A Project Report on

E-commerce Based Sales Prediction Framework

Bachelor of Engineering

in

Information Technology

by

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

Name of Guide: Prof. Rujata Chaudhari
Name of Co-Guide: Prof. Vishal Badgujar

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

CERTIFICATE

This is to certify that the project Synopsis entitled ***" E-commerce Based Sales Prediction Framework"*** Submitted by ***Tejal Tandel (16104047),Sayali Wagal (17204016),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

(Prof. Vishal Badgujar)
Co-Guide

(Prof. Rujata Chaudhar)
Guide

Prof. 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.

_____

(Tajel Tandel 16104047)

_____

(Sayali Wagal 17204016)

_____

(Nisha Singh 17204011)

Date:

Abstract

There are several e-commerce sites that have emerged in the market but despite of that not many retailers have integrated their services with the same. The said e-commerce systems operate with their own name and the merchant's name is not highlighted in the process. Therefore, business owners who want to operate with their own name are hesitant to use such systems even though they can improve their sales. In addition to this, traditional e-commerce websites are consumer centric. The sellers are only presented with reports of current sales or notified if restocking is required.

The proposed system is of an e-commerce platform, with an additional feature for sales prediction and product recommendation for retailers. The consumers will get all the benefits of online purchasing but will be focused upon one organization only. The system will also gather consumer feedback for better analysis and performance evaluation. This system can help retailers organise their inventory efficiently and avoid financial and commercial losses. In addition to this retailers will be provided with a sales prediction and product recommendation feature to boost their sales. This will be done using data mining. Data mining is the process of extracting relevant data from data sources and then processing it to generate desired outputs. Data mining techniques have become full-fledged in the business market since 2015 and have proven to give accurate predictive analysis.

Introduction

Internet has become a vital part in day-to-day life today. 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. 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.

Our proposed system is an amalgamation of both these trends which will make the whole process more convenient. In addition to these features, to make the entire activity even more captivating, our system has recommendation feature for both sellers as well as customers. The customers will receive recommendations based on their and other user's purchase trends whereas the sellers will be recommended to display the products that are high on trend for the current market or that are most frequently purchased from them.

This can be brought about by using market basket analysis. It is a type of data mining technique where relations between two or more data can be identified. Thus by implementing this technique, products that are interrelated or complimentary to each other can be given as 'recommendations" to both, the customers and the sellers.

In addition to this retailers will be provided with a sales prediction and product recommendation feature to boost their sales. This will be done using data mining. Data mining is the process of extracting relevant data from data sources and then processing it to generate desired outputs. Data mining techniques have become full-fledged in the business market since 2015 and have proven to give accurate predictive analysis.

The product recommendation output will be brought about by using association rule mining using the Apriori algorithm. Association rule mining is a procedure of finding some correlation or association between the data items present in the database. Data from different data sources like relational databases, transactional databases or other form of data repositories can be analysed using this. Association rules can be generated for the products that are similar or complementary to each other. The results will be a suggestion report based on customer buying patterns. Using this the retailer will be able to take product based decisions and update the inventory accordingly.

Objectives

The main objective of this is to help retailers to improve there sales by giving them a sales prediction of there sales in the graphical from so to utilized there overall sales and minimize there losses by purchasing the required quantity products and maximize there profit.And with the market basket analysis they attract customer to buy more product buy using the market basket strategies.It will also generate a report on monthly sales so that retailers can analyse their expenses.

Literature Review

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

- **"Performance prediction using modified clustering techniques with fuzzy association rule mining approach for retail," C. Ezhilarasan and S. Ramani, 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.

- **"Demographic transformation and clustering of transactional data for sales prediction of convenience stores," 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.

- **"Profit Prediction Using Regression Model for Travel Agents," R. P. Santi and M. L. Khodra, 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 itemsets 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 overfitting of data if the independent itemsets are not linear.

- **"Data Mining Applications for Sales Information System Using Market Basket Analysis on Stationery Company," A. Setiawan, G. S. Budhi, D. H. Setiabudi and R. Djunaidy, 2017 International Conference on Soft Computing, Intelligent System and Information Technology (ICSIIT), Denpasar, 2017**

The system proposed in the paper is for an online stationery. The system aims at improving the current functioning method of the same by adding association rules and then giving suggestions to the owner on what products should be included with the already existing ones.

Association rules are if-then statements that help you find relationship between the items in the data set. It is dependent on two factors- support and confidence. Support is an indication of how frequently the items appear in the data. Confidence indicates the number of times the if-then statements are found true. In Apriori algorithm, candidate itemset is generated from larger itemset from the previous pass. In every pass, the items which do not satisfy the support are discarded.

Association rules are generated by using the Apriori algorithm. Books, pencils, book covers, erasers, labels are linked to one another for example books, book covers, labels, pencils, erasers, books book covers, labels, pencils. Using this information, product recommendation is provided to stationery owner so he can prompt the buyers the possible complimentary items to purchase.

The drawback for this algorithm is that it gets dismissed if no more itemsets are prolonged. In addition to this, the performance time is more as it consumes memory resources at every rule creation. It needs more exploration space and computing cost may be more depending on the size of itemset and the support-confidence conditions.

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.

Existing System Architecture/Working

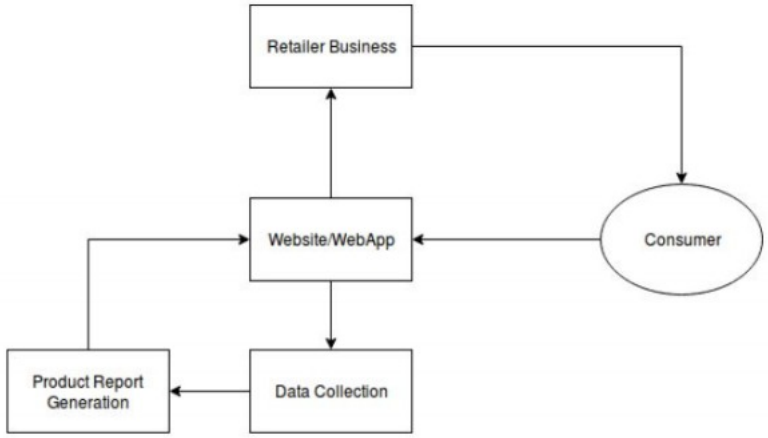


Figure 1: Existing system of an e-commerce platform

The existing e-commerce platform will give the feature of the buying and selling. In this system the person can sign-up as a seller by providing information about the business and the products the seller proposes to sell. Also, it will give the customer recommendation about the product based on there buying history. The sellers get an order notification or a restock alert. The system seller then delivers the product to the consumer via some middle man. The financial transaction is done online but there are no reports generated out of it to provide some insights to the retailer.

Proposed System Architecture/Working

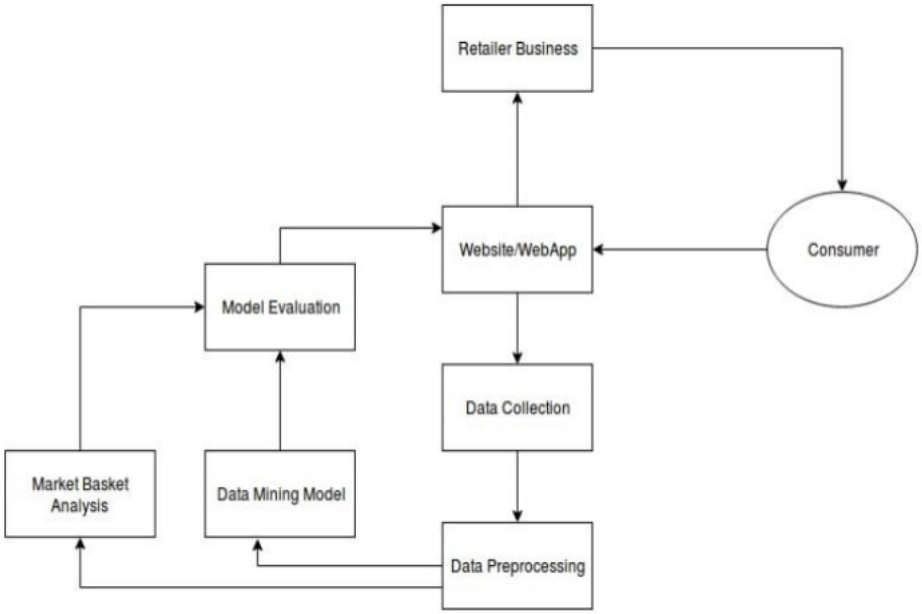


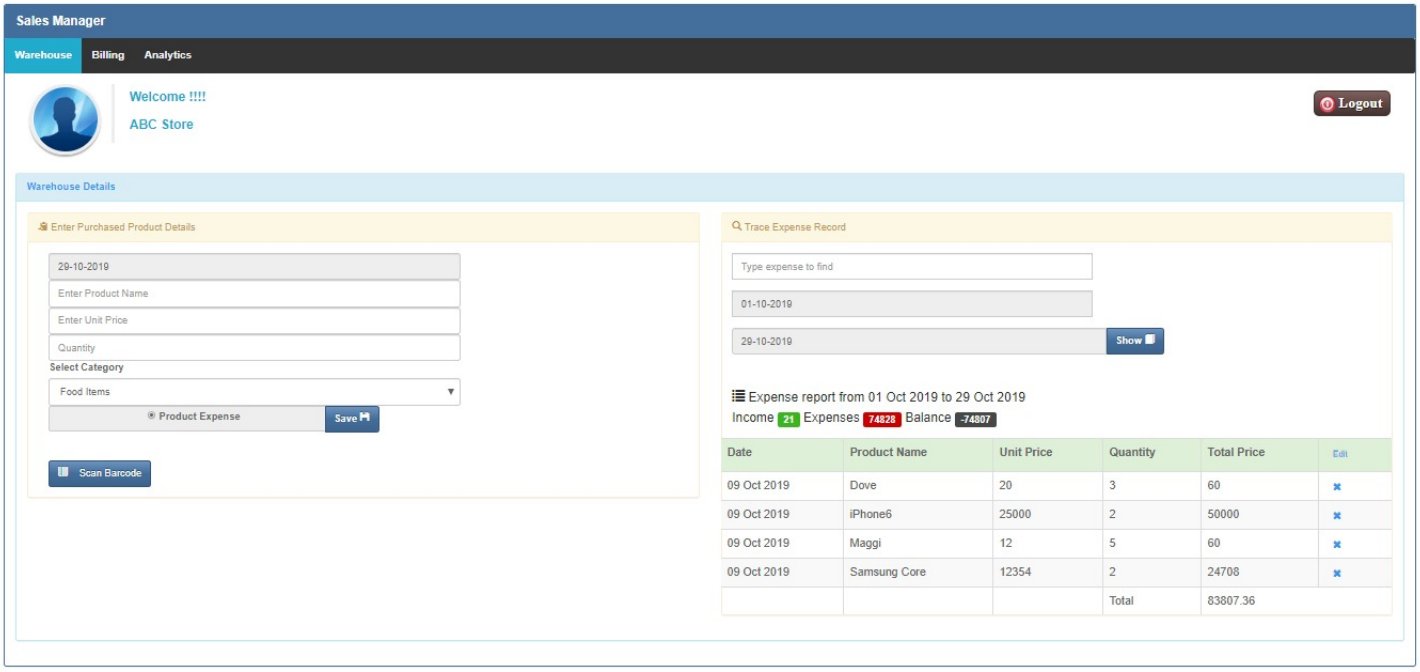
Figure 2: Proposed system of an e-commerce platform

The proposed e-commerce platform will give general features of the same along with hosting to the business organizational website. Using classification and clustering algorithms, sales analysis will be performed and a predictive result will be generated from it. We have also included market basket analysis which will use association rules to recommend related products to customers as well as retailers. A collective analysis of the said reports will help retailers make business decisions which are convenient and cost effective.

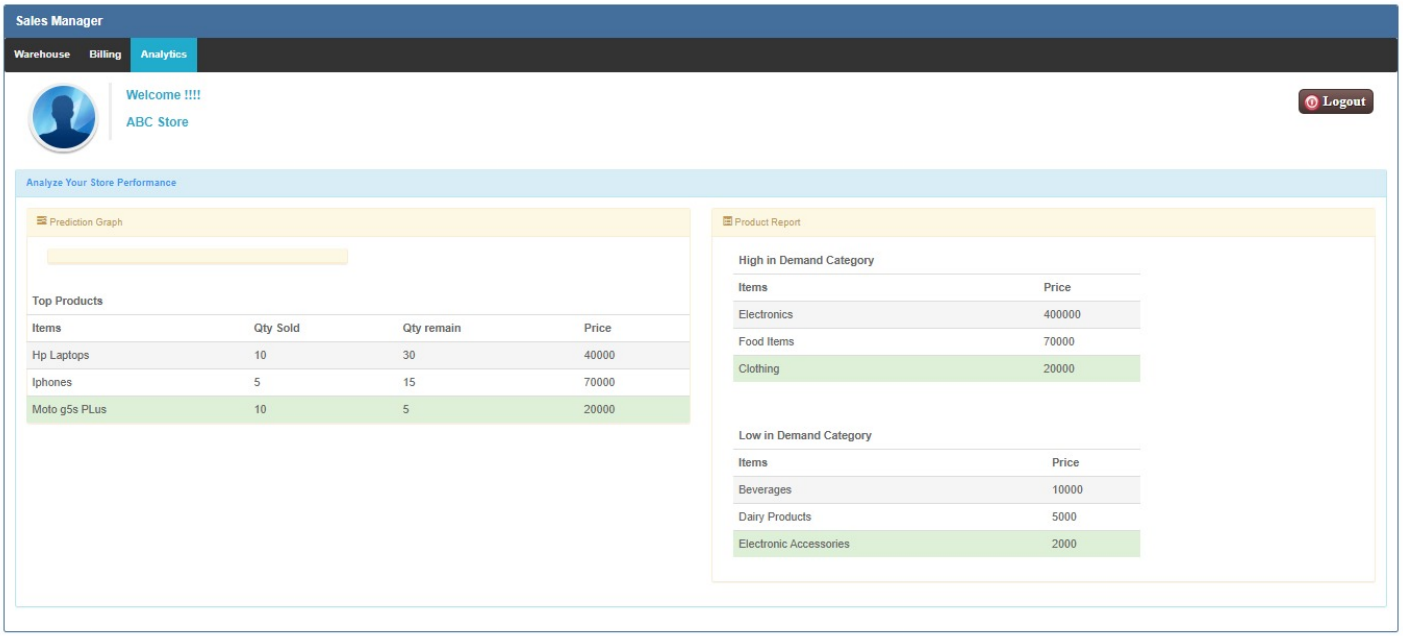
In this system they can add more description about product. It will product description should impress them without making them spend much time. Another important element of catalog that can help seller to increase their sales by adding there product in the right category and sub-category. To increase their sales and become a successful seller, they should always work towards customers satisfaction and receive their commitment. The feedback feature will ensure that customer concerns are understood and resolved speedily

- **Desing**

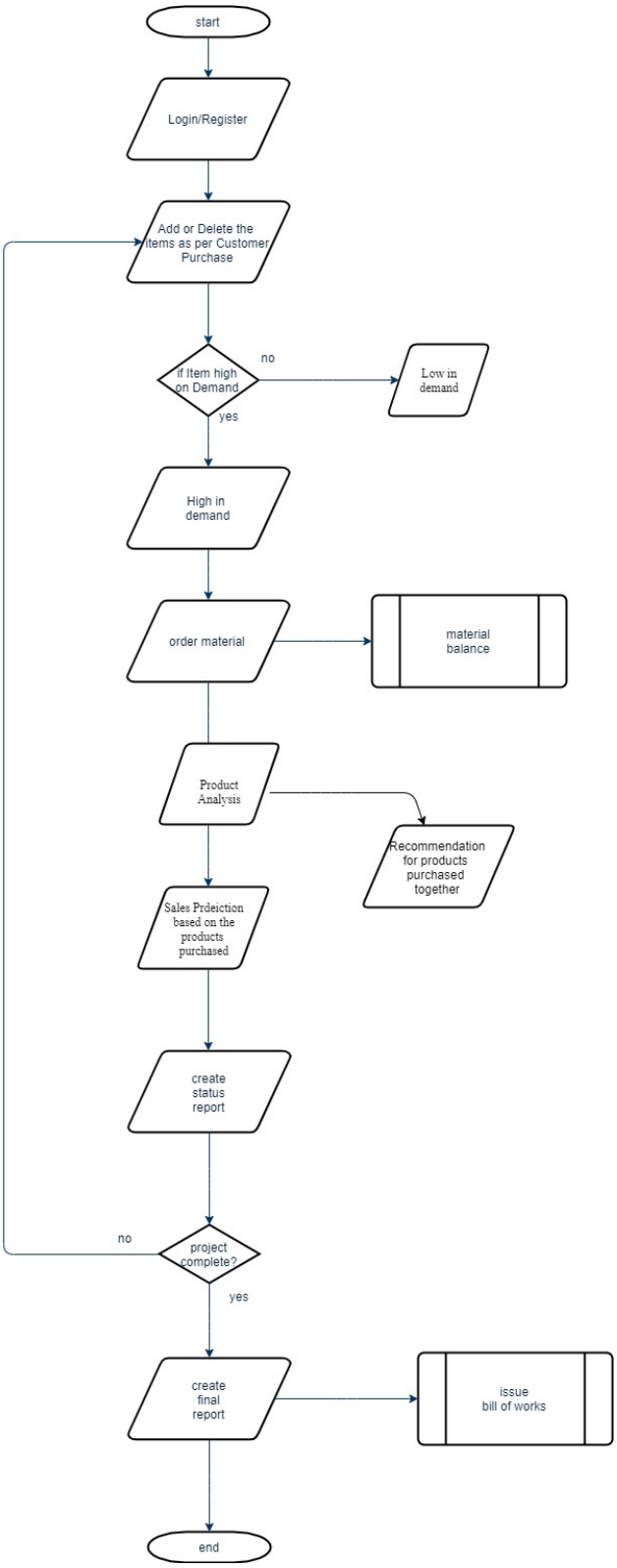
1. Homepage



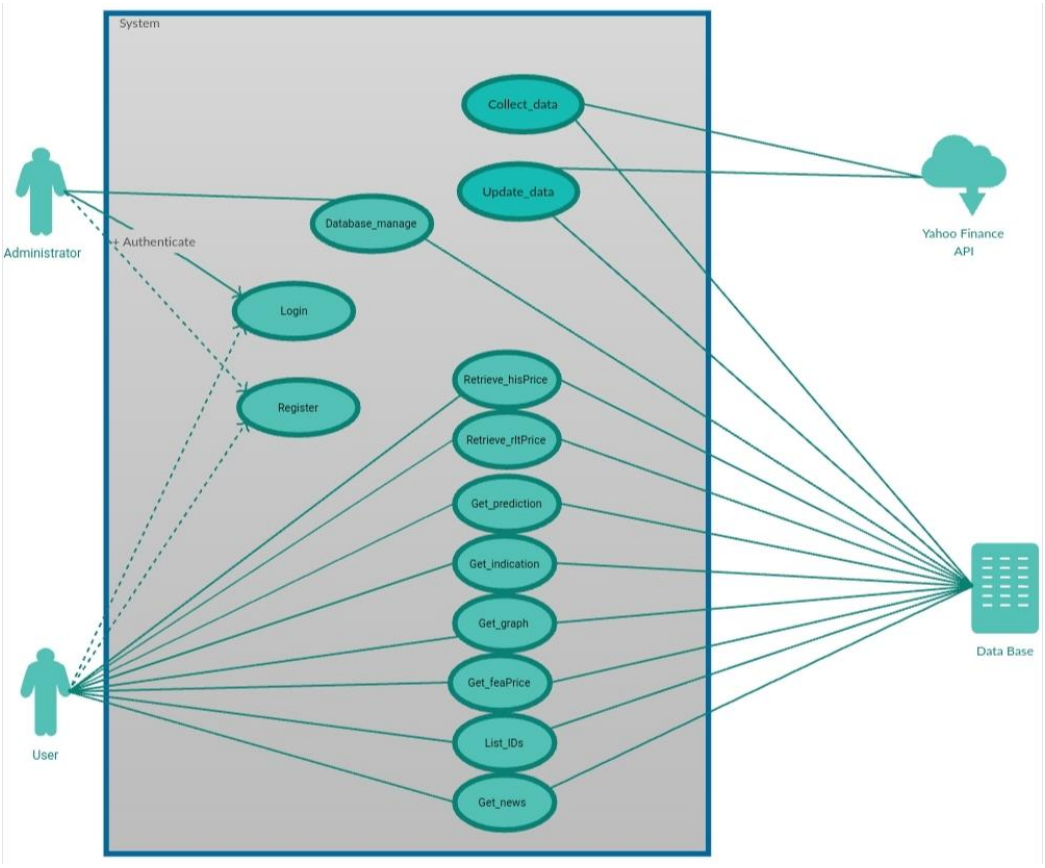
1. Analytics Page



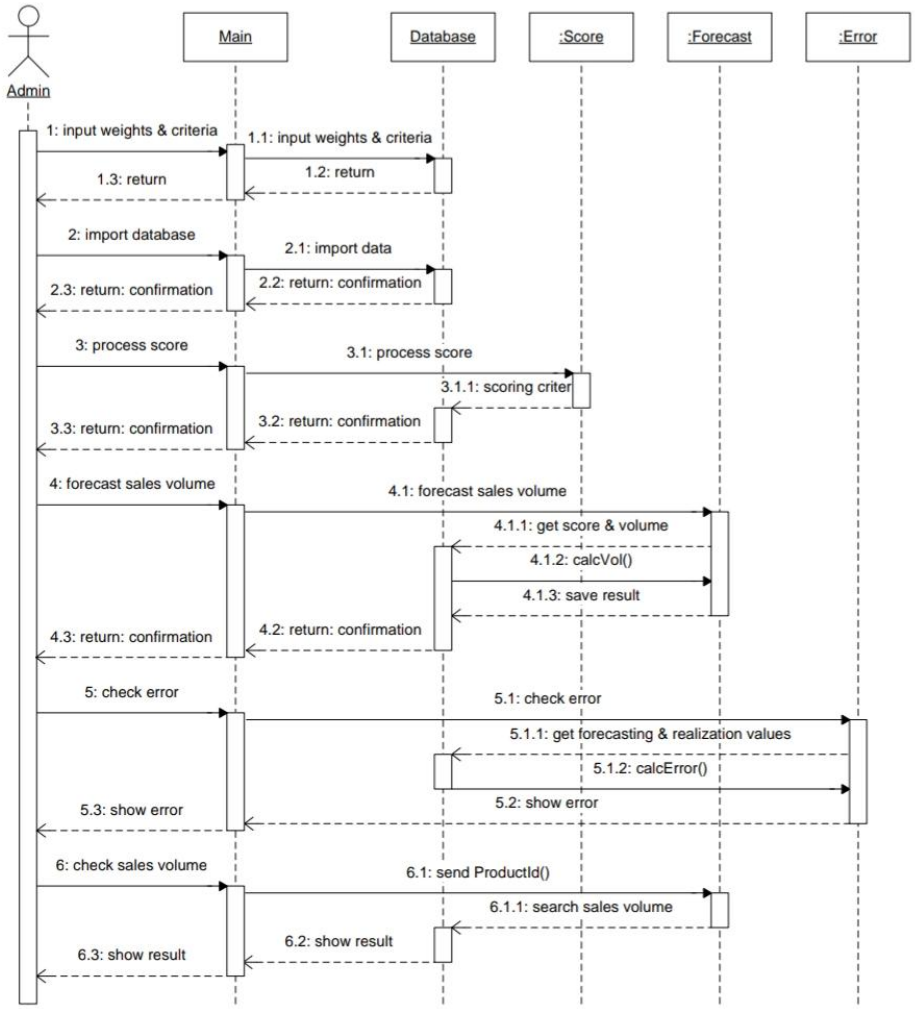
- **Desing Flow**


start
Login/Register
Add or Delete the Items as per Customer Purchase
if Item high on Demand
no
Low in demand
yes
High in demand
order material
material balance
Product Analysis
Recommendation for products purchased together
Sales Prdeiction based on the products purchased
create status report
no
project complete?
yes
create final report
issue bill of works
end

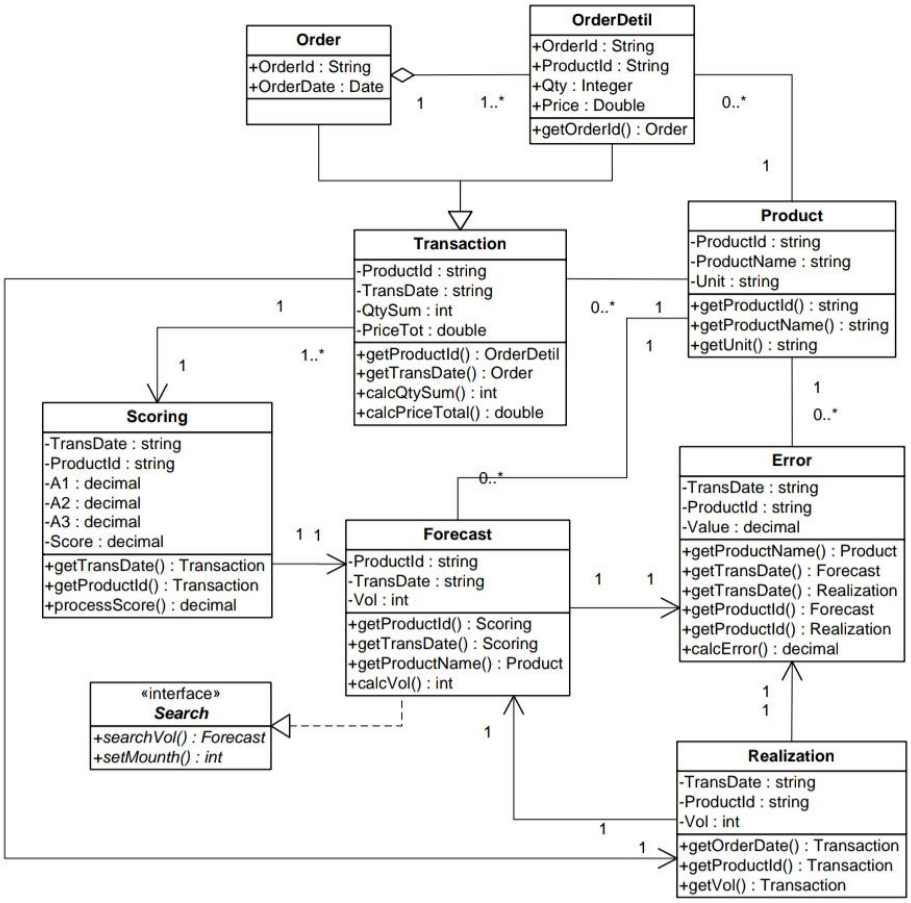
- usecase


System
Collect_data
Update_data
Database_manage
Authenticate
Administrator
Login
Register
Retrieve_hisPrice
Retrieve_rltPrice
Get_prediction
Get_indication
Get_graph
Get_feaPrice
List_IDs
Get_news
User
Yahoo Finance
API
Data Base

- sequence


Admin
Main
Database
:Score
:Forecast
:Error
1: input weights & criteria
1.1: input weights & criteria
1.2: return
1.3: return
2: import database
2.1: import data
2.2: return: confirmation
2.3: return: confirmation
3: process score
3.1: process score
3.1.1: scoring criter
3.2: return: confirmation
3.3: return: confirmation
4: forecast sales volume
4.1: forecast sales volume
4.1.1: get score & volume
4.1.2: calcVol()
4.1.3: save result
4.2: return: confirmation
4.3: return: confirmation
5: check error
5.1: check error
5.1.1: get forecasting & realization values
5.1.2: calcError()
5.2: show error
5.3: show error
6: check sales volume
6.1: send ProductId()
6.1.1: search sales volume
6.2: show result
6.3: show result

- Class


Order
+OrderId : String
+OrderDate : Date
OrderDetil
+OrderId : String
+ProductId : String
+Qty : Integer
+Price : Double
+getOrderId() : Order
Transaction
-ProductId : string
-TransDate : string
-QtySum : int
-PriceTot : double
+getProductId() : OrderDetil
+getTransDate() : Order
+calcQtySum() : int
+calcPriceTotal() : double
Product
-ProductId : string
-ProductName : string
-Unit : string
+getProductId() : string
+getProductName() : string
+getUnit() : string
Scoring
-TransDate : string
-ProductId : string
-A1 : decimal
-A2 : decimal
-A3 : decimal
-Score : decimal
+getTransDate() : Transaction
+getProductId() : Transaction
+processScore() : decimal
Forecast
-ProductId : string
-TransDate : string
-Vol : int
+getProductId() : Scoring
+getTransDate() : Scoring
+getProductName() : Product
+calcVol() : int
Error
-TransDate : string
-ProductId : string
-Value : decimal
+getProductName() : Product
+getTransDate() : Forecast
+getTransDate() : Realization
+getProductId() : Forecast
+getProductId() : Realization
+calcError() : decimal
«interface»
Search
+searchVol() : Forecast
+setMounth() : int
Realization
-TransDate : string
-ProductId : string
-Vol : int
+getOrderDate() : Transaction
+getProductId() : Transaction
+getVol() : Transaction

Summary

E-commerce is ought to flourish and attract a wider customer base over a period of time. Thus it is advantageous to have an e-commerce system that not only assists with online transactions 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.

This model will help the retailers to plan their investment and business tasks more effectively. The sales report for future sales will help retailers analyse the product that is high on demand and they can add more items into their stocks to boost the sales. The product wise report will avoid stocking of products that do not significantly contribute to the overall earnings and the retailers can discard them for a better financial stability.

- Sales prediction
- Datamining
- Market basket analysis
- Generate a report on monthly sales so that retailers can analyse their expenses.
- Give customers as well as retailers recommendations as to which other items can be purchased along with the selected one
- Help retailers make product-wise decisions like reordering or discarding the product.
- Help retailers stock appropriate products based on the in-demand product sale.

References

[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] Alexander Setiawan, Gregorius Satia Budhi, Djoni Haryadi Setiabudi, Ricky Djunaidy, "Data Mining Applications for Sales Information System Using Market Basket Analysis on Stationery Company," 2017 International Conference on Soft Computing, Intelligent System and Information Technology, 2017.

[5] 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.

[6]] 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.

[7] 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.

1 Publication

Paper entitled **E- Commerce Website With Sales Prediction** is presented at **"2019 International Conference on Innovations in Computer Technology and Applications (IC2TA-2019), January 4-5, 2019, Thane ,Mumbai"** by
Tejal Tande.
Sayali Wagal.
Nisha Singh.
Prof. Rujata Chaudhari.
Prof. Vishal Badgujar.