:A.P.Shah Institute of Technology, Thane
Date:

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, We have adequately cited and referenced the original sources. We also declare that We have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We 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 has not been taken when needed.

_____

(Tejal Tandel and 16104047)

_____

(Sayali Wagal and 17204016)

_____

(Nisha Singh and 17204011)

Date:

Abstract

The retail sector has widely adapted different inventory management applications and some retail chains even employ prediction software to analyze future sales. However, a lot of day-to-day shopping in India happens through local shops. The owners of such mom-and-pop shops do not necessarily have the capital to invest in proprietary applications for setting up an inventory management system. Needless to say that same is the case for any sales prediction software. As a result, many of the shopkeepers end up hoarding a lot of irrelevant and nonprofitable products that lead to financial losses. A very cost-effective and accessible solution for this problem is a Website that provides all the features of a pointof-sale system as well as gives future sales insights.

It will enable shopkeepers to manage their current product purchases and invoicing. The predictive sales analysis will help them to modify their investments on products and supplies thereby ensuring maximum profits. If a shop houses relevant products that cater to customer needs, its customer reach will increase. The Economic Times published an article in the May of 2019, which stated that the number of online platform users in India is expected to rise by 84 to 859 million by 2022 from 468 million in 2017. It is safe to assume that a large population of shop owners will have online platform in the following years. Hence, equipping the local shopkeepers with a mobile application will prove instrumental since it will give them exposure to all the aforementioned benefits.

Contents

Chapter 1

Introduction

1.1 Introduction

The e-commerce trend has taken businesses online and has proved to be beneficial for them. In a basic e-commerce system,the merchants put their products on display over the website and the customers searching for that product can place an order using the same website. The financial transaction is done by payment portals and then a delivery service delivers the products to the designated customer. E-commerce websites hold the potential to showcase a wide variety of products at once and therefore is equally convenient to buyers and sellers. The sellers can generate a report of their sales or product demands either manually or through a data mining software. Even with so much backbone in the e-commerce industry, many shopkeepers in India have chose to stay completely offline. This scenario raises a need to understand the reason for the same. A major contributing factor for local shopkeepers to not take their businesses online is the lack of monetary funds and resources. It may seem a personal drawback from afar but if we aim at achieving social development this problem needs to be addressed and a viable solution must be found. Accessing a point-of-sale system via their own Website is definitely one of the most feasible solution to this problem. Website platform are free to download and can provide an equally good user interface as that of a computer based inventory management system.

A seller can manually list down all the products and investment and tally it with to total sales to produce a profit report but that is not to say tedious and monotonous. Through data mining techniques the same results can be achieved more quickly and one can even get a graphical representation for better understand makes the process more engaging. Traditionally to perform any product based analysis, a different software is needed to be purchased. A website with an amalgamation of both these trends will make the whole process more convenient. The website should comprise all of the following functionalities for catering to the need of the shopkeepers under consideration:

1. Add/ Update/ Delete Category of products.
2. Add/ Update/ Delete Products in every category.
3. Display existing inventory list.
4. Barcode scanning for incoming products and invoice via phone camera.
5. Sales report and predictions.

1.2 Objective

The main objective of our project is to develop a easy-to-use web based application for local shopkeepers that helps them improve their finances by providing them a predictive analysis of their sales. Some other objectives of the project are:

- To provide a platform which is budget friendly.

  The entire system is developed using HTML5, materialized CSS, JavaScript, PHP and Python. The use of said languages makes the web frame extremely compatible. The local shopkeepers need not buy high-end computer systems or additional hardware in order to use this solution. Cost-cutting enables the owners to avail modern technologies like inventory management and sales prediction without straining their resources.

- To help retailers stock appropriate products based on the in-demand product sale.

  Tabular reports are generated based on the buying-selling transactions. Data fetched from the databases is organized to give insights about the best selling category of products and the items of those categories respectively. This tabular representation will make it easy to interpret which products add to the shop's profit and which do not. Therefore the task of planning the appropriate inventory is simplified.

- To generate a report on expected monthly sales so that retailers can analyse their expenses.

  By employing a predictive data mining model, a sales forecast report is generated in graphical format. The number of products sold and the summation of their respective selling price are given as attributes and results are obtained using regression analysis. Any shop owner can grasp the variations in graph and analyze the future sales of products without having to pursue over the top analytical solutions.

1.3 Problem Definition

A retailer can participate in general buying and selling on traditional e-commerce systems but they do not get customized sales reports for their products.It is beneficial for the sellers if they are presented with some statistical reports about their future sales so that they can plan their business accordingly. Keeping the business offline is a drawback as it reduces the customer reach. Recommendation of potential products is not given to the retailers.

Many of the businesses face losses because their customer reach is limited. There can be conditions where proper inventory planning is not achieved because lack of understanding of product demand. This in turn results into a loss for the business owner. The local retailers cannot do investment and management activities beforehand because they do not have the statistics for it. Improper product planning can result in decrease in the revenue. There is no dedicated model for analysing the comments or feedback from the users. Lack in transparency between the buyer and seller can hinder the sales for that organization

1.4 Technology Stack

FRONTEND:

- HTML5, CSS3, materialized CSS
- BootStrap
- JavaScript, jQuery
- PHP (server side scripting)

BACKEND:

- PHP (user side scripting)
- Python3, Flask
- MySQL (database)

ALGORITHM:

- Linear Regression for Predictive Analysis

Chapter 2

Literature Review

The papers referred to while developing the system are mentioned below.

Paper Title 1: Performance prediction using modified clustering techniques with fuzzy association rule mining approach for retail.

Authors: C. Ezhilarasan and S. Ramani

Published in: 2017 International Conference on Intelligent Computing and Control (I2C2), Coimbatore, 2017.

In this paper, the performance of a hosted website is predicted using data mining techniques. The website traffic and conversion rates are considered as the attributes for the data mining logic. Website traffic is the number of users who visit that website. Web traffic is measured in "visits" and is a common way to measure the effectiveness of the website and in-turn the business, in gathering a customer base. Conversion rate is the percentage of a website's visitors out of the total visitors that have completed a desired goal for that online business. Data analysis can be used to design extraction models which define future data trends. The technique used in this paper for data mining and prediction reports is fuzzy logic. Fuzzy logic is used when the outcome is uncertain. In cases where it cannot be determined whether a condition is true or false, fuzzy logic is used as it provides flexibility in logical reasoning. This makes it efficient for predictive analysis in particular clusters. The proposed method for this project is web hosting. Since, it targets web traffic and conversion rates, it can be used in any domain. Web hosting makes it a real-time environment. Another advantage is that fuzzy logic can work with any kind of input i.e. even noisy or distorted can be given as input and yet the result accuracy 8 will not be hampered. The drawback for this approach is that fuzzy requires large data to compute the uncertainties. Therefore, small amount of data cannot be processed and hence the initial training results can be inaccurate. Precise conversion rates are required every time during processing the attributes otherwise the results generated can be inappropriate. There is no systematic approach to solve a problem statement using fuzzy which causes ambiguity during if the product is distributed across branches.

Paper Title 2: Demographic transformation and clustering of transactional data for sales prediction of convenience stores.

Authors: Xiaojun Zhang, Jisheng Pei and Xiaojun Ye

2016 IEEE International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), Chengdu, 2016. Demographics are statistical representation of the characteristics of a population. These are mostly the socio-economic features of an individual. The age, educational level, occupation, income, marital status, average size of family etc are all considered as demographics. In terms of a website, these demographics are extracted from the visitors. In data analysis process, these demographics are categorised into groups and are mapped with their frequent activities on the website.

The paper follows a technique where the demographic clusters are combined with corresponding transactional data clusters to generate an input for the data prediction model. The data mining task is done using k-means algorithm. K-means algorithm categorises input itemset into "k" number of clusters based on their similarity. The similarities are calculated using Euclidean distance method. It is an effective algorithm where large input itemsets are available. Using k-means is beneficial as it is easily implemented and forms firmer clusters than hierarchical clustering. Moreover, if the number of clusters is small, the model computes faster. With the proposed model, association rules can be integrated easily and can be used to give product recommendation. The predictive analysis approach used in this paper was centred around broader categories. Therefore product wise analysis cannot be done. The whole approach considers predictive analysis of main categories for the overall outcome. This can be inefficient as sales for a business cannot be analysed by only considering the categories of items.

Paper Title 3: Profit Prediction Using Regression Model for Travel Agents.

Authors: R. P. Santi and M. L. Khodra

Publishedin: 2018 International Workshop on Big Data and Information Security (IWBIS), Jakarta, 2018. In this paper, profit for a tour agency which predominantly focuses on air travel is predicted. Air travel has become a sort of some frenzy in the recent years. Therefore many travel agencies have focused their business activity on air travel to increase transactions and corporate profit. Online ticket sales are proportional to the profits earned for any agency. To analyse this, ticket fare is clustered into groups based on the price range for various different airlines. Historic data from airlines can be used to extract ticket fare and other profit affecting factors, for the trading model. and linear regression is used to generate the results. Since there is a proportionality in the attributes and outcome, regression technique is used for the data mining task. Linear regression constructs statistical model considering dependent and independent data items. Ticket fare clusters are given as input item set to a prediction model which uses linear regression to generate the results. The inputs are independent data sets and the result of regression equation are dependent data sets. Through these reports, the travel agents can analyse their return-of-investment

and plan packages accordingly.

This gives the cost estimation to travel agent and also describes the value for target completion. Linear regression provides less space complexity as it saves data weights at the end of training. Through this technique the agents can handle feature selection.

The problem with this system arises if the data is not normalised. The linear regression model presumes that data inputs are normally distributed and thus if such is not the case, the output obtained is inaccurate and ineffective. Moreover, the method used in this paper takes only ticket fares as a viable input data set.

The profit prediction for an agency cannot be determined as efficiently if only one factor is taken into consideration. This model is also prone to over fitting of data if the independent item set are not linear.

Chapter 3

Proposed System Architecture/Working

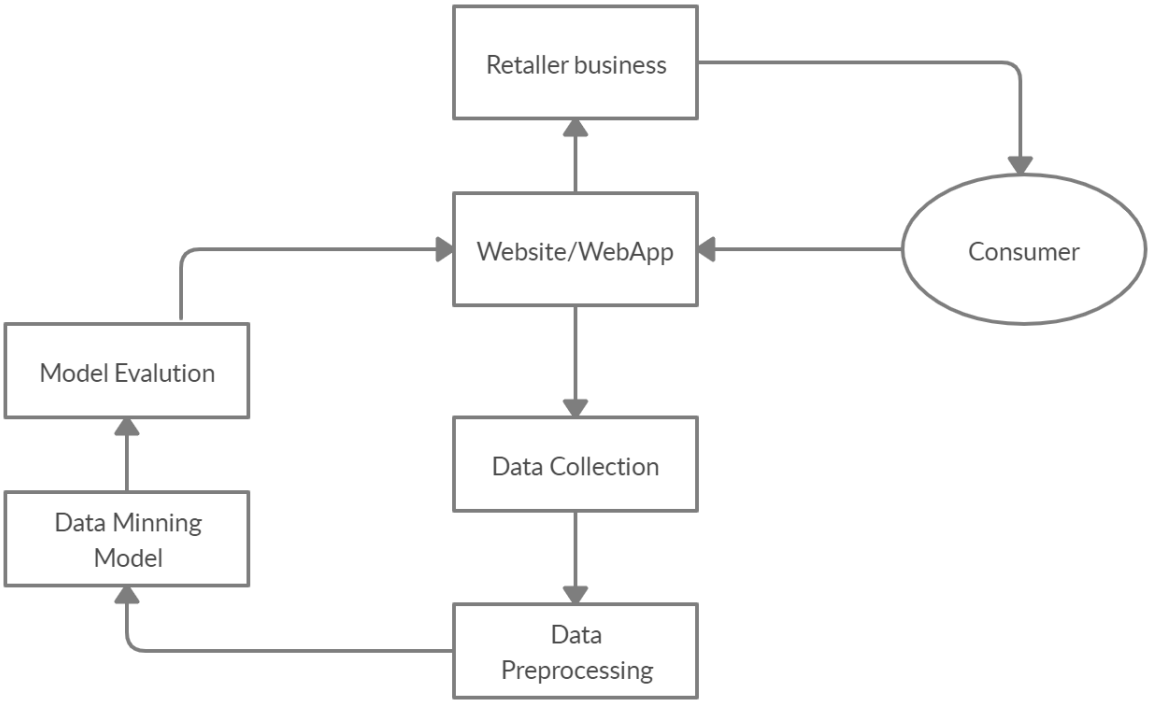


Figure 3.1: Proposed system

PROPOSED SYSTEM ARCHITECTURE:

The diagram above represents the chain of activities that the proposed system is based on. The central element of this system is a website which serves inventory management features like product listing and invoice generation. The data obtained from all the transactions between the seller and the consumer is stored in a database and is then a segment of it is fed into a data preprocessing loop. After this data is cleaned, it is given as input for the predictive data mining model. First a training model is initiated for the predictive analysis. The

model uses linear regression for the same. During the training phase regression constants are generated and the errors are resolved. Once all the desired parameter are obtained, the model is evaluated before all the real time data is fed into the final model. The data mining model keeps learning as more and more data is received and the accuracy improves. The results will then be fetched into the website in graphical format for users to analyse and understand.

EXISTING SYSTEM AND ITS DRAWBACKS:

There are already existing and fully functional e-commerce and inventory management websites available to the users. They effectively provide provide list of products, assist in easier bill generation and calculations and can be modified as per user's need.
However, there is no provision for readily available sales report or future sales analysis. If business owners want to avail this facility, they have to purchase a third-party software or a licensed product to do so securely. While big companies and businesses can afford this solution, many of the local shopkeepers cannot due to limitation on capital investment.
In addition to this certain existing websites or software require special configuration or hardware to operate smoothly. Local vendors refrain from such upgrading and therefore are at loss from availing sales prediction features.

PROPOSED SYSTEM WORKING:

The proposed system is focused on overcoming the drawbacks of existing system. To bring about these functionalities, data mining model is integrated with the inventory management. The modules at the entry point of the system are that of registration and login. Data abstraction and isolation as well as session control are done through these. The user interface then has modules that will help the shop owners to maintain the inventory records, update product details and generate customer bills. The invoice generation is an automatic loop that fetches existing values of parameters like selling price, discount (if any) and calculates the overall amount.
The database is updated dynamically and a python loop fetches data from the same for the preprocessing task. Once the data is cleaned, the data mining task is initiated.

DATA MINING ALGORITHM: LINEAR REGRESSION

Linear regression is a basic and commonly used type of predictive analysis. The overall idea of regression is to examine two things: (in our case) the summation of selling price and the summation of quantities sold and in what way do they impact the outcome variable. These regression estimates are used to explain the relationship between one dependent variable and one or more independent variables. The simplest form of the regression equation with one dependent and one independent variable is defined by the formula y = a + bx, where y = estimated dependent variable score, a = constant, b = regression coefficient, and x = score on the independent variable.
In order to get the values of "a" and "b" a test run is made by using a fragment of data set in the database and passing it through a test model. The sklearn.model.selection library has a package named train.test.split which is used to find the values for said constants. The outcome of that model looks as follows:

```
const 6.948683
electronics 0.054564
dtype: float64
```

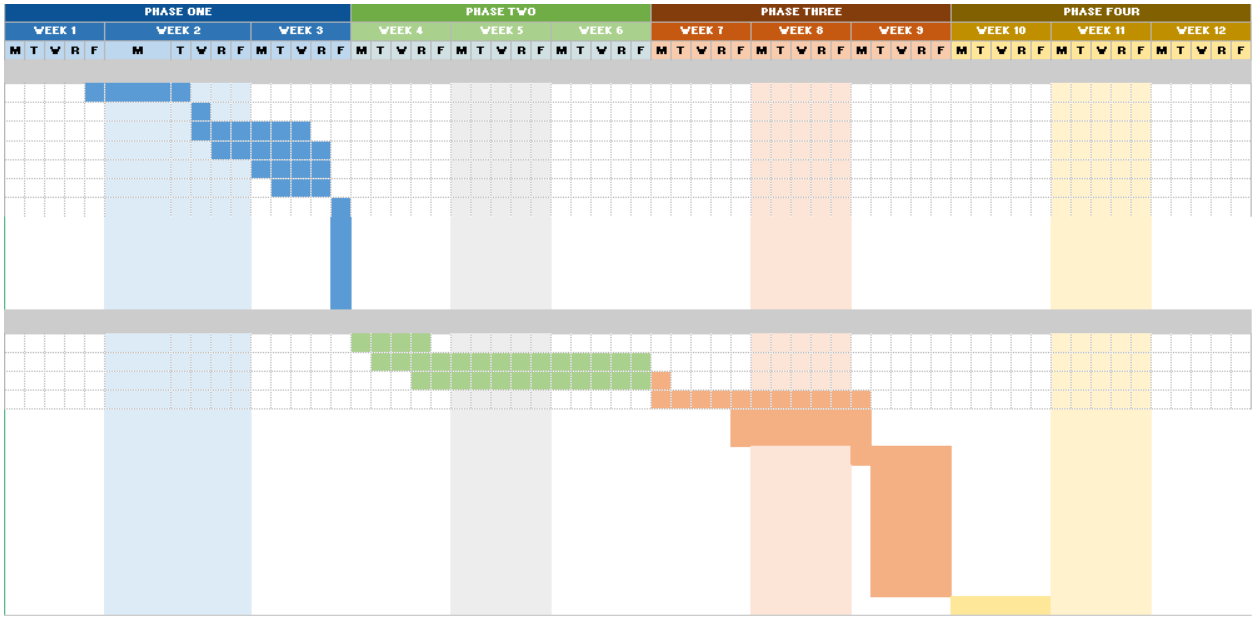
Figure 3.2: Figure 3.2: Output of training model

From the parameters we got our linear equation for the products falling under the category "electronics" becomes Sales=6.948+0.054×(total revenue generated from electronics). Similarly very independent category is processed and constants are generated. In the end all the values obtained for "Sales" varibale are plotted in a time-series graph using matplotlib library.

Poject Timeline Chart

WBS NUMBER	TASK TITLE	TASK OWNER	START DATE	DUE DATE	DURATION(Wee kr)	PCT OF TASK COMPLETE
1	**Project Conception and Initiation**					
1.1	Research paper search	Tejal,Sayali,Nisha	7-10-19	7-26-19	3	100%
1.1.1	Research paper finalization	Tejal ,Sayali	7-10-19	7-26-19	3	100%
1.2	Project Title	Tejal,Sayali,Nisha	7-10-19	7-26-19	1	100%
1.3	Abstract	Nisha	8-23-19	8-30-19	1	100%
1.4	Objectives	Nisha	8-23-19	8-30-19	1	100%
1.5	Literature Review	Sayali,Nisha	8-23-19	8-30-19	1	100%
1.6	Problem Definition	Tejal,Sayali,Nisha	3-23-18	8-30-19	1	100%
1.7	Scope	NIsha	8-23-19	8-30-19	1	100%
1.8	Technology stack	Tejal	8-23-19	8-30-19	1	100%
1.9	Benefits for environment	Nisha	8-23-19	8-30-19	1	100%
1.1	Benefits for society	Sayali	8-23-19	8-30-19	1	100%
1.11	Applications	Tejal	8-23-19	8-30-19	1	100%
2	**Project Design**					
2.1	Proposed System	Tejal,Sayali,Nisha	9-19-19	9-27-19	1	100%
2.2	Design(Flow Of Modules)	Tejal,Sayali,Nisha	9-19-19	9-27-19	1	100%
2.3	Activity Diagram	Sayali	9-19-19	9-27-19	1	100%
2.4	Use Case Diagram	Sayali	9-19-19	9-27-19	1	100%
2.5	Description Of Use Case	Sayali	9-19-19	9-27-19	1	100%
2.6	Modules	Tejal,Sayali,Nisha	9-25-19	10-1-19	1	100%
2.6.1	Module-1	Nisha	9-25-19	10-1-19	1	100%
2.6.2	Module-2	Nisha	9-25-19	10-1-19	1	100%
2.6.3	Module-3	Nisha	9-25-19	10-1-19	1	100%
2.6.4	Module-4	Sayali	9-25-19	10-1-19	1	100%
2.6.5	Module-5	Sayali	9-25-19	10-1-19	1	100%
2.6.6	Module-6	Sayali	10-25-19	10-1-19	1	100%
2.6.7	Module-7	Tejal	10-25-19	10-1-19	1	100%
2.6.8	Module-8	Tejal	10-25-19	10-1-19	1	100%
2.7	Preparation Of Report	Sayali,Nisha	10-5-19	10-12-19	1	100%

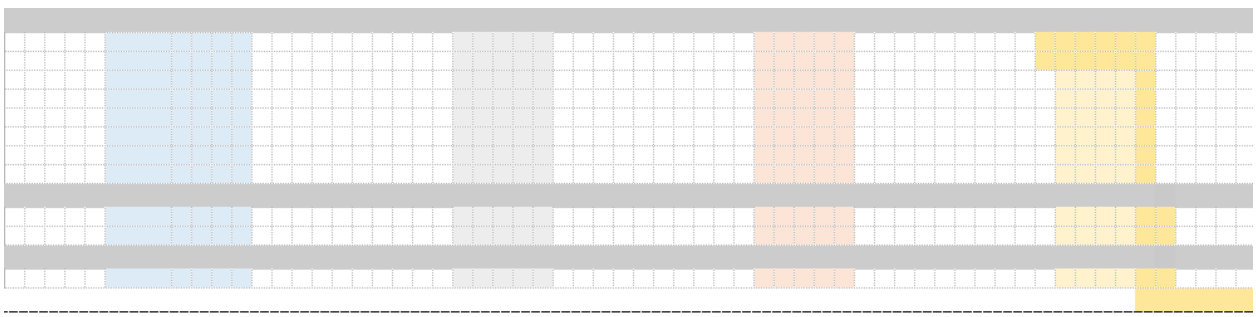
Gantt Chart 1



Gantt Chart Weekly Report 1

3	**Project Implementation**					
3.1	Module-1	Nisha	12-2-19	12-7-19	1	100%
3.2	Module-2	Nisha	12-8-19	12-13-19	1	100%
3.3	Module-3	Nisha	12-15-19	3-29-20	2	100%
4.3	Module-4	Sayali	12-5-19	12-24-19	2	90%
5.3	Module-5	Sayali	1-6-20	1-10-20	1	100%
6.3	Module-6	Sayali	1-11-20	1-16-20	1	90%
7.3	Module-7	Tejal	12-5-19	12-10-19	1	100%
8.3	Module-8	Tejal	12-14-19	1-15-20	4	80%
4	**Testing**					
4.1	Design of Test Cases	Tejal,Sayali,Nisha	1-17-20	2-5-20	4	100%
4.2	Testing	Tejal,Sayali,Nisha	2-10-20	3-3-20	3	100%
5	**Results and Analysis**					
5.1	Analysis Of Results	Sayali	3-7-20	3-21-20	2	100%
5.3	Report Preparation	Tejal,Sayali,Nisha	4-1-20	4-6-20	1	100%

Gantt Chart 2



Gantt Chart Weekly Report 2

Chapter 4

Design

4.1 Usecase Diagram

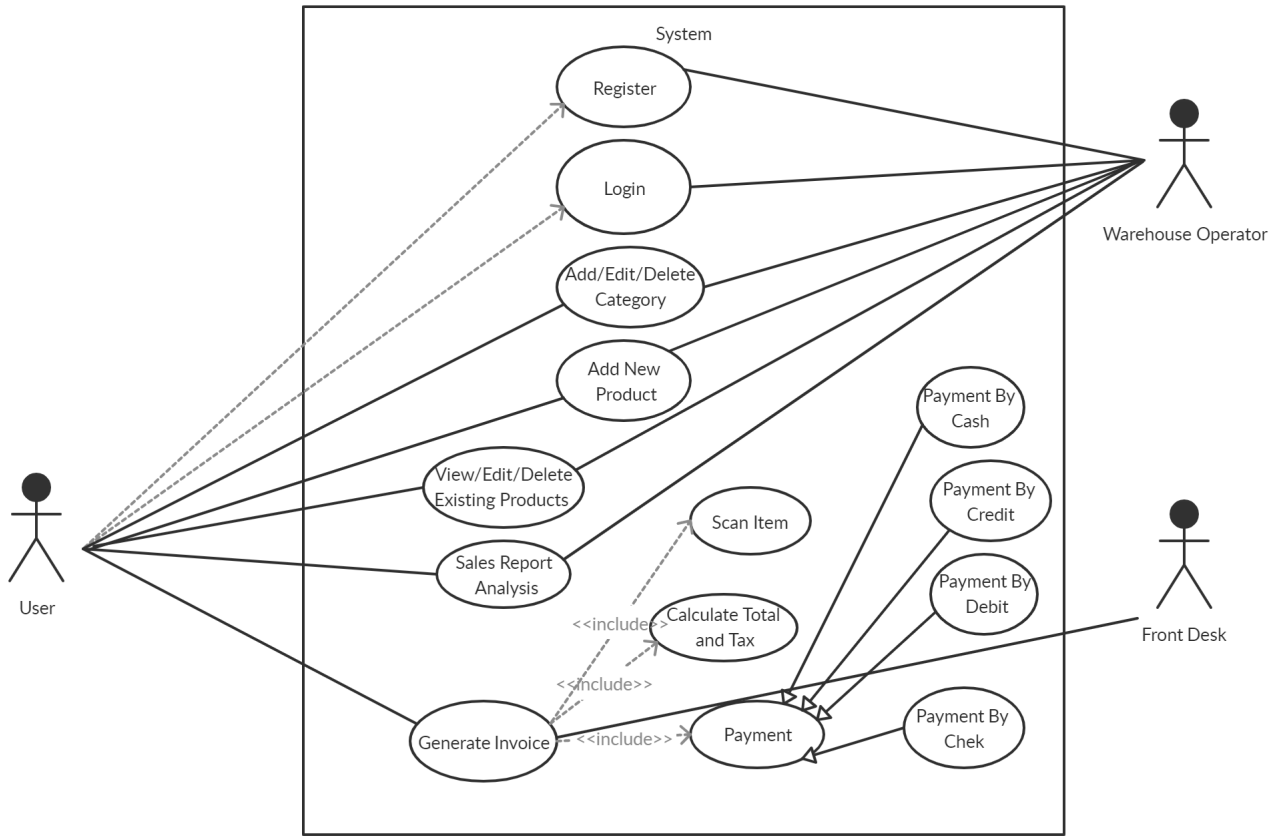


Figure 4.1: Usecase Diagram

Our system has three actors: Front Desk operator, Warehouse operator and User(Shop Owner). They can perform all the different tasks they are associated with in the diagram. The entire user system has database support. The shop owner can monitor and maintain all the tasks performed by front desk operators and warehouse operators.

4.2 Activity Diagram

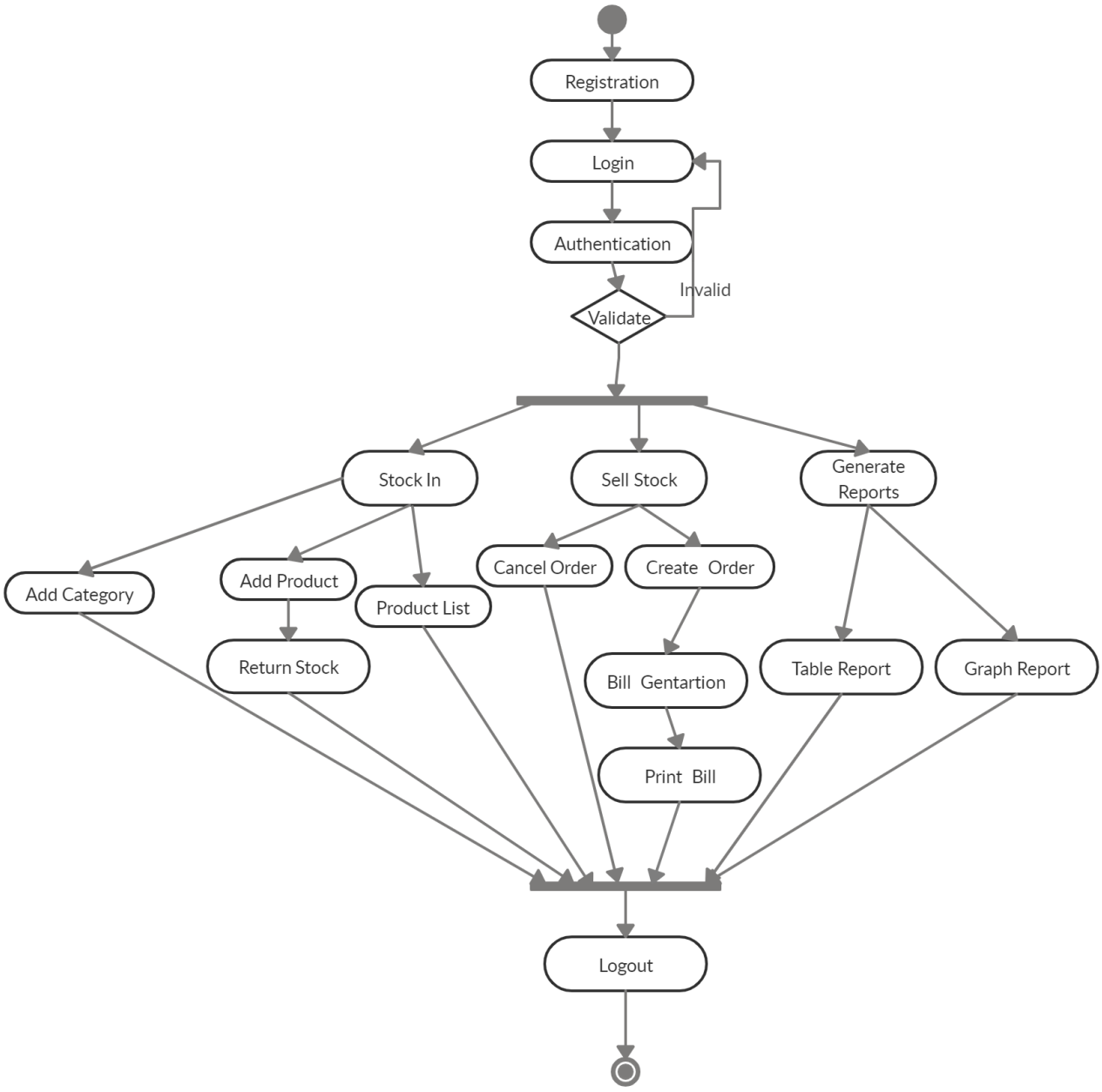


Figure 4.2: Activity Diagram

In this activity diagram, there is the basic flow of our website from like it will start from oping the website-¿data collection-¿dataprocesing-¿regrationbased analysis-¿module evolution-¿website this is the basic flow of this system

4.3 Class Diagram

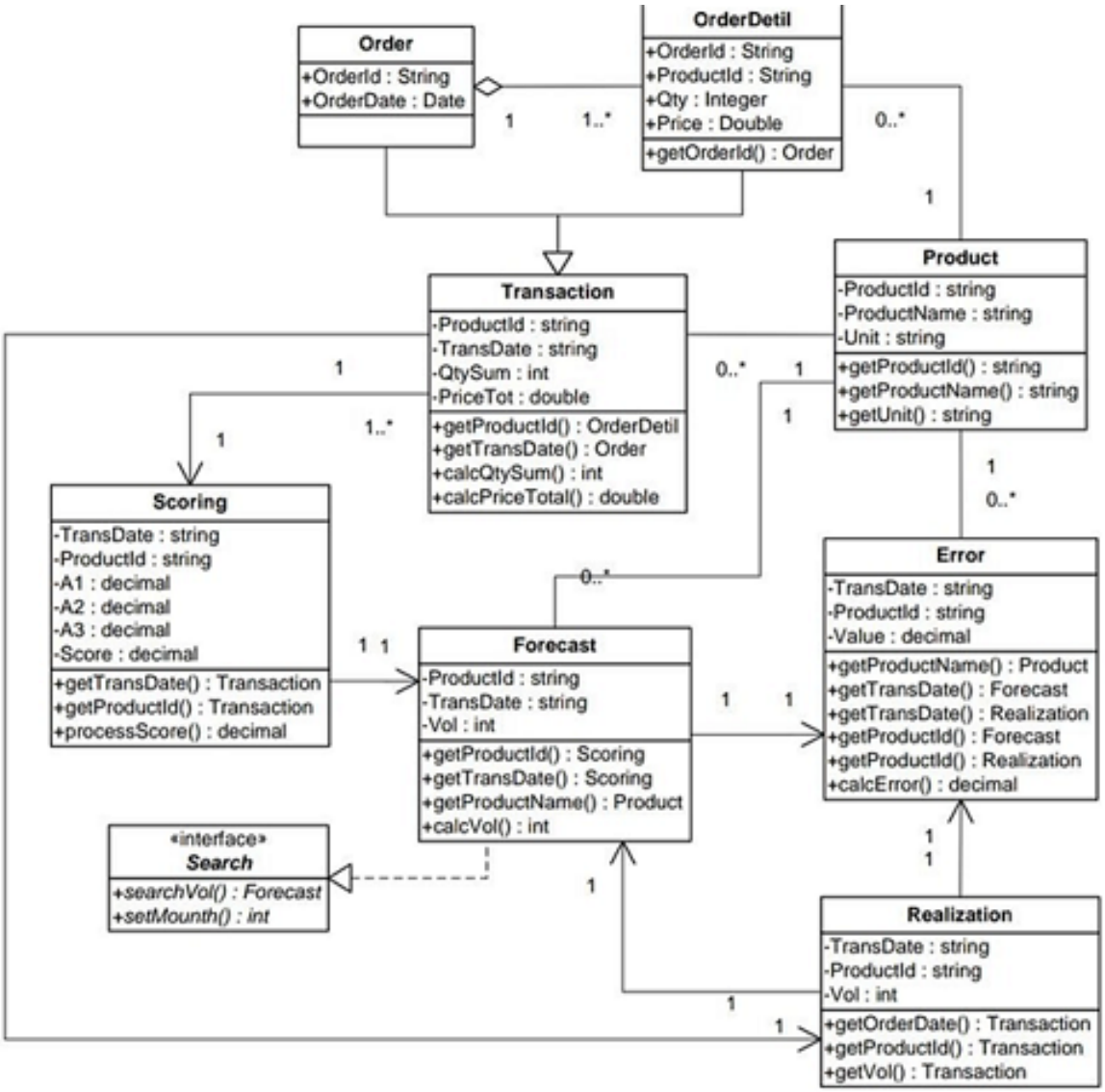


Figure 4.3: Class Diagram

The transaction class is a central class through which data is updated in the database and then used by other classes. The forecast class relies on the data modified by the product class. Every class more or less relies on some shared attributes as seen in the figure since all of them target a single database.

4.4 Sequence Diagram

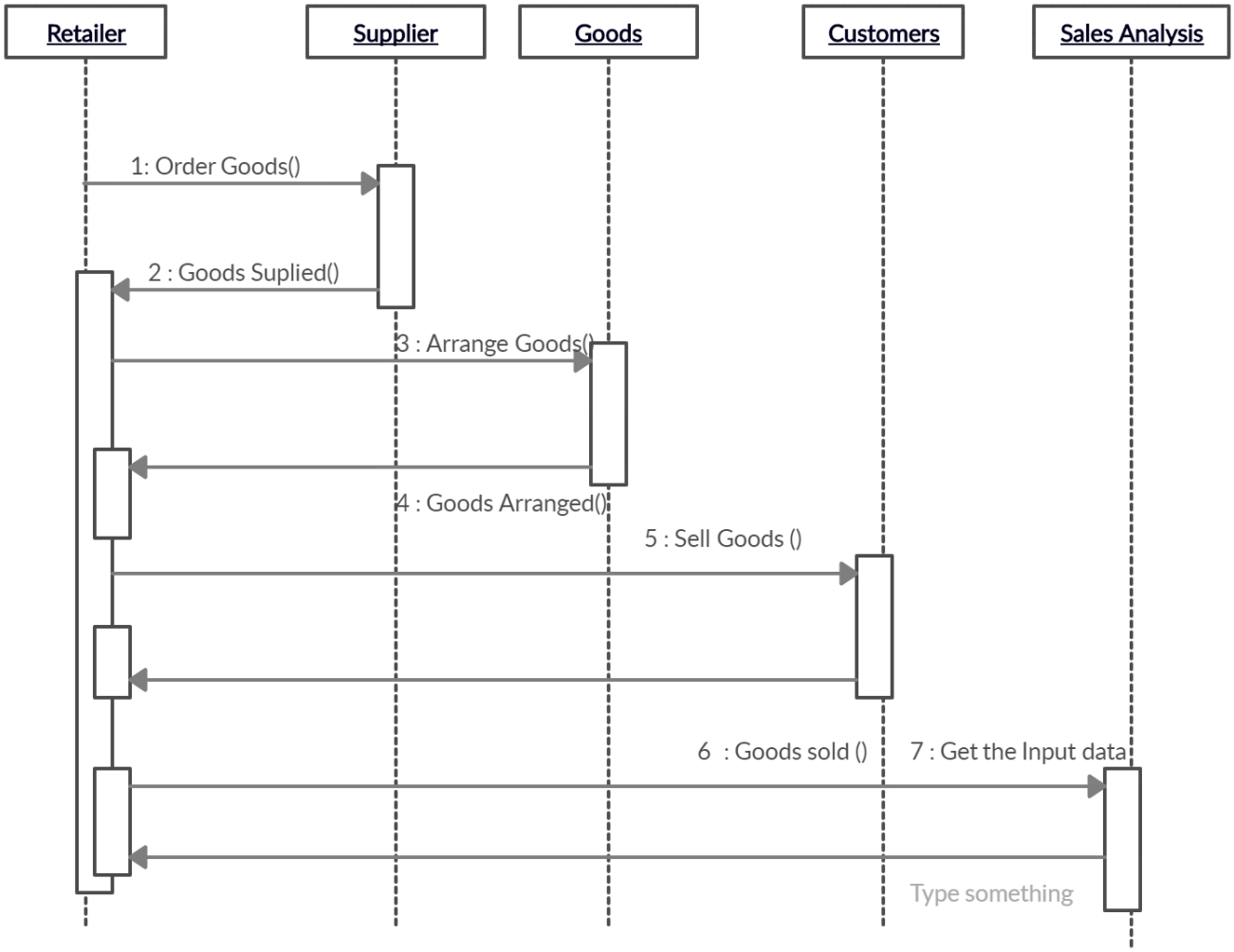


Figure 4.4: Sequence Diagram

Different interactions between the participants are depicted. Retailer is the main handler and function initiator, therefore, he interacts with all the other elements of the system. Interaction between retailer and supplier can be interpreted as retailer places orders for goods and then the supplier sends those goods to the retailer.

Chapter 5

Implementation

5.1 Code sinppets

Following is a code snippet of "create order" module. This snippet shows the code that fetches corresponding data from the database when a product is chosen or entered in the GUI.

```php
if(isset($_POST['btnsaveorder'])){

    $customer_name=$_POST['txtcustomer'];
    $order_date=date('Y-m-d',strtotime($_POST['orderdate']));
    $subtotal=$_POST['txtsubtotal'];
    $tax=$_POST['txttax'];
    $discount=$_POST['txtdiscount'];
    $total=$_POST['txttotal'];
    $paid=$_POST['txtpaid'];
    $due=$_POST['txtdue'];
    $payment_type=$_POST['rb'];
    ////////////////////////////////

    $arr_productid=$_POST['productid'];
        $arr_productname=$_POST['productname'];
        $arr_stock=$_POST['stock'];
        $arr_qty=$_POST['qty'];
        $arr_price=$_POST['price'];
        $arr_total=$_POST['total'];



    $insert=$pdo->prepare("insert into
    tbl_invoice(customer_name,order_date,subtotal,tax,discoun
    t,total,paid,due,payment_type)values(:cust,:orderdate,:st
    otal,:tax,:disc,:total,:paid,:due,:ptype)");

    $insert->bindParam(':cust',$customer_name);
    $insert->bindParam(':orderdate',$order_date);
    $insert->bindParam(':stotal',$subtotal);
    $insert->bindParam(':tax',$tax);
    $insert->bindParam(':disc',$discount);
    $insert->bindParam(':total',$total);
    $insert->bindParam(':paid',$paid);
    $insert->bindParam(':due',$due);
    $insert->bindParam(':ptype',$payment_type);

    $insert->execute();


    ////2nd insert query for tbl_invoice_details
```

Figure 5.1.1: Create Order Code

Following is a code snippet from the "product list" module. This is the script that fetches product name, category, cost price, selling price, etc from the database and displays the same of the GUI in a tabular manner.

```
            <?php
           $select=$pdo->prepare("select * from tbl_product order by pid desc");
            $select->execute();

           while($row=$select->fetch(PDO::FETCH_OBJ) ){

               echo'
               <tr>
               <td>'.$row->pid.'</td>
               <td>'.$row->pname.'</td>
               <td>'.$row->pcategory.'</td>
               <td>'.$row->purchaseprice.'</td>
               <td>'.$row->saleprice.'</td>
               <td>'.$row->pstock.'</td>
               <td>'.$row->pdescription.'</td>
               <td><img src="productimages/'.$row->pimage.'" class="img-rounded" width="40px" height="40px"?></td>

               <td>
               <a href="viewproduct.php?id='.$row->pid.'" class="btn btn-success" role="button"><span class="glyphicon glyphicon-eye-open"
               style="color:#ffffff" data-toggle="tooltip"  title="View Product"></span></a>
               </td>

               <td>
               <a href="editproduct.php?id='.$row->pid.'" class="btn btn-info" role="button"><span class="glyphicon glyphicon-edit"
               style="color:#ffffff" data-toggle="tooltip"  title="Edit Product"></span></a>
               </td>

               <td>
               <button id='.$row->pid.' class="btn btn-danger btndelete"><span class="glyphicon glyphicon-trash" style="color:#ffffff" data-
               toggle="tooltip" title="Delete Product"></span></button>
               </td>
                </tr>
              ';
           }
            ?>


         </tbody>
          </table> </div>


  </div>
</div>
```

Figure 5.1.2: Product List Code

This code shows packages imported for the data mining model and the logic used to obtain the constant for linear regression equation. This logic is operated in loops for every category and different constants are generated accordingly.

```
import warnings
warnings.filterwarnings('ignore')

# Import the numpy and pandas package

import numpy as np
import pandas as pd

# Data Visualisation
import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y,
train_size = 0.7, test_size = 0.3, random_state = 100)
X_train.head()
y_train.head()
import statsmodels.api as sm
X_train_sm = sm.add_constant(X_train)

# Fit the resgression line using 'OLS'
lr = sm.OLS(y_train, X_train_sm).fit()
lr.params
plt.scatter(X_train, y_train)
plt.plot(X_train, 6.948 + 0.054*X_train, 'r')
plt.show()
X_test_sm = sm.add_constant(X_test)
```

Figure 5.1.3: Constant Generation Code

The constants generated in the above logic substituted in the linear regression equation. first the model is tested against over fitting and under fitting errors. Then these irregularities are fixed before plotting the actual graph points. The code below shows sales prediction logic for individual categories. The final results are plotted on a line graph with every category being represented by a different line.

```
# Fit the resgression line using 'OLS'
lr = sm.OLS(y_train, X_train_sm).fit()
lr.params
plt.scatter(X_train, y_train)
plt.plot(X_train, 6.948 + 0.054*X_train, 'r')
plt.show()
X_test_sm = sm.add_constant(X_test)

# Predict the y values corresponding to X_test_sm
y_pred = lr.predict(X_test_sm)
y_pred.head()
from sklearn.metrics import mean_squared_error
from sklearn.metrics import r2_score
#Returns the mean squared error; we'll take a square root
np.sqrt(mean_squared_error(y_test, y_pred))
plt.scatter(X_test, y_test)
plt.plot(X_test, 6.948 + 0.054 * X_test, 'r')
plt.show()
# Plotting Values and Regression Line
max_x = np.max(X) + 100
min_x = np.min(X) - 100
# Calculating line values x and y
x = np.linspace(min_x, max_x, 1000)
y = c + m * x

# Ploting Line
plt.plot(x, y, color='#52b920', label='Regression Line')
```

Figure 5.1.4: Constant Generation Code

5.2 Result

To execute all the operations described in the proposed system, we have designed a user friendly GUI. All these modules will assist the shopkeepers in performing some activities with ease and accuracy. Following is an overview of all the modules of the system.

5.2.1 Registration

Registration is one of the primary modules in any data management system. Registration is the process by which the retailer identically creates their account. Registered users normally provide some sort of credentials (such as a username or e-mail address, and a password) to the system in order to prove their identity.

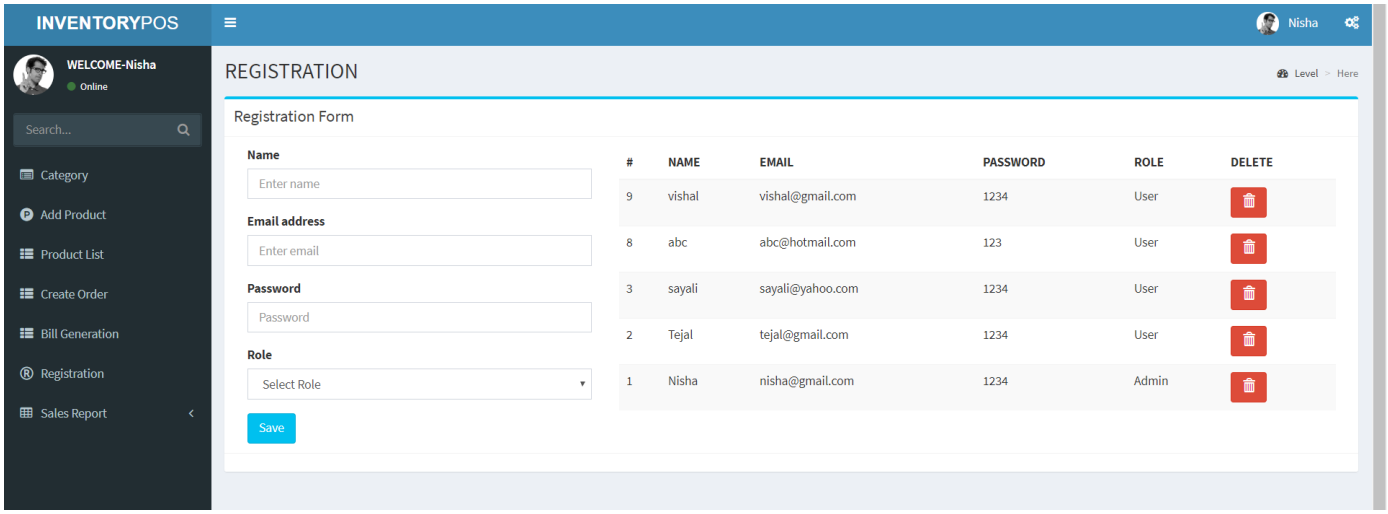


Figure 5.2.1: Registration

5.2.2 Login

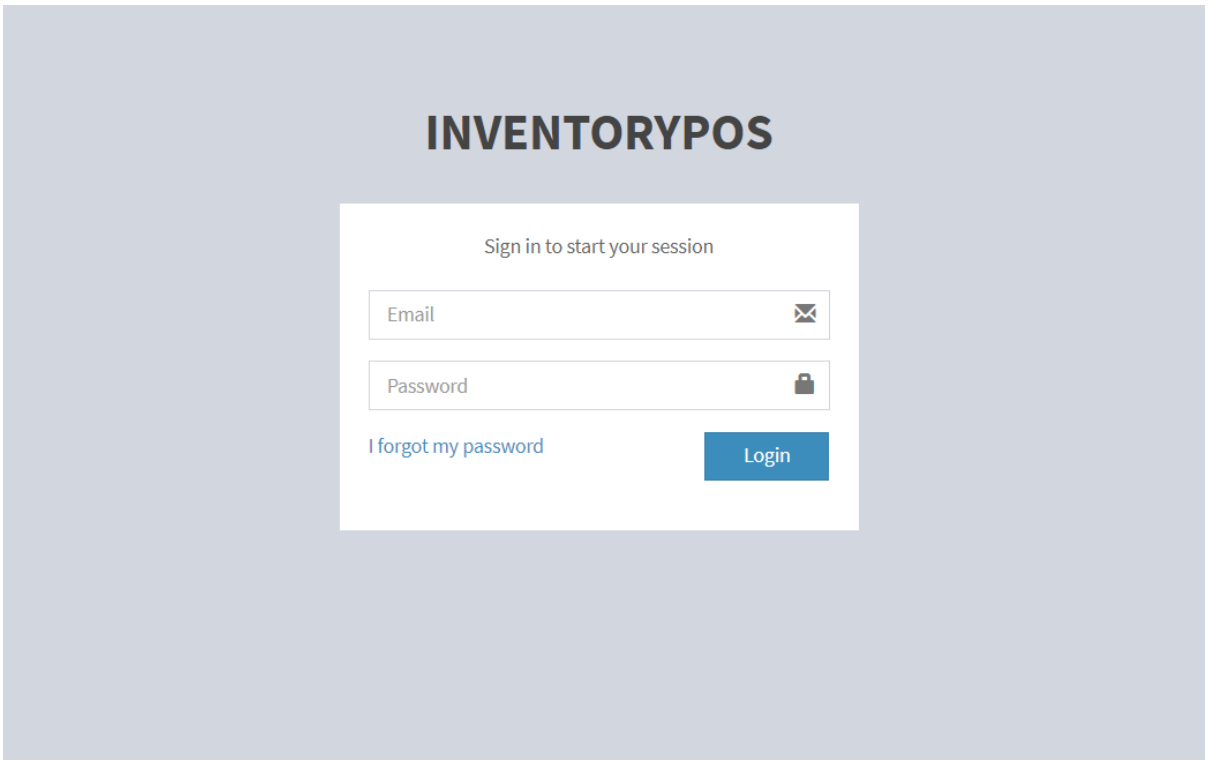


Figure 5.2.2: Login page

Every store has their login credentials.This ensures segregation of data.If the login cre-

dentials are incorrect access is denied.Sessions are managed through these credentials.

5.2.3 Category

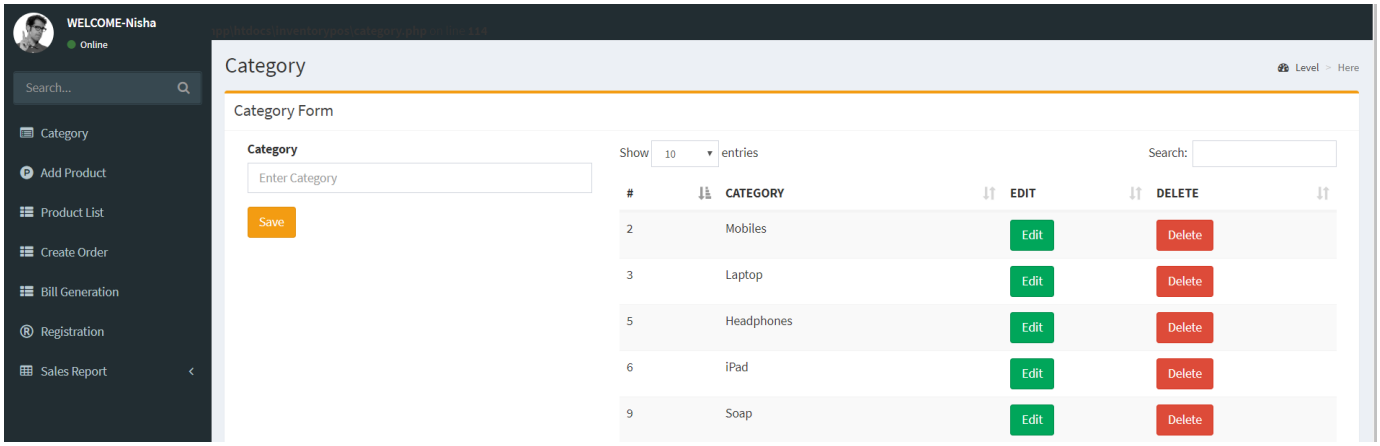


Figure 5.2.3: Category

A category is a module by which the retailer can add a different category to sort their products. They can add or delete the category as they want. This feature is helpful for them to easy search of any product detail.

5.2.4 Add Product

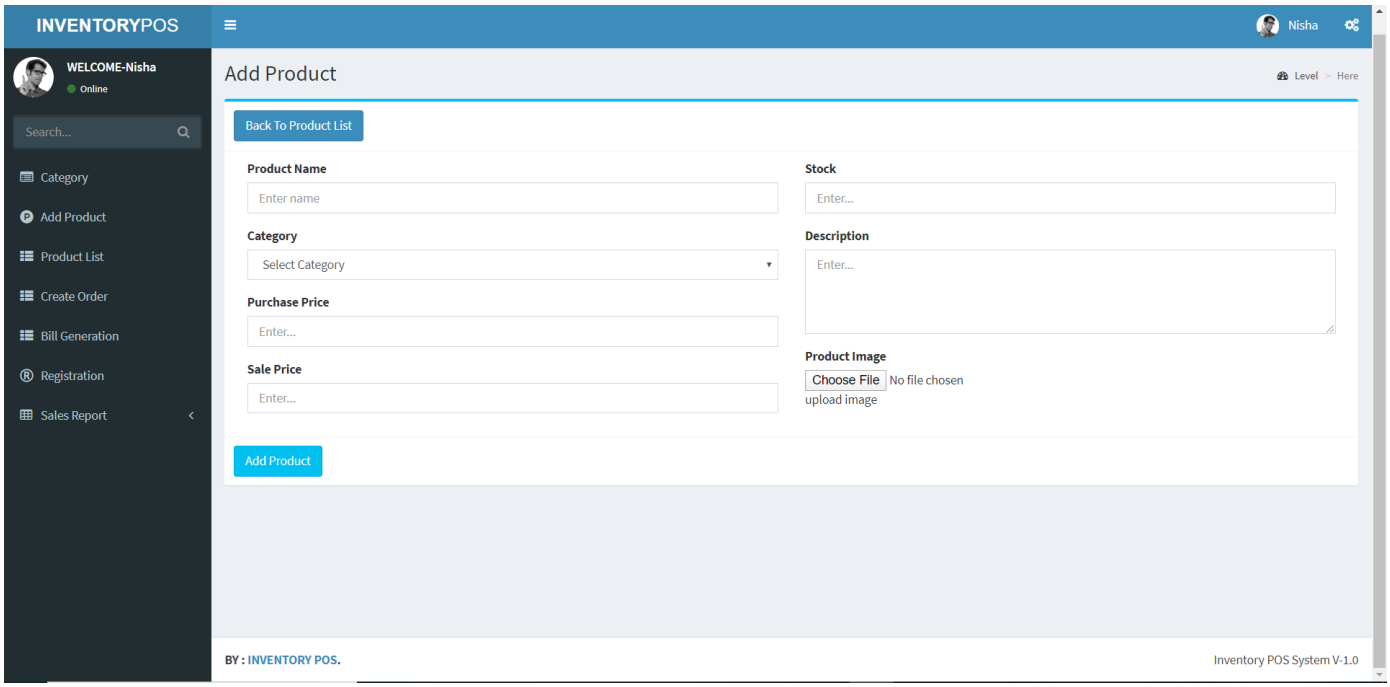


Figure 5.2.4: Add Product

Add product is the field by which the user can add their product detail to the system. The user can add the product detail such as name, category,prices price, sale price,stock, description,image.

5.2.5 Product List

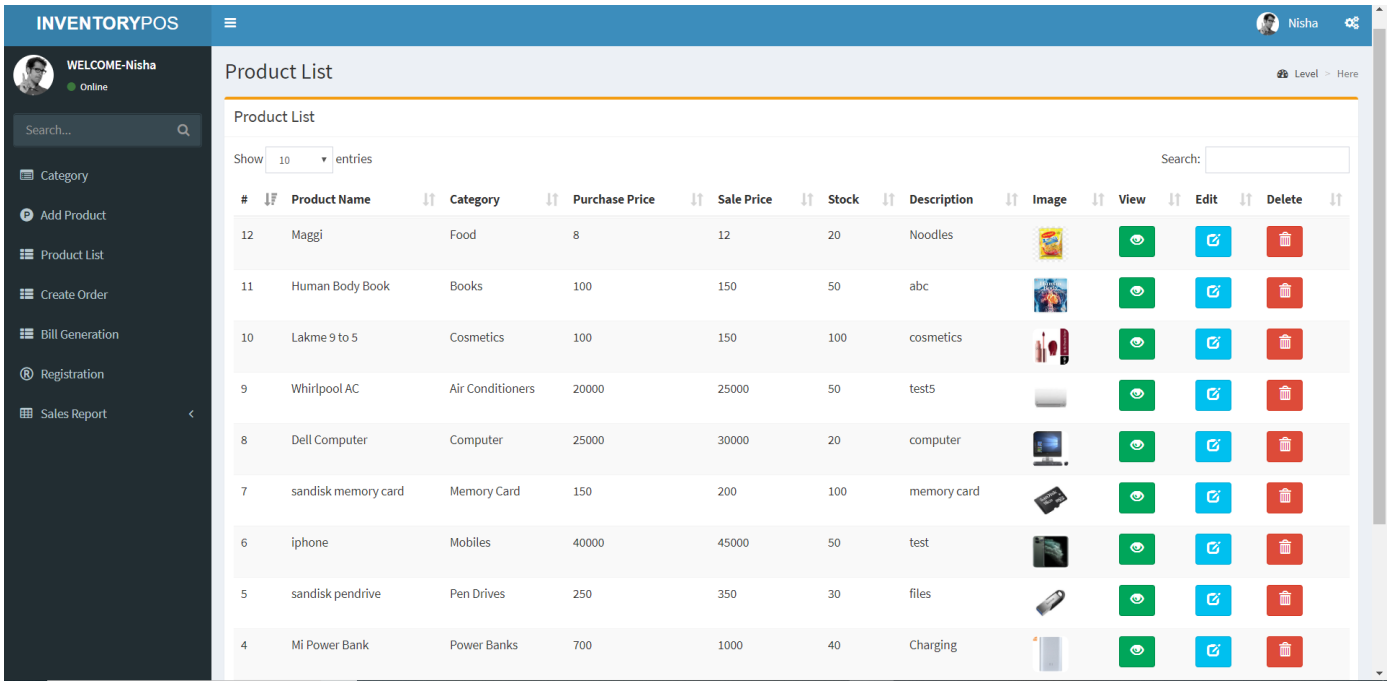


Figure 5.2.5: Product List

It is the main display of the product which is added by the user. Here user can check their products, add or delete them. They can edit the product if they want to.

5.2.6 Create Order

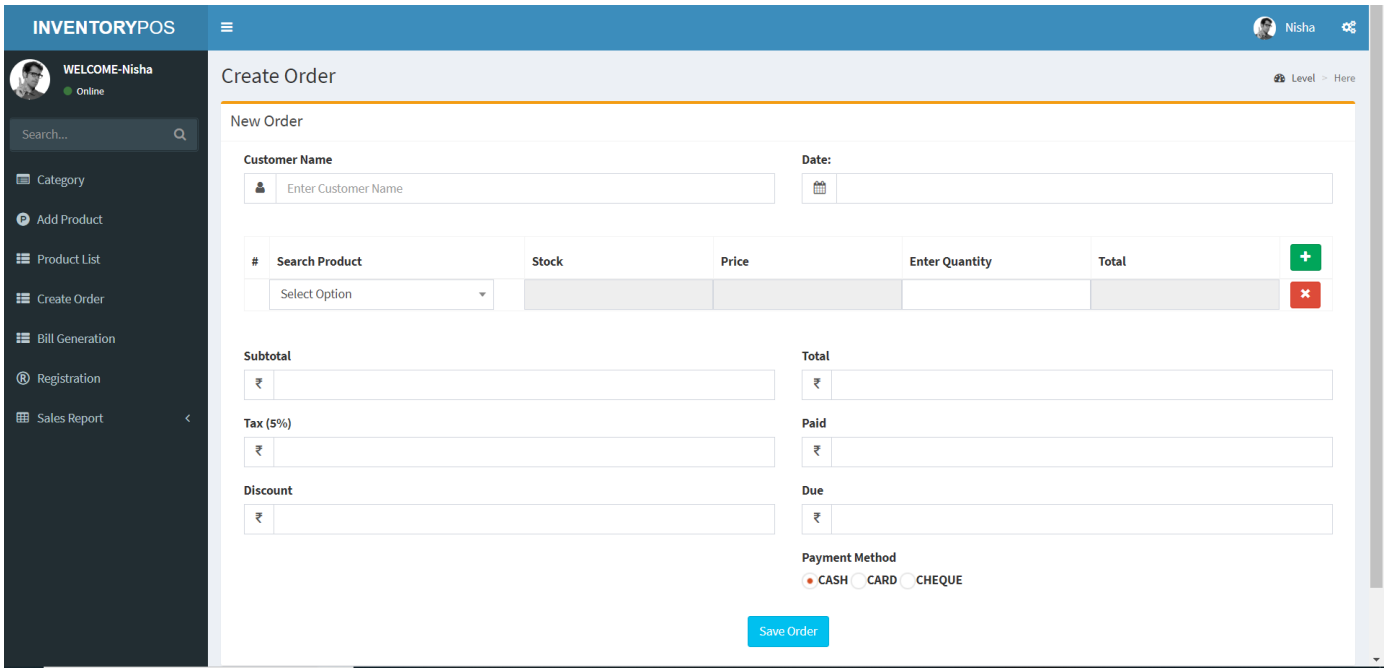
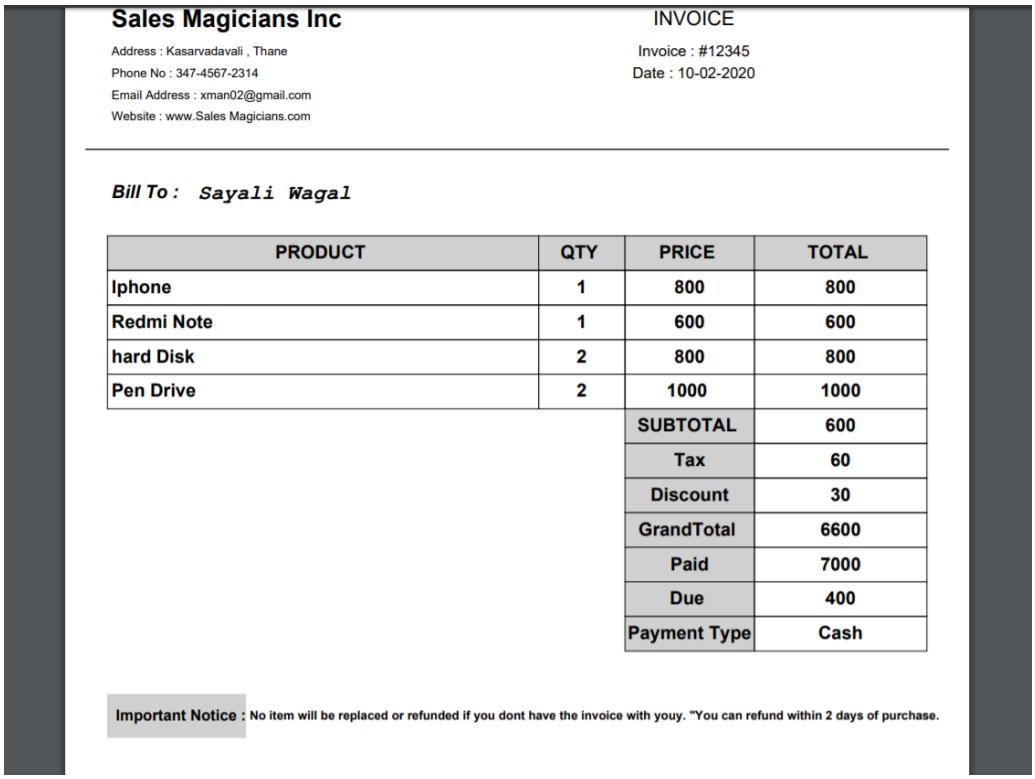


Figure 5.2.6: Create Order

This tab takes input for incoming purchases. The owners can keep a record of their inventory items through this tab. It has date filters to sort inventory items.

5.2.7 Bill Generation



Sales Magicians Inc

Address : Kasarvadavali , Thane
Phone No : 347-4567-2314
Email Address : xman02@gmail.com
Website : www.Sales Magicians.com

INVOICE

Invoice : #12345
Date : 10-02-2020

Bill To : ***Sayali Wagal***

PRODUCT	QTY	PRICE	TOTAL
Iphone	1	800	800
Redmi Note	1	600	600
hard Disk	2	800	800
Pen Drive	2	1000	1000
		SUBTOTAL	600
		Tax	60
		Discount	30
		GrandTotal	6600
		Paid	7000
		Due	400
		Payment Type	Cash

Important Notice : No item will be replaced or refunded if you dont have the invoice with youy. "You can refund within 2 days of purchase.

Figure 5.2.7: Bill Generation

The shopkeepers can generate bills by adding product name and quantity. The generated bill can be printed if required. The purchase details of customers can be viewed through this tab.

5.2.8 Sales Report

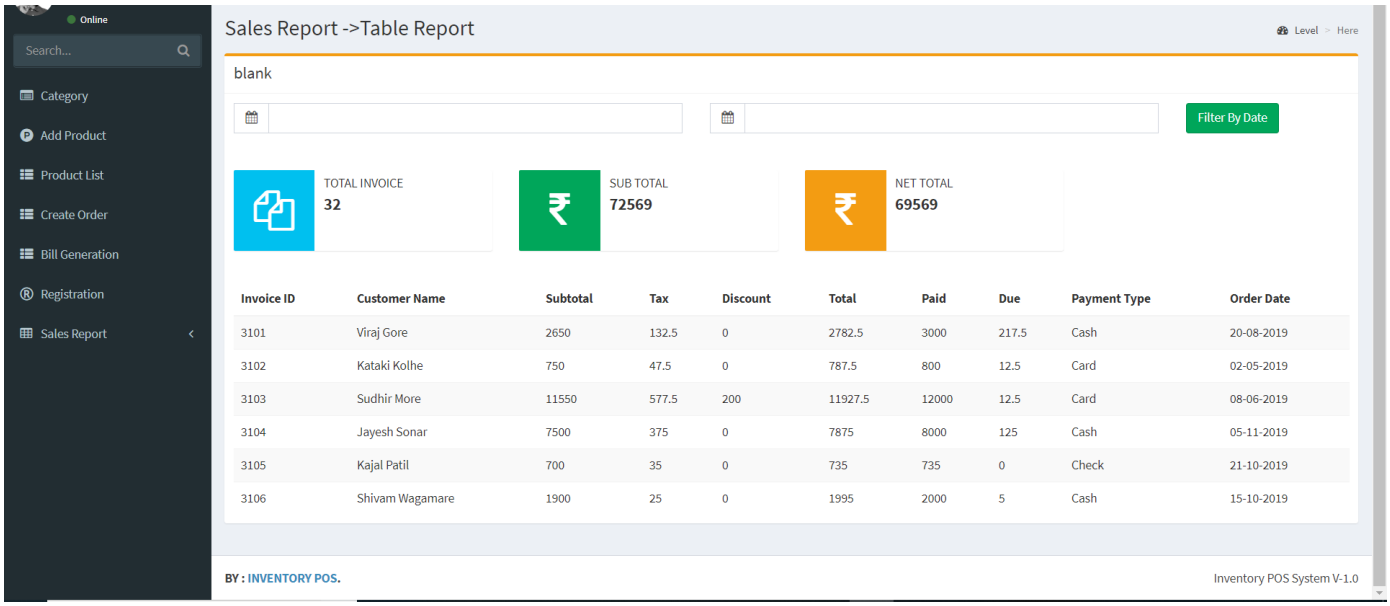


Figure 5.2.8: Sales Report

In this tab, there are two different views of the monthly sales of the shopkeeper's sales. First, It will show the table report of his sales in the targeted month. Second is by a graphical view of their sales data. Sales prediction is viewed through the line graph. Product analysis is given through PHPCharts.

Chapter 6

Testing

We have applied multiple testing methods to verify user inputs on each of the module. Following are the explanation of the different types of testing done.

SOFTWARE TESTING

Software Testing is defined as an activity to check whether the actual results match the expected results and to ensure that the software system is Defect free. It involves execution of a software component or system component to evaluate one or more properties of interest.Software testing also helps to identify errors, gaps or missing requirements in contrary to the actual requirements. It can be either done manually or using automated tools. Some prefer saying Software testing as a White Box and Black Box Testing. In simple terms, Software Testing means Verification of Application Under Test (AUT).

SOFTWARE TESTING TYPES:

(I)**MANUAL TESTING:**

Manual testing is the process of testing software by hand to learn more about it, to find what is and isn't working. This usually includes verifying all the features specified in requirements documents, but often also includes the testers trying the software with the perspective of their end user's in mind. Manual test plans vary from fully scripted test cases, giving testers detailed steps and expected results, through to high-level guides that steer exploratory testing sessions.

(II)**AUTOMATION TESTING:**

Automation testing is the process of testing the software using an automation tool to find the defects. In this process, testers execute the test scripts and generate the test results automatically by using automation tools. Some of the famous automation testing tools for functional testing are QTP/UFT and Selenium.

UNIT TESTING

Unit Testing is a type of software testing where individual units or components of a software are tested. The purpose is to validate that each unit of the software code performs as expected. Unit Testing is done during the development (coding phase) of an application by the developers. Unit Tests isolate a section of code and verify its correctness. A unit may be an individual function, method, procedure, module, or object.

In SDLC, STLC, V Model, Unit testing is first level of testing done before integration testing. Unit testing is a WhiteBox testing technique that is usually performed by the developer. Though, in a practical world due to time crunch or reluctance of developers to tests, QA engineers also do unit testing.

Unit Testing Techniques:

Black Box Testing - Using which the user interface, input and output are tested.
White Box Testing - used to test each one of those functions behaviour is tested.
Gray Box Testing - Used to execute tests, risks and assessment methods.

Unit testing is important because it is one of the earliest testing efforts performed on the code and the earlier defects are detected, the easier they are to fix. Early bug-detection is also the most cost-effective for a project, with code fixes becoming more expensive the later they're found in the life cycle

INTEGRATION TESTING

Integration Testing is defined as a type of testing where software modules are integrated logically and tested as a group. A typical software project consists of multiple software modules, coded by different programmers. The purpose of this level of testing is to expose defects in the interaction between these software modules when they are integrated

Integration Testing focuses on checking data communication amongst these modules. Hence it is also termed as 'I T' (Integration and Testing), 'String Testing' and sometimes 'Thread Testing'. Integration tests determine if independently developed units of software work correctly when they are connected to each other. The first part of testing would be unit testing, which would test that module on its own, against the specification that had been done in the design phase.

Types of Integration Testing:

(I) Bottom-Up Integration Testing:

In bottom-up testing, each module at lower levels is tested with higher modules until all modules are tested. The primary purpose of this integration testing is, each subsystem is to test the interfaces among various modules making up the subsystem. This integration testing uses test drivers to drive and pass appropriate data to the lower level modules.

(II) Top-Down Integration Testing:

Top-down integration testing technique used in order to simulate the behaviour of the lower-level modules that are not yet integrated.In this integration testing, testing takes place from top to bottom. First high-level modules are tested and then low-level modules and finally integrating the low-level modules to a high level to ensure the system is working as intended.

RECORD OF TEST CASES PERFORMED

TESTCASE NO.	CONDITION	INPUT	EXPECTED	ACTION
1.	Product added in add product module should be refelected in the product list	Product: Drawing Sheets Quantity: 20	The Product name and Quantity should be displayed in the product list	Product name and Quantity is displayed in the product list
2.	Reset Password option should prompt after clicking on Forgot Password	After clicking Forgot Password the Reset Password option prompts inorder to reset the New Password	Reset Option will then allow to reset the New Password	The New Password is then set.
3.	Fetching the product details by selecting the product name in create order module	Selecting the Product Name from the drop down Ex. Dell Computer	The product details are fetched from the database the details provided by database are Stock, Price, Quantity and Total	The details are fetched from the database and displayed for creating the order for the user

Chapter 7

Conclusion and Future Scope

CONCLUSION:

A good percentage of people in India have access to Website and that percentage will greatly increase in coming 2-3 years. With such favorable circumstances, an Website is ought to flourish and attract a wider customer base over a period of time. Thus it is advantageous to have a Web application that not only assists with inventory and invoice operations but also helps with sales analysis.

Since sellers are presented with performance reports and product analysis, they can make necessary changes in their policies or way of operation after thoroughly studying and understanding all the factors that impact their sales. Similarly, customer's will avail the experience of accessing the right products at right time and will stay informed about new products simultaneously. With time, as more and more data is fed into the database the accuracy of the data mining model will improve and regression analysis will be presented with over a 96

From the perspective of profits for the vendor company, they can present in-app purchases to the shopkeepers like unlimited sales report storage, e-mail sharing of invoices with customers, etc. Therefore this system will not only assist in bringing about social empowerment and development but will also present profitable business opportunities to web development companies.

FUTURE SCOPE:

Android App:
This project can be extended as an Android application. Since a wide percentage of local shop owners use Android devices, it is convenient to have the operate the same facilities through a mobile phone.

Cloud Software:
For local brands that operate as a franchise, a cloud software will be an appropriate option to keep data collaborated with other branches while still having data isolation from other stores.

Bibliography

[1] C. Ezhilarasan and S. Ramani, "Performance prediction using modified clustering techniques with fuzzy association rule mining approach for retail,"2017 International Conference on Intelligent Computing and Control (I2C2), Coimbatore, 2017

[2] Xiaojun Zhang, Jisheng Pei and Xiaojun Ye, "Demographic transformation and clustering of transactional data for sales prediction of convenience stores,"2016 IEEE International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), Chengdu, 2016.

[3] R. P. Santi and M. L. Khodra, "Profit Prediction Using Regression Model for Travel Agents,"2018 International Workshop on Big Data and Information Security (IWBIS), Jakarta, 2018.

[4] Y. Kaneko and K. Yada, "A Deep Learning Approach for the Prediction of Retail Store Sales,"2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW), Barcelona, 2016.

[5] T. Chen et al., "TADA: Trend Alignment with Dual-Attention Multi-task Recurrent Neural Networks for Sales Prediction," 2018 IEEE International Conference on Data Mining (ICDM), Singapore, 2018.

[6] S. Cheriyan, S. Ibrahim, S. Mohanan and S. Treesa, "Intelligent Sales Prediction Using Machine Learning Techniques," 2018 International Conference on Computing,Electronics and Communications Engineering(iCCECE),Southend, United Kingdom, 2018.

Acknowledgement

We have great pleasure in presenting the report on **E-commerce Based Sales Prediction Framework** We take this opportunity to express our sincere thanks towards our guide **Ms.Rujata Chaudhari** & Co-Guide **Mr.Vishal Badgujar** Department of IT, APSIT thane for providing the technical guidelines and suggestions regarding line of work. We would like to express our gratitude towards his constant encouragement, support and guidance through the development of project.

We thank **Mr. Kiran B. Deshpande** Head of Department,IT, APSIT for his encouragement during progress meeting and providing guidelines to write this report.

We thank **Ms. Anagha Aher** & **Ms.Apeksha Mohite** BE project co-ordinator, Department of IT, APSIT for being encouraging throughout the course and for guidance.

We also thank the entire staff of APSIT for their invaluable help rendered during the course of this work. We wish to express our deep gratitude towards all our colleagues of APSIT for their encouragement.

Tejal Tandel:
16104047:

Sayali Wagal:
17204016:

Nisha Singh:
17204011:

Publication

[1.]Paper entitled **"Case Study on an Android App for Inventory Management System with Sales Prediction for Local Shopkeepers"** is presented at **"2020 6th IEEE International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)"** by **"Tejal Tandel, Sayali Wagal, Nisha Singh, Ms. Rujata Chaudhari, Mr. Vishal Badgujar".**