BACHELOR OF SCIENCE IN COMPUTER SCIENCE
AND ENGINEERING

Online Shopping System with Prediction Analysis

by

Md. Nasid Hasan

Jysan Aziz

Supervised by

Md. Moniruzzaman

Department of Computer Science and Engineering (CSE)
Islamic University of Technology (IUT)
November, 2016

Abstract

Online shopping is the process whereby consumers directly buy goods, services etc. from a seller interactively in real-time without an intermediary service over the internet. Our system has a special feature that is predictive analysis of each product based on past data analysis.

There is two different views in this system one is for user and another is for admin. In user view user can see all the products and choose among them. Here we try to implement a clothing system. The system has some categories like Man.Woman,Kids etc. Each category has also some subcategories. In admin view admin can update or delete any data. Admin can see the predictive analysis of each product. The unique feature of our system is the future prediction of each product. Here the prediction works based on the past and present data analysis of each product. Generally in a big system Big Data,Cloud computing etc. used for analyzing the huge number of data. But we have made a small protocol for less amount of data.In the KNN learning algorithm,the decision classifier will detect by itself according to the event and data. But in our small protocol we will build a system with static classifier,we will take the decision according to the static range.

Acknowledgment

At the very beginning We express our heartiest gratitude to Almighty Allah for His divine blessings which has allowed us to bring this project work to life.

We are grateful and indebted to our supervisor MD. Moniruzzaman, senior lecturer, Department of Computer Science and Engineering, IUT. His supervision, knowledge and relentless support have time and again proved to be invaluable and allowed us to complete this project successfully.

Lastly we would like to thank the faculty members of CSE Department, IUT who have helped to make our working environment a pleasant one, by providing a helpful set of eyes and ears when problems arose.

Table of Contents

Chapter 1

Introduction

In this chapter, we present a basic outline of our project including problem definition and its application. Then we include our project objectives.

1.1 Background

We present our work with an online retailer,as an example of how a retailer can use its wealth of data to optimize pricing decisions on a daily basis.One of the retailers main challenges is pricing and predicting demand for products that it has never sold before, which account for the majority of sales and revenue. To tackle this challenge, we use machine learning techniques to estimate historical lost sales and predict future demand of new products. Nonparametric structure of our demand prediction model, along with the dependence of a products demand on the price of competing products,pose new challenges on translating the demand forecasts into a pricing policy. We develop an algorithm to efficiently solve the subsequent multi-product price optimization that incorporates reference price effects,and we create and implement this algorithm into a pricing decision support tool for daily use.We conduct a field experiment and find that sales does not decrease due to implementing tool recommended price increases for medium and high price point products.

Helping to effectively balance a retailers supply and demand, demand forecasting software is a must-have solution that can make or break a season(s). By minimizing inventory investment and increasing revenue opportunities, best-in-class demand forecasting software enables a look into the future; forecasting

customer demand for new and current products and allowing for inventory level planning to meet demand, as well as creating the best possible replenishment orders. Demand forecasting tools can help you to minimize costly inventory errors and stock-outs.

An end-to-end process, demand forecasting is not a stand-alone activity. The fact is that actual customer demand needs to drive the entire business, as consumers today are more sophisticated, fickle and have many more options than just traditional brick and mortar which used to be the only choice available. Now, consumers have the ability to compare prices on the web, research product availability at the click of a mouse or on their mobile devices and more. Therefore, it is imperative for companies to utilize a demand forecasting software solution that can help them to overcome these new challenge.

1.2 Problem Statement

In a given online shopping system,we will focus on the predictive analysis of product which is included the revenue management system,pricing system and many other analytical questions. This thing is basically done by the administrative perspective. According to the given data and event analysis we will clarify the decision making. Determining discrete and dependent event,extracting data related to it and making decision clarifier table with the learning experience is the challenge. Main problem is developing this system with the desired classifier.Along with the classifier we have to also develop the whole system with proper design.Organizing the data is also a challenge for this problem.

1.3 Project Objective

- To shop while in the comfort of your own home, without having to step out of the door. One can buy all the things by sitting in home using the internet.

- Future prediction system on the administration side. Admin can see all the past records of a product and analyzing those records admin can get a predicted result of a product.Here the system will analyze the data and give the output result of certain question.

- By analyzing the past data and present data admin can easily understand the demand of a product in the market.So he or she can easily maintain the purchase amount of any product.

- Future predictive analysis helps the owner to make a decision about a product whether he needs to buy the product more or he should not buy the product anymore.

- It will also maintain an optimize system which will balance the consumer price satisfaction along with maximum profit for the administrator.

Chapter 2

Related Works

In this chapter, we described the all related works for accomplishing the future prediction of a product.We have two parallel system here.One is for front end the online shopping system and another is the backend part where admin will manage the inventory part.Our main focused part is here the future prediction which will be developed through a decision classifier.

2.1 User perspective

Usual online shopping system where user can choose and buy any product. Users can visit web stores from the comfort of their homes and shop as they sit in front of the computer. Consumers buy a variety of items from online stores. In fact, people can purchase just about anything from companies that provide their products online. Books, clothing, household appliances, toys, hardware, software, and health insurance are just some of the hundreds of products consumers can buy from an online store.That means user can only see the front end of the system.

Features for user

- User login
- Feedback
- Shopping cart
- Online Transaction

2.2 Admins perspective

Generally admin handle all the back end designing and contents handling.

2.2.1 Manipulation of the database

Admin can add or delete products from the database. Admin has all the power to add,update or delete any kind of data. Public demands varies day by day so he cannot run the system with the same data for whole life he needs to change the data after some time.And this changes can be done only by the admin.

2.2.2 Future prediction of a product

The unique feature of our system is the future prediction of each product. Here the prediction works based on the past and present data analysis of each product. Generally in a big system Big Data,Cloud computing etc. used for analyzing the huge number of data. But we have made a small protocol for less amount of data.In the KNN learning algorithm,the decision classifier will detect by itself according to the event and data. But in our small protocol we will build a system with static classifier,we will take the decision according to the static range.

2.3 Database Design

The basic design of database which is known as data dictionary is given below.The tables are generated through phpmyadmin localhost.

Table	Rows	Type	Collation	Size	Overhead
category	3	InnoDB	latin1_swedish_ci	16 KiB	-
decision	4	InnoDB	latin1_swedish_ci	16 KiB	-
details	2	InnoDB	latin1_swedish_ci	16 KiB	-
fdbk	3	InnoDB	latin1_swedish_ci	32 KiB	-
inventory	0	InnoDB	latin1_swedish_ci	16 KiB	-
items	3	InnoDB	latin1_swedish_ci	16 KiB	-
odersdetail	20	InnoDB	latin1_swedish_ci	16 KiB	-
orders	2	InnoDB	latin1_swedish_ci	16 KiB	-
orderscart	100	InnoDB	latin1_swedish_ci	16 KiB	-
product	10	InnoDB	latin1_swedish_ci	32 KiB	-
register	10	InnoDB	latin1_swedish_ci	32 KiB	-
subcategory	16	InnoDB	latin1_swedish_ci	16 KiB	-
trash	36	InnoDB	latin1_swedish_ci	16 KiB	-
13 tables	**209**	**InnoDB**	**latin1_swedish_ci**	**256 KiB**	**0 B**

category

Column	Type	Null	Default	Links to	Comments	MIME
cat_id *(Primary)*	varchar(30)	No				
category	varchar(50)	No				

Indexes

Keyname	Type	Unique	Packed	Column	Cardinality	Collation	Null	Comment
PRIMARY	BTREE	Yes	No	cat_id	3	A	No	

decision

Column	Type	Null	Default	Links to	Comments	MIME
Message	varchar(1000)	No				
Name *(Primary)*	varchar(300)	No				

Indexes

Keyname	Type	Unique	Packed	Column	Cardinality	Collation	Null	Comment
PRIMARY	BTREE	Yes	No	Name	4	A	No	

details

Column	Type	Null	Default	Links to	Comments	MIME
name *(Primary)*	varchar(30)	No				
pass	varchar(30)	No				

Indexes

Keyname	Type	Unique	Packed	Column	Cardinality	Collation	Null	Comment
PRIMARY	BTREE	Yes	No	name	2	A	No	

fdbk

Column	Type	Null	Default	Links to	Comments	MIME
name	varchar(30)	No				
phone no *(Primary)*	varchar(30)	No				
email	varchar(30)	No				
subj	varchar(30)	No				
mesg	varchar(30)	No				

Figure 2.2: Tables Design

Indexes

Keyname	Type	Unique	Packed	Column	Cardinality	Collation	Null	Comment
PRIMARY	BTREE	Yes	No	phone no	2	A	No	
email	BTREE	Yes	No	email	2	A	No	

inventory

Column	Type	Null	Default	Links to	Comments	MIME
Date *(Primary)*	date	No				
itemno	varchar(30)	No		items -> itemno		
sell_amnt	int(11)	No				
sold_tk	int(11)	No				

Indexes

Keyname	Type	Unique	Packed	Column	Cardinality	Collation	Null	Comment
PRIMARY	BTREE	Yes	No	Date	0	A	No	

items

Column	Type	Null	Default	Links to	Comments	MIME
catg	varchar(40)	No				
subcatg	varchar(40)	No				
img	varchar(30)	No				
itemno *(Primary)*	varchar(30)	No				
price	int(30)	No				
buy_price	int(11)	No				
stock	int(11)	No				
sold_qnty	int(11)	No				
desc	varchar(300)	No				
info	varchar(500)	No				
dat	varchar(100)	No				

Indexes

Keyname	Type	Unique	Packed	Column	Cardinality	Collation	Null	Comment
PRIMARY	BTREE	Yes	No	itemno	3	A	No	

odersdetail

Figure 2.3: Tables Design

Column	Type	Null	Default	Links to	Comments	MIME
productid	int(11)	No				
ordersid	int(11)	No				
price	float	No				
quantity	int(11)	No				

orders

Column	Type	Null	Default	Links to	Comments	MIME
pname	varchar(30)	No				
itemno	varchar(30)	No				
price	varchar(30)	No				
size	varchar(30)	No				
uname	varchar(30)	No				
ac_no	varchar(30)	No				
mob_no	varchar(30)	No				
add	varchar(300)	No				
bank	varchar(30)	No				
city	varchar(30)	No				
order_no	varchar(30)	No				

orderscart

Column	Type	Null	Default	Links to	Comments	MIME
id	int(10)	No				
name	varchar(100)	No				
datecreation	date	No				
status	tinyint(1)	No				
username	varchar(100)	No				

product

Column	Type	Null	Default	Links to	Comments	MIME
id	int(11)	No				
Name	varchar(200)	No				
day	int(11)	No				
month *(Primary)*	int(11)	No				
year	int(11)	No				
Sld_qnty	int(11)	No				
Remain_qnty	int(11)	No				

Figure 2.4: Tables Design

profit	int(11)	No				
Decision	varchar(1000)	No				

Indexes

Keyname	Type	Unique	Packed	Column	Cardinality	Collation	Null	Comment
PRIMARY	BTREE	Yes	No	month	10	A	No	
month	BTREE	Yes	No	month	10	A	No	

register

Column	Type	Null	Default	Links to	Comments	MIME
name	varchar(300)	No				
phone	varchar(13)	No				
email *(Primary)*	varchar(300)	No				
addr	varchar(300)	No				
pass	varchar(24)	No				

Indexes

Keyname	Type	Unique	Packed	Column	Cardinality	Collation	Null	Comment
PRIMARY	BTREE	Yes	No	email	10	A	No	
email	BTREE	Yes	No	email	10	A	No	

subcategory

Column	Type	Null	Default	Links to	Comments	MIME
cat_id	varchar(30)	No				
subcategory	varchar(50)	No				

trash

Column	Type	Null	Default	Links to	Comments	MIME
catg	varchar(50)	No				
subcatg	varchar(50)	No				
img	varchar(60)	No				
itemno	varchar(30)	No				
price	int(30)	No				
desc	varchar(300)	No				
dat	varchar(30)	No				

Figure 2.5: Tables Design

Chapter 3

The Proposed Mechanism

3.1 Introduction

This chapter of the project is about the different ways about the future prediction handling.

3.2 Predictive Analytics and Online Retail

To attempt understanding what predictive analytics could do to the retail industry it is vital for us to understand what predictive analytics is and the delta that it offers when compared to traditional Business Intelligence solutions that the whole retail industry was obsessed about when it came into the scene. Predictive analytics, in essence, is a process of developing data mining techniques that use analytical models to discover hidden patterns and apply them to predict future trends and behaviors. Predictive Analytics takes history into account to forecast and learn to foresee emerging patterns that can be used as a basis to take decisions for the future. The process of prediction involves the following steps: 1. Problem Identification 2. Determining the Outcome and Predictors 3. Explore data and segregate data 4. Test the models 5. Apply models to an identified population to predict behavior and evaluate. In most situations, analysts face the challenge of defining the business problem. Successful predictive business problems have quantitative goals.

Its an understatement to say that online retail is a competitive space. As a

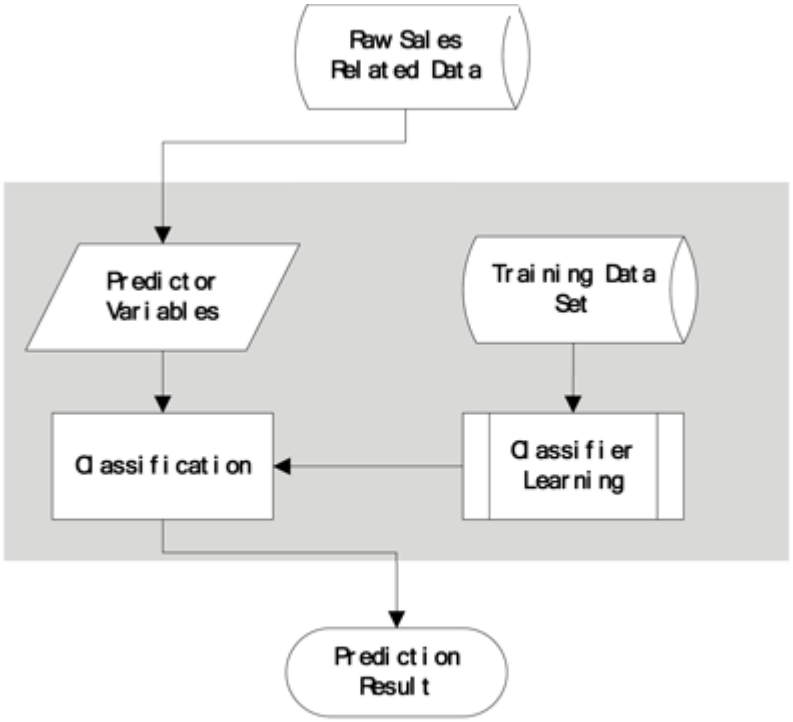


Figure 3.1: Prediction Analysis Proposed System

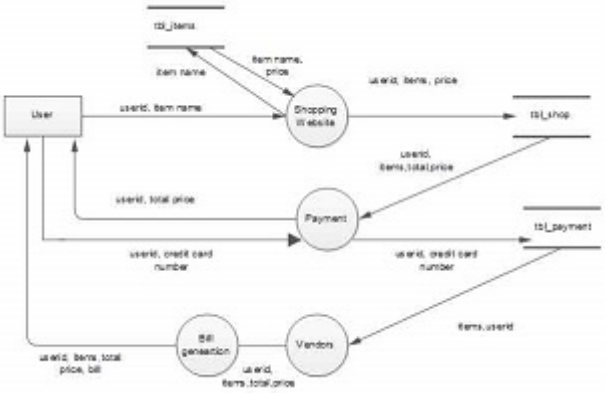


Figure 3.2: Online Retail System

recent Deloitte report points out, consumers are more empowered than ever. Todays consumer has vastly different and more sophisticated expectations of product, service, value, and environment than five or even three years ago, says Ian Geddes, Deloitte partner and head of UK retail.Shoppers want personalized ex-

periences, on-demand information, and unprecedented customer service. Retail leaders are under immense pressure to keep upit takes less than two minutes for unsatisfied customers to write negative reviews and flock to competitor websites. The future of online retail depends on customer connections. Well now we will give a simple outlook of online shopping system.

3.3 Key Analytical Shifts in Data

Retail giants rely on information about their customers to make strategic decisions.What retailers need to understand, however, is that data has value beyond information-gathering and surveillance. Retailers should leverage data to uncover and pursue untapped opportunities to drive growth and efficiency. The strategy is to target customers who are problem-aware, but who are not necessarily solution-aware. Millions of consumers visit e-commerce sites, but only few thousands visitors may buy the products. The E-retailer wants to improve the customer experience and would like to improve the conversion rate. The forecast of potential sites is based on sales data on shop level for existing stores and a broad variety of spatially aggregated geographical, socio-demographical and economical features describing the trading area and competitor characteristics. The model was built from a-priori expert knowledge and by analytic knowledge which was discovered during the data mining process.This linear classifier is called the Optimal Separating Hyperplane. Kernel functions are used to handle instances where the data points are not linearly separable which works by transforming the input space containing the training instances into a new, higher-dimensional feature space, in which it becomes possible to separate the data.

The data used in this approach focuses on online shopping data.he approach involves manipulating raw data into available forms and then a trigger model is proposed to do the classification. The classification result indicate the best prediction model for each item. Finally, by use of the most appropriate model, the prediction is accomplished.This approach involves applying two typical fore-

casting models and several dimensions to the trigger model through training and testing the classification model with real sales data and focuses on the correlation of two subjects and ignores the causal relationship between them.

The development of any search application in Exalead CloudView requires defining what data to include in the index schema, and then configuring one or more search logics to control how the documents are presented to users as search results. Index fields are used for searching and to display data in hit content in the results that are returned. Categories store static facet values. These values display in the Refinements panel of the search results as well as in hit content. Static facets allow users to narrow their search results by focusing on a certain aspect of the results, such as a particular country or product line, price value within a range etc.

3.4 Prioritizing the Point of Sale

Long-term shopping history, product preferences, and social media activity are all valuable customer data points. What matters most though is the retail point of sale. Is there a way to design a shopping experience that optimizes for that 30-second moment? Sales are the shining light of predictive analytics. Thats the idea behind Inkirus strategy to target active shoppers when theyre most ready to buy. Retailers are leading the pack in data collection from point-of-sale and customer loyalty data to demographics and are amassing huge data stores. Wal-Mart handles close to 1 million customer transactions per hour, which is imported into databases estimated to contain more than 2.5 PetaBytes of data - the equivalent of 167 times the information contained in all the books in the US Library of Congress. Unfortunately, collecting data is not enough. Tech savvy retailers are looking towards predictive analytics to unleash the power data. Access to the right data mining and predictive analytics solutions can help a retailer take insightful decisions in todays volatile economic climate. Retailers use predictive analytics to set the bar in customer retention, inventory optimization and low-

cost promotions which drive increases in profitability and market share. There can be created a set of property types like alphanumeric, numeric, geographic. After scanning the data sources, the settings which have been defined for the properties generate new index fields and categories inside the index schema and also several meta elements or prefix handlers according to the data logic. When a configuration is applied, the high-level view provided by the Data Model is expanded into the multiple index and search elements. A common challenge when creating properties is to know which metas belong to the corpus. There are several ways in which the available metas can be explored using the CloudView data model. When creating a data model property, it can be chosen one of the following field types: either an index field or a category 316 316 facet only, or both.

3.5 Building 1:1 Relationships

Branding is mission-critical to sales. But how do companies pinpoint an exact brand persona? The answer is straightforward: address exact shopper needs. The problem is, though, that shoppers come from many walks of life.Thats why Nordstrom, for example, employs a rigorous customer segmentation process. Statistical tools help marketers make sense of their customer segments to deliver personalized messaging. Data helps companies become everybodys brand.

3.6 Optimize Goals

The success of a data strategy depends on your analytical framework. Make data make sense, and leverage predictive analytics to address a specific company need.Think of the benefits of understanding demand and being able to more effectively manage purchasing, planning, assortment, storing and not to mention, optimization and discounts.Its all about understanding the consumers lifetime value.This business-focus should guide your consumer analytics team.

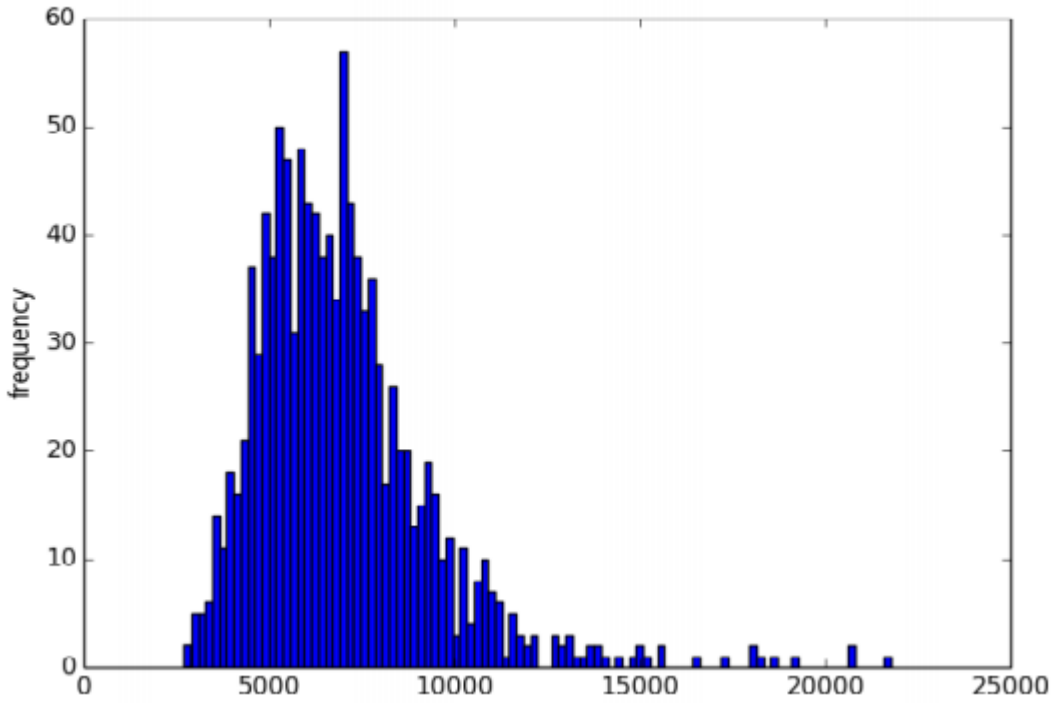


Figure 3.3: Online Retail System

Data Description 1) Store - Each store in the dataset has a unique ID associated with it.

2) Sales - The turnover for a store on a given day.

3) Customers - the number of customers who visited the store on a given day.

4) Open - indicates whether the store was open(0) or closed(1).

5) StateHoliday - indicates a state holiday. There are 4 classes -¿ a = public holiday, b = Easter holiday, c = Christmas, 0 = None.

6) SchoolHoliday - indicates if the was affected by the closure of public schools. These holidays vary from state to state.

7) StoreType - differentiates between 4 different store models: a, b, c, d. Different kinds of stores sell different products.

8) Assortment - describes an assortment level: a = basic, b = extra, c = extended. Indicates the variety in items the store sells.

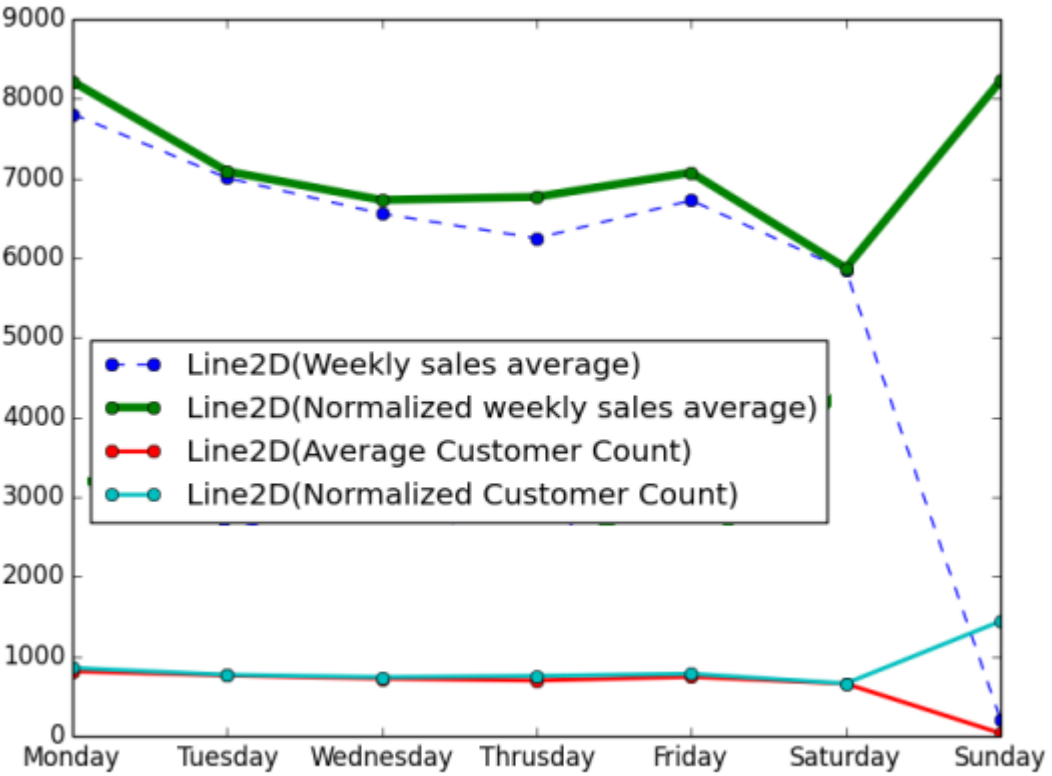


Figure 3.4: Online Retail System

9) CompetitionDistance - distance in meters to the nearest competitor store.

10) CompetitionOpenSince[Month/Year] - The year and month the nearest competitor was opened.

11) Promo - indicates whether a store is running a promo on that day : 1 indicates a promo, 0 indicates no promo.

12) Promo2 - a continuing and consecutive promotion for some stores: 1 = store is conducting the promo, 0 = store is not conducting the promo.

13) Promo2Since[Year/Week] - the year and week when the store started participating in Promo2.

14) PromoInterval - Promo2 runs during certain months of the year, this field indicates this event.

15) DayOfWeek - Varies from Monday to Sunday. Most stores are closed on Saturday and Sunday.

Fig 1 depicts the histogram of the mean sales per store when the stores werent closed. From this graph we can infer that most stores have very similar sales and that there is a small percentage of outliers.
Fig 2 shows how the number of customers and sales average vary based on the day of the week taken into consideration. As expected average sales tend to be more on the Sunday as compared to other days.

3.7 Summary

These are the different methods by which one can do the predictive analysis of a online shopping system.In our project we build a small protocol based on small amount data.In the KNN learning algorithm,the decision classifier will detect by itself according to the event and data. But in our small protocol we will build a system with static classifier,we will take the decision according to the static range.

Chapter 4

Implentation

Here we will give the view of our implemented part. Basically there are two side on implementation.One is user side, another is administrator side.The view will be following.

4.1 User View

The front end part, comprising of pages which users can view, is created by using templates so that it can expedite design and loading processes. This part, in the appearance, the main menu is consistent, thus it is easy to use. Considered as the most frequently access when comparing with other parts of the web site, front end pages have to use template, and in turn, saving bandwidth and number of database accesses. For database operation issue, aiming for minimizing number of accesses, it needs just four queries to display all details. Those queries consist of jointed SQL statements, which inquire merely main information from the database.The basic view and the implemented part will be given below.

4.2 Admin View

This part provides facility for each store owner to edit and modify information in his own store. Providing validation check for member,store identification and future prediction the back end system can securely protect users proprietary information. In addition, all page views employ session variables to deter manually defined variables by users. Applying user friendly approach, and focusing on web

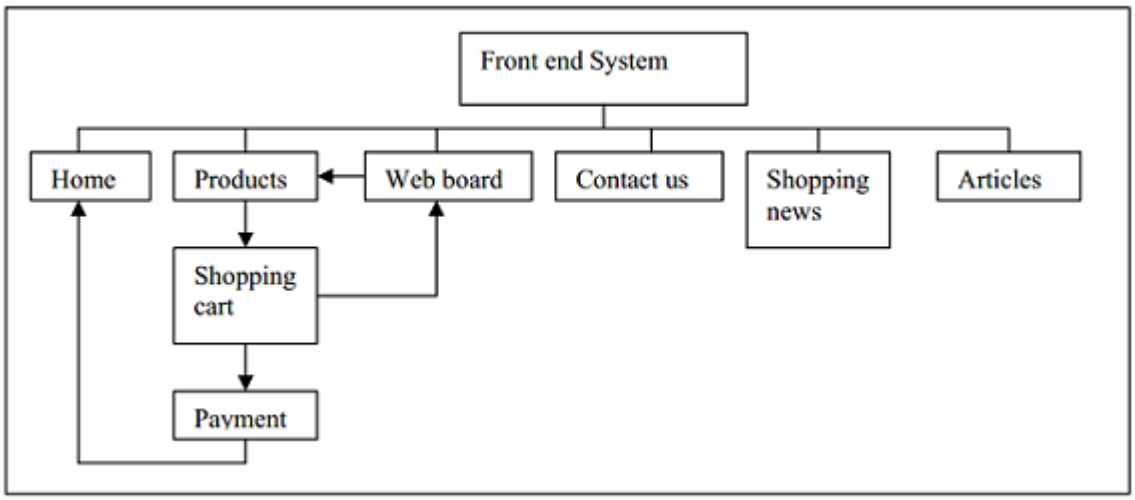


Figure 4.1: Online Retail System Front End Design

(a) Homepage

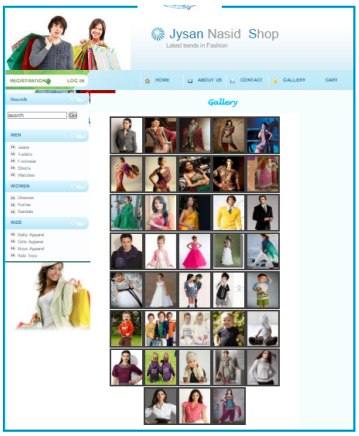
(b) Category wise view

Figure 4.2: User view

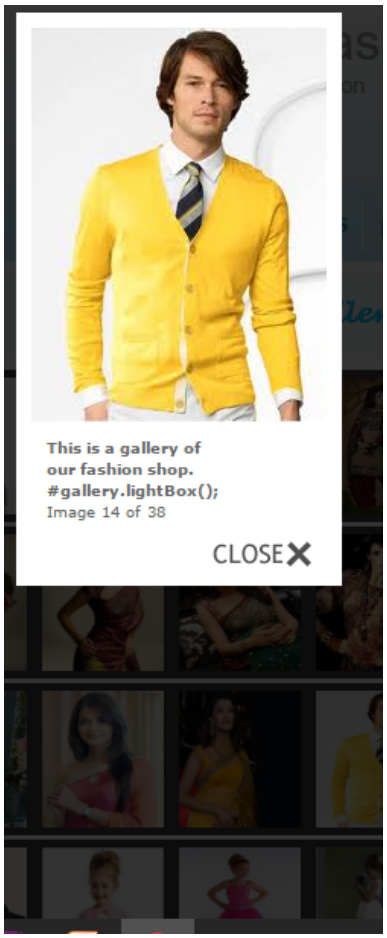
programming inexperience, the user can effortlessly manage his back end information. Inside the back end, users can control and view all store information. Beside that, using content management design, the back end part encompasses with these modules: product management, shop information management, web board management, shopping cart management, member management, promotion management, banner management, plaza management, poll management, currency management and article management.

(a) View page of product

(b) User Login

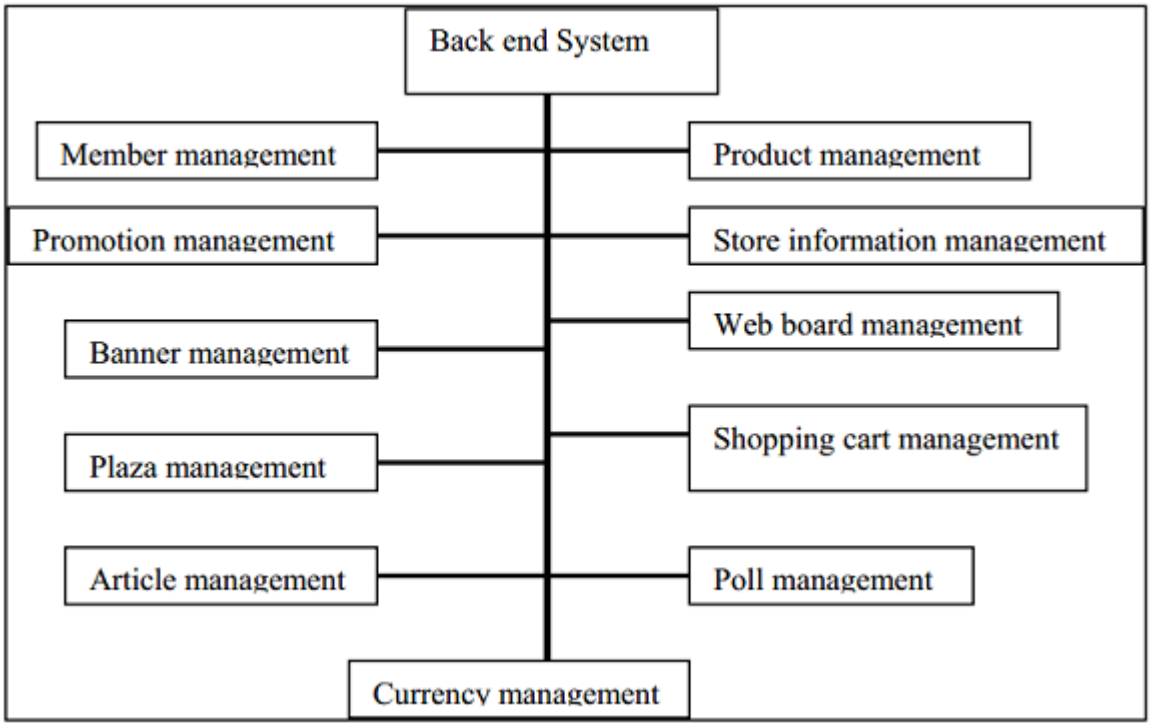


(c) Gallery of Products



(d) Close view on gallery

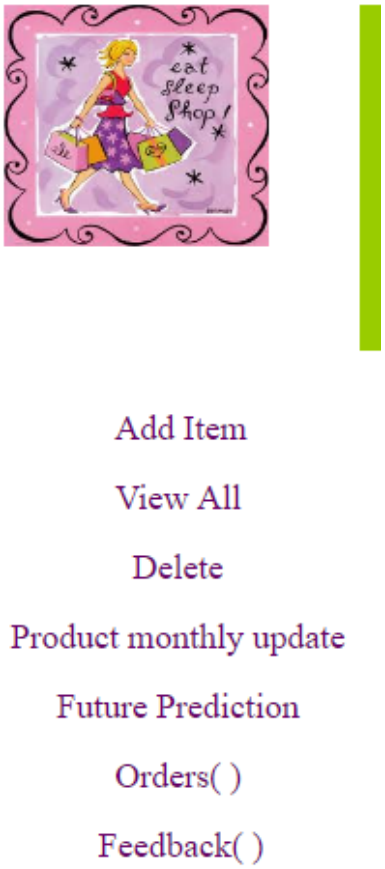
Figure 4.3: User view



Online Retail System Back End Design

(a) Admin Login

(b) Adding product

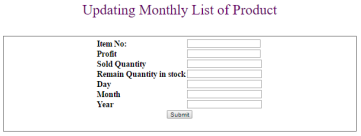


(d) Menu List of Admin

(c) Admin view of Product

Figure 4.4: Admin view

(a) Deleting item



(b) Monthly Update of Products

(c) Future Prediction of a Product

(d) Predicted System Output

Figure 4.5: Admin view

Chapter 5

Conclusions

5.1 Summary of the Project

Online shopping is the process whereby consumers directly buy goods, services etc. from a seller interactively in real-time without an intermediary service over the internet. Admin can add or delete products from the database. Admin has all the power to add,update or delete any kind of data.The unique feature of our system is the future prediction of each product.There are many ways to handle the future prediction of a online shopping system.Prioritizing the Point of Sale,Matchmaking through Algorithms,Optimize Goals etc. are the ways to handle the future prediction of products. Here the prediction works based on the past and present data analysis of each product. Generally in a big system Big Data,Cloud computing etc. used for analyzing the huge number of data. But we have made a small protocol for less amount of data.In the KNN learning algorithm,the decision classifier will detect by itself according to the event and data. But in our small protocol we will build a system with static classifier,we will take the decision according to the static range.

5.2 Future Work

Here we have build the online shopping system with predictive analysis based on small amount of data.In our project we build a small protocol based on small amount data.In the KNN learning algorithm,the decision classifier will detect by itself according to the event and data. But in our small protocol we will build

a system with static classifier,we will take the decision according to the static range.This is not suitable for huge amount of data.We researched about many methods by which the future prediction can be handled.So in future we will be able to handle large amount of data.We will try our best to improve this project in such a way that it will work for large amount of data.

Bibliography

[1] Huailin Dong, Lingwei Xie, Zhongnan Zhang, "Research on Statistics-based Model for E-commerce User Purchase Prediction"

[2] Bisheng DuChristian Larsen and Alan Scheller-Wolf, "Optimal Inventory Decisions in a Capacitated Online Retailing System with Customer Priorities"

[3] Mona Gupta , Happy Mittal , Parag Singla , Amitabha Bagchi,"Characterizing Comparison Shopping Behavior: A Case Study"

[4] Tian-Hsiang Huang,, Jian-De Lu and Vladimir Nikulin."Mining Shoppers Data Streams to Predict Customers Loyalty",2015 International Conference on Intelligent Systems and Knowledge Engineering

[5] Tiger Tyagarajan, Building a Maniacal Customer-centric Culture. White Paper of Genpact Limited. Available: http://www.genpact.com/docs/resource-/building-a-maniacalcustomer-centric-culture

[6] J. Ben Schafer, Joseph A. Konstan and John Riedl, E-commerce recommender application, Data mining and knowledge discovery, vol. 5, pp. 115 153, 2001.

[7] S. Kok, M. Sumner, M. Richardson, P. Singla, H. Poon, D. Lowd, J. Wang, and P. Domingos, The Alchemy system for statistical relational AI, University of Washington, Tech. Rep., 2008, http://alchemy.cs.washington.edu.

[8] Wedad Elmaghraby and Pinar Keskinocak,"Dynamic Pricing in the Presence of Inventory Considerations: Research Overview, Current Practices and Future Directions"