

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

A Project Report on

E-commerce Based Sales Prediction Framework

Submitted in partial fulfillment of the degree of
Bachelor of Engineering(Sem-7) in
INFORMATION TECHNOLOGY

Tejal Tandel(16104047)

Sayali Wagal(17204016)

Nisha Singh(17204011)

Under the Guidance of

Project Guide: Prof. Rujata Chaudhari

Project Co- guide: Prof. Vishal Badgujar

1. Project Conception and Initiation

1.1 Abstract

Data mining is a strong new technology with great potential to help companies to focus on the most important information in their data warehouses. This will help in marketing and sales.

The proposed system uses clustering and market basket analysis types of algorithm with modification which will make algorithm more efficient. The analyst can perform data mining and extraction and finally conclude the result and make appropriate decision for retailer businesses.

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.

1.2 Objectives

- To help retailers stock appropriate products based on the in-demand product sale.
- To help retailers make product-wise decisions like reordering or discarding the product.
- To give customers as well as retailers recommendations as to which other items can be purchased along with the selected one.
- To generate a report on monthly sales so that retailers can analyse their expenses

1.3 Literature Review

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. 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

1.3 Literature Review

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. 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.

1.3 Literature Review

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. 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

1.4 Problem Definition

Retailers face a loss in revenue due to improper investment and management activities. Many incur losses because their customer reach is limited. They cannot properly identify the demand of products and therefore end up buying limited products. On the other hand they can underestimate the growth of some products and remove them from their shop.

All these cases lead to a financial loss and stagnate the shop's growth. If the shopkeepers are equipped with tools and applications that give them insights of their sales and recommend products to them, the shopkeepers can improve their finances.

1.5 Scope

- This project will be beneficial to bigger supermarkets as well, having vivid range products. It will save them the extra expenditure of purchasing a third party sales prediction software.
- The project can be expanded on Android and iOS platforms to make it more easy to use and thereby reduce the hardware space.
- This project can be exported as a bundle which can be installed privately into individual shops.

1.6 Technology stack

Frontend:

Bootstrap, HTML5, CSS3, JavaScript

Backend:

MySQL

Algorithms:

Classification: Linear Regression

Market Basket Analysis: Apriori

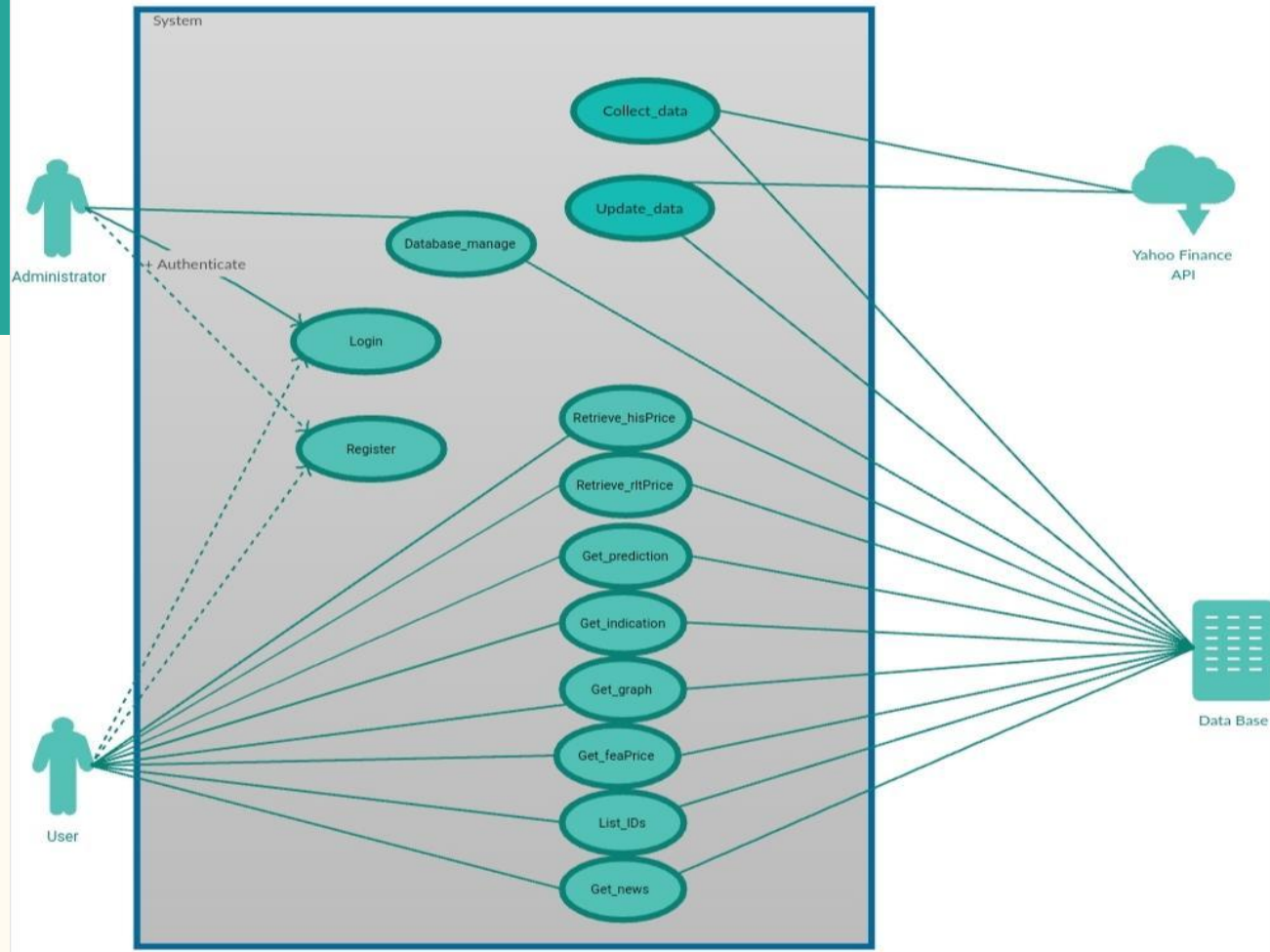
2. Proposed System

The proposed 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.

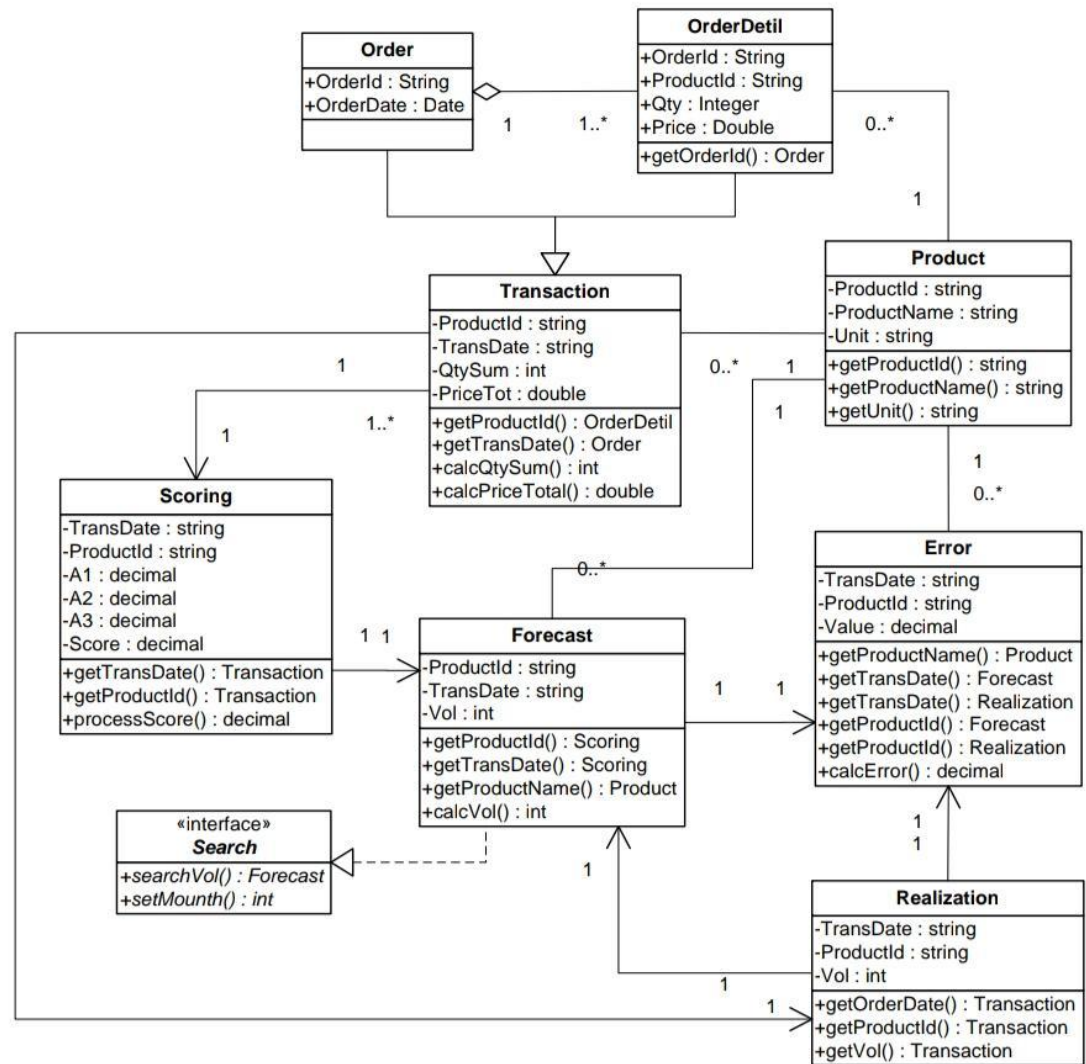
A collective analysis of the said reports will help retailers make business decisions which are convenient and cost effective.

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 .

2.1 Use Case Diagram



2.2 Class Diagram



2.3 Module 1

Login

- Every store has their login credentials. This ensures segregation of data.
- If the login credentials are incorrect access is denied.
- Sessions are managed through these credentials.

2.4 Module 2

Warehouse

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

2.5 Module 3

Billing

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

2.5 Module 4

Analytics

- Through this tab shopkeepers will get a graphical view of their sales data.
- Sales prediction is viewed through the line graph.
- Product analysis is given through PHPCharts.

2.6 Module 5

Prediction Model

- This prediction model is based on Linear regression.
- The previous sales data is given as input for the dependent variable and the quantity sold is used to compute the constant.
- .The independent variable thus generated is the forecasted sales.

2.7 Module 6

Association Rules

- Market Basket analysis is done through Apriori algorithm.
- The algorithm scans the database and searches for frequently paired products.
- The probability of these pairs is calculated and product recommendation is generated.

2.8 References

- 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
- 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
- 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
- A. Setiawan, G. S. Budhi, D. H. Setiabudi and R. 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 (ICSIIT), Denpasar, 2017

3. Planning for next semester

- Implementing Data mining Tasks
- Designing prediction model using Liner Regression algorithm
- Recommendation model design using Apriori algorithm.
- Enchancing security features.

Thank You