Prediction Of Consumer Purchasing In Retail Store

Submitted By

Shivani Dholakiya

16MCEC04

DEPARTMENT OF COMPUTER ENGINEERING

INSTITUTE OF TECHNOLOGY

NIRMA UNIVERSITY

AHMEDABAD-382481

May 2018

Prediction Of Consumer Purchasing In Retail Store

Major Project

Submitted in partial fulfillment of the requirements

for the degree of

Master of Technology in Computer Science and Engineering

Submitted By

Shivani Dholakiya

(16MCEC04)

Guided By

Dr. Zunnun Narmawala

DEPARTMENT OF COMPUTER ENGINEERING

INSTITUTE OF TECHNOLOGY

NIRMA UNIVERSITY

AHMEDABAD-382481

May 2018

Certificate

This is to certify that the major project entitled **"Prediction Of Consumer Purchasing In Retail Store"** submitted by **Shivani Dholakiya (16MCEC04)**, towards the partial fulfillment of the requirements for the award of degree of Master of Technology in Computer Science and Engineering of Nirma University, Ahmedabad, is the record of work carried out by him under my supervision and guidance. In my opinion, the submitted work has reached a level required for being accepted for examination. The results embodied in this major project part-I, to the best of my knowledge, haven't been submitted to any other university or institution for award of any degree or diploma.

Dr. Zunnun Narmawala
Guide & Associate Professor,
CE / IT Department,
Institute of Technology,
Nirma University, Ahmedabad.

Dr. Priyanka Sharma
Professor,
Coordinator M.Tech - CSE
Institute of Technology,
Nirma University, Ahmedabad

Dr. Sanjay Garg
Professor and Head,
CE Department,
Institute of Technology,
Nirma University, Ahmedabad.

Dr Alka Mahajan
Director,
Institute of Technology,
Nirma University, Ahmedabad

Statement of Originality

I, **Shivani Dholakiya**, **16MCEC04**, give undertaking that the Major Project entitled "**Prediction of consumer Purchasing In Retail Store**" submitted by me, towards the partial fulfillment of the requirements for the degree of Master of Technology in **Computer Science & Engineering** of Institute of Technology, Nirma University, Ahmedabad, contains no material that has been awarded for any degree or diploma in any university or school in any territory to the best of my knowledge. It is the original work carried out by me and I give assurance that no attempt of plagiarism has been made.It contains no material that is previously published or written, except where reference has been made. I understand that in the event of any similarity found subsequently with any published work or any dissertation work elsewhere; it will result in severe disciplinary action.

_____

Signature of Student

Date:

Place:

Endorsed by

Dr. Zunnun Narmawala

(Signature of Guide)

Acknowledgements

It gives me immense pleasure in expressing thanks and profound gratitude to **Dr. Zunnun Narmawala**, Associate Professor, Computer Engineering Department, Institute of Technology, Nirma University, Ahmedabad for his valuable guidance and continual encouragement throughout this work. The appreciation and continual support he has imparted has been a great motivation to me in reaching a higher goal. His guidance has triggered and nourished my intellectual maturity that I will benefit from, for a long time to come.

It gives me an immense pleasure to thank **Dr. Sanjay Garg**, Hon'ble Head of Computer Engineering Department, Institute of Technology, Nirma University, Ahmedabad for his kind support and providing basic infrastructure and healthy research environment.

A special thank you is expressed wholeheartedly to **Dr. Alka Mahajan**, Hon'ble Director, Institute of Technology, Nirma University, Ahmedabad for the unmentionable motivation he has extended throughout course of this work.

I would also thank the Institution, all faculty members of Computer Engineering Department, Nirma University, Ahmedabad for their special attention and suggestions towards the project work.

- Shivani Dholakiya

16MCEC04

Abstract

In Retail Industry, to maintain and increase sales, retailer has to do some observation based on past sale data. But, sometimes instead of increase in sales, retailer faces exponential decrease in sales due to some parameters like, unsatisfied customers due to product quality, behavior of store attendees, etc. Decrease in sales causes loss to retailer. To retain and increase sales and productivity by Customer satisfaction, Oracle Retail provides solution i.e. "Oracle Retail Sales & Productivity and Loss Prevention" to Retailers.In this thesis, I have worked on part of the solution i.e. "Prediction of Consumers Purchasing Pattern in Retail Stores" using data mining technique called Association Rule Mining. Based on past consumers purchasing transactions sample data. Using this prediction we can generate individual report for a particular product and link those individual reports based on different filters and parameters to predict likely-hood of buying particular product with other specific product or set of products . That will help to analyze the area where retailer can improve sales productivity and reports also help us identify unusual occurrences of transactions which shows fraud occurrence in retail store.

Abbreviations

POS	Point Of Sale
KPI	Key Performance Index
OJET	Oracle Javascript Extended Toolkit
JRAF	JET Retail Architecture Framework
BI	Business Intelligence
ORPOS	Oracle Reatail Point of Sale

Contents

List of Figures

Chapter 1

Introduction

1.1 Company Introduction

Oracle RGBU is Retail Global Business Unit which covers the retail-industry providing merchandise management. Oracle RGBU provides solution to meet customer satisfaction and run business in profitable margin. RGBU provides best-of-breed business applications, Open, integrated, cloud services which includes daily merchandising activities, including distribution, purchasing,financial close and order fulfillment. Customers of Oracle RGBU are leading fashion, grocery and specialty retailers, hotels, food and beverage facilities providers, etc.

Oracle Retail's Software and hardware products empowers retailers to predict market shifts with placing and data science, simplifies operations and delivers an extraordinary quality customer experience across channels.

Oracle Reatil has mainly 5 channels[3] :-

1. Merchandising
2. Omni channel
3. Planning
4. Supply chain
5. Retail Learning subscriptions

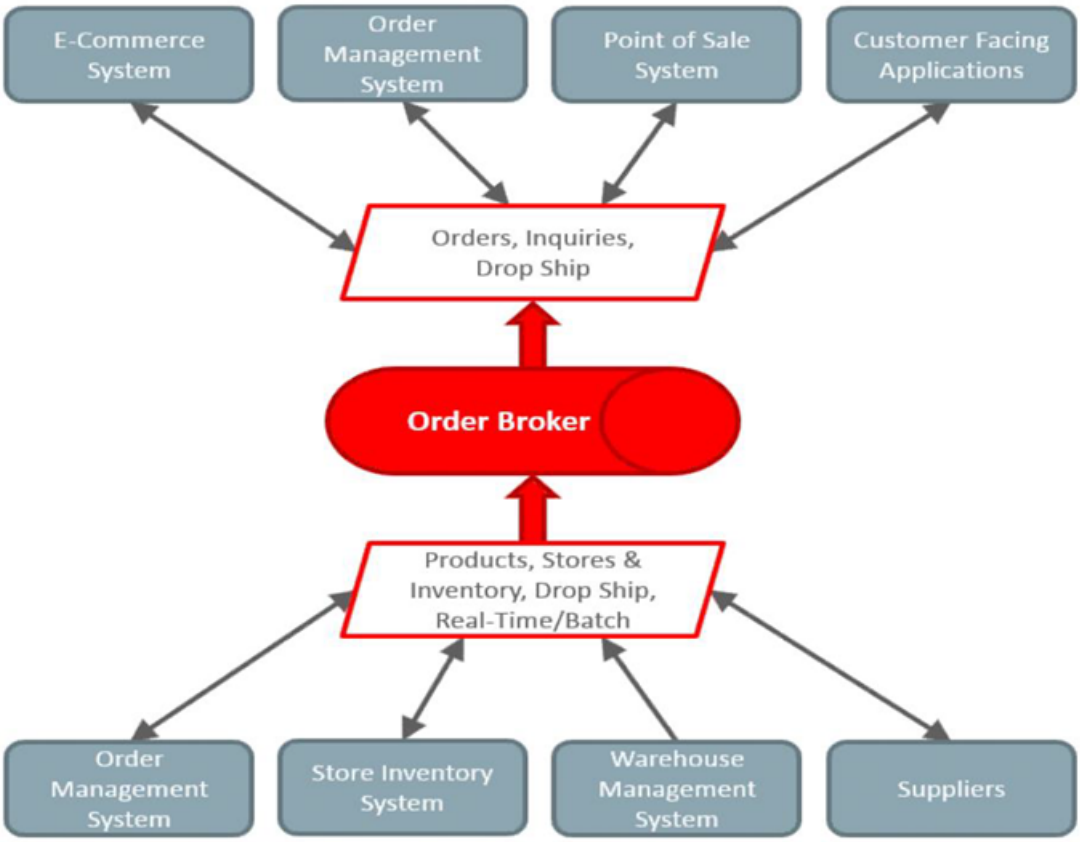

Figure 1.1: Retail Chain

As shown in Fig. 1.1, For each area Oracle Retail Omni channel provides Solution. Some of the products of omni-channel are mentioned bellow.

1. Oracle POS- Robust point-of-service functionality and platform enables retail staff to deliver on brand promise in store with inventory visibility, customer intelligence, and seamless transactions.

2. Oracle Store Inventory Management- ORSIM enables real-time, accurate and accessible store-stock data to execute commerce anywhere through consistent, efficient and effective in-store processes and procedures.

3. Order Broker - Order broker Access real-time inventory information to complete transactions on demand through alternative fulfillment methods, such as shipping directly from another warehouse or store or scheduling a store pickup.

4. XBRi Sales & Productivity and Loss Prevention - This product is used to increase retailer sales and reduce loss and prevention of causes of loss in Retail Industry. This product is used to analyze exceptions can happen at retail stores

5. Brand Compliance-Brand Compliance enables brand owners to collaborate with their supply chain in the sourcing, development, marketing, and quality control of their products using a scalable suite of integrated modules.

Oracle Retail Omni-channel provides total 14 such solution to Retailer to run their business in profitable margins.

1.2 Problem Statement

In Retail Industry, to maintain and increase sales, retailer has to do some observation based on past sale data. But, sometimes instead of increase in sales, retailer faces exponential decrease in sales due to some parameters like, unsatisfied customers due to product quality, behavior of store attendees, etc. Decrease in sales causes loss to retailer. To retain or increase sale Retailers ask Oracle Retail for the solutions.[3]

1.3 Objective

Objective of this project is to provide solution to meet customer satisfaction and run retail business in profitable margin.

1.4 Scope

The solutions given by the organization are aimed to cover whole retail industry.

1.5 My Role in Organization

I am working on the Omni-channel product which gives solution to retailer to increase their sales and productivity and loss prevention by analyzing consumers different activities. I am working with development team of product. I am working as full-stack developer.

Chapter 2

Tools And Technologies Used

2.1 Oracle JET

Oracle JavaScript Extension Toolkit (JET) certifies developers by providing a modular open source toolkit based on modern HTML5, CSS3 and JavaScript design and development principles. [1]

Oracle JET is used at intermediate to advanced JavaScript developers who are working on client side applications. It is a collection of open source JavaScript libraries along with a set of Oracle defined JavaScript libraries that makes it as simple and efficient because it possibles to build applications that consume and communicate with Oracle products and services, mostly Oracle Cloud services.[1]

2.1.1 Features & Benefits of OJET [1]

- Complete JavaScript development toolkit
- Leverages popular open-source technologies
- Full lifecycle management for template based SPA
- Built in accessibility support
- Support for internationalization (28 languages and 180+ locales) Rich set of UI components
- Advanced two-way binding with a common model layer

- Powerful routing system supporting single-page application navigation
- Smart resource management
- Built-in mobile support
- Rich set of UI components

2.2 JRAF

JRAF is an acronym for JET Retail Application Framework. JRAF provides a base framework for retail applications built on JET enabling an accelerated development for their client side development. JRAF provides a basic UI Shell and some Navigation/Collaboration components which are implemented based on the Standards and the Patterns Guide published by the RGBUS UEX team.
JRAF augments the capabilities of Oracle JET to include JS libraries that will help teams develop views (workflows) and other UI components with consistent style and behavior. Retail Application Team leveraging JET framework are expected to develop application navigational workflows using JRAF.JRAF would enable applications to render a web view of business application workflows within its UI Shell.

2.3 Cloud Services

A cloud service is any service made available to users on demand via the Internet from a cloud computing provider's servers as opposed to being provided from a company's own on-premises servers. Cloud services are designed to provide easy, scalable access to applications, resources and services, and are fully managed by a cloud services provider.

A cloud service can dynamically scale to meet the needs of its users, and because the service provider supplies the hardware and software necessary for the service, theres no need for a company to provision or deploy its own resources or allocate IT staff to manage the service. Examples of cloud services include online data storage and backup solutions, Web based e-mail services, hosted office suites and document collaboration services, database processing, managed technical support services and more.

2.4 Java Script

JavaScript often abbreviated as JS, is a high-level, dynamic, weakly typed, object based, multi-paradigm, and interpreted programming language. Alongside HTML and CSS, JavaScript is one of the three core technologies of World Wide Web content production. It is used to make web pages interactive and provide online programs, including video games. The majority of websites employ it, and all modern web browsers support it without the need for plug-ins by means of a built-in JavaScript engine.

2.5 Knockout

Knockout is a JavaScript library that helps you to create rich, responsive display and editor user interfaces with a clean underlying data model. Any time you have sections of UI that update dynamically (e.g., changing depending on the users actions or when an external data source changes), KO can help you implement it more simply and maintainability.

2.6 Database Oracle 12 C

Oracle Database (commonly referred to as Oracle RDBMS or simply as Oracle) is an object-relational database management system produced and marketed by Oracle Corporation. The Oracle Database 12c is a high-performance, enterprise-class database. According to Oracle, this is "the first database designed for the cloud." Oracle Database 12c also introduces 500 new features to the database, most notably plug-gable databases and multitenant architecture. The project used plug-gable databases and since the product is cloud enabled, Oracle 12c is preferred over other databases.

2.7 Web Logic 12.1.2

This is the server used to run the code in the eclipse IDE. WebLogic Server contains Java 2 Platform, Enterprise Edition (J2EE) technologies. J2EE is the standard platform for developing multi-tier enterprise applications based on the Java programming language. WebLogic Server is an application server which is based on JavaEE(Enterprise Edition).it

has framework bolsters the sequence of many types of circulated programs and is ca-pable of building applications in view of Service Oriented Architectures (SOA).

2.8 Supported Client Browsers and Mobile Devices

Browsers	Comments
Internet Explorer 11	Compatibility view is not supported
Safari (iPad)	Basic analyst functionality supported on iPad
Firefox	Version 46.0
Google Chrome	Version 50.0
Mobile Devices	**Comments**
Apple iPad	iOS8, iOS9, iOS10

Table 2.1: Browsers And Mobile devices supported

2.9 Machine Learning techniques

Machine learning is a field of software engineering that gives Computers the capacity to learn without being unequivocally customized.Machine learning is firmly identified with computational insights, which likewise concentrates on forecast making using PCs.

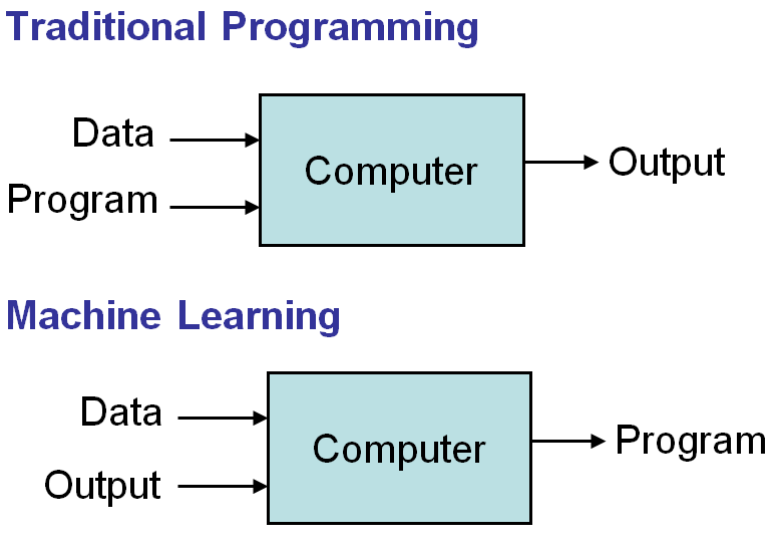

Figure 2.1: Machine Learning

Machine learning is a sort of artificial intelligence (AI) that enables programming applications to wind up noticeably more exact in foreseeing results without being expressly

modified.

Process of Machine Learning and Data mining are similar with each other.Both require hunting through information to look down examples and altering program activities as needs be. Many individuals know about machine gaining from shopping on the web and being served promotions identified with their buy. This happens in light of the fact that suggestion motors utilize machine figuring out how to customize online advertisement conveyance continuously. Past customized promoting, other basic machine learning use cases incorporate misrepresentation location, spam sifting, arrange security risk identification, prescient support and building news encourages.

There are major two technique of machine learning

1. Supervised Learning - This technique of machine learning is used to train the machine on bases of past experiences i.e. by analyzing past dataset and generating some pattern for prediction. In supervised Learning, we label the data.

2. Unsupervised Learning - In this technique of machine learning interactive approach is used.Unsupervised learning is a sort of machine learning calculation used to draw deductions from datasets comprising of info information without named reactions.In unsupervised learning, we do not label dataset. This technique is mostly used in clustering

3. Reinfocement Learning - Fortification learning is a type of machine learning propelled by behaviorist brain science, worried about how programming operators should take activities in a domain in order to amplify some idea of total reward.

Chapter 3

Product Introduction

Oracle Sales Productivity and Loss prevention is Omni-channel Product. Which is used to prevent loss of retailer by increasing in sales and reducing causes of Loss.
In Retail industry, Retailer need to look after the analysis of Sales over period. Because ultimately, for retail industry to run the store profitably is main motive.For the retailer to run their business in profit oracle retail provides solution called "Sales & Productivity and Loss Prevention"
Product has mainly 2 Module.

1. Sales & Productivity
2. Loss Prevention

Above two modules can be buy individually and retailer can also buy both of the modules in integration.

3.1 Sales and Productivity

Sales & Productivity module offers robust and highly configurable reporting across all levels of the retail organization hierarchy (Salesperson, Store, District, Region, and so on), merchandise hierarchy (item, class, dept., and so on), and/or by geographic attributes. Through a comprehensive set of grid and graph reports, documents and interactive dashboards, users can compare same store sales to past performance and custom goals, measure sales members productivity, and evaluate the impact of merchandise characteristics

on productivity.

Sales and Productivity are further divided 4 sub-modules.

1. Merchandise Productivity - Merchandise Productivity provides retailers with sales analysis across their merchandising hierarchy. A merchandise (SKU Master) table has been incorporated, which provides the levels of summarization needed within these hierarchies along with added attributes associated with the item such as vendor or manufacturer. This allows retailers to analyze their merchandise sales by categories and by items across their merchandise and operational hierarchies. It also allows for Margin analysis as well as, Return, Discount, and Voids summary information through these same hierarchies.

2. Salesperson/Employee Productivity - Salesperson Productivity reporting tracks key sales and productivity KPIs. These KPIs are attuned to a true selling environment in which employees are held accountable for sales and the quality and focus of what they sell. Another component of salesperson productivity is measuring sales against hours worked. Hours worked can be classified as selling or non-selling hours. These metrics are used to create and monitor sales/labor hour comparatives. A comprehensive set of new reports is provided in the Productivity category, as well as a new Salesperson Productivity dashboard. Using the new Salesperson Custom Stats page in Administration, Project Defaults, users can identify up to 10 custom count, transaction count, and amount statistics to use in reporting.

3. Sales Flow by Period - The Store Flow by Period components of this module provide customers with added business analytics focusing on the flow of sales transactions throughout business days, by hour, or by day part custom categorizations as well as by traffic counts and conversion rates.

4. Comparative Sales - The purpose of the Comparative Store Sales category is to report on sales by store by day. These sales include This Year vs. Last Year (TY/LY) transformation metrics, based on retailers predefined fiscal calendars. A second set of transformation metrics track comparative (Comp), or same store sales. These metrics add further logic to the TY/LY comparison, by comparing store sales only to when the store was open at the same time this year and last; with additional business logic. The Comp settings are maintained within the application

using a new Store Status page in Administration, Project Defaults. In addition, the customer can define and load sales goals or budgets to compare a stores sales performance to as many as three different sales goals. This is accomplished within the application, using a new Upload Goals/Sales page in Administration, Project Defaults. The goal file can also be loaded from a customer provided file feed.

3.2 Loss Prevention

The Loss Prevention module is an intuitive, intelligent and global analytical reporting solution that is designed to quickly identify suspicious trends, transactions, and other data anomalies. The Loss Prevention module allows easy user access, dynamic functionality, and forensic analysis to make more-informed decisions with timely, data-driven answers to business questions and to protect the bottom line.[4]

3.3 Functionality & Features

Functionality of Sales & Productivity and Loss Prevention as listed bellow.[5]

- Real-Time Analytics
- Dynamic Core Reporting
- Smart Links
- Interactive dashboards
- Self Service BI
- Mobile App

3.3.1 Real-Time Analytics

It provides real-time processing to support intra-day flash sales reporting for the new Sales and Productivity module. This is accomplished by processing data in specified time increments throughout the day rather than once at the end of day. Additional business logic supports the inclusion of post voids, clock in, clock out, and adjustments to normal business. Additional Loss Prevention analysis followed by no sale and no match

transactions will be updated through the end of day process. Here the implementation of Real Time Processing:

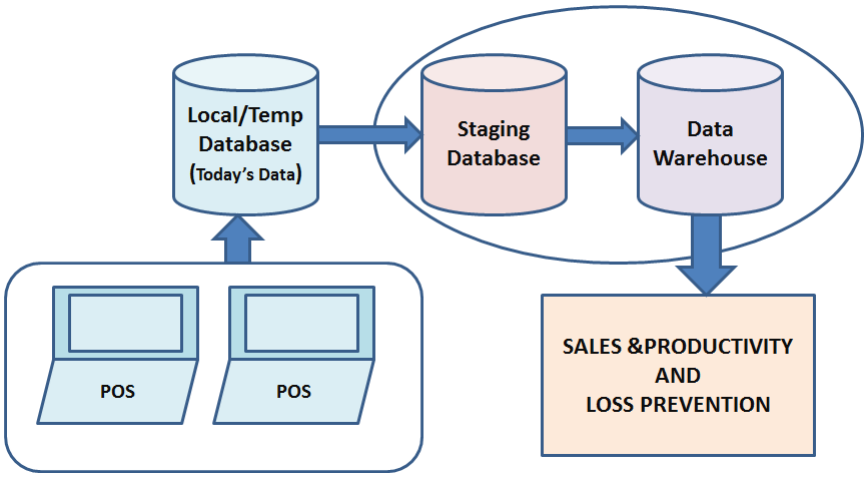

Figure 3.1: Real Time Analysis

In Real-time analysis ELT Flow also involved for real time data analysis on cloud. Data flow take major 4 steps for completion:

1. In first step, it will extract all the data from the customer transaction.
2. Then, it will lead to deliver data on Cloud through a protocol called SFTP.
3. Data available on cloud loaded into local database of system
4. Loaded data now transformed to a valuable format.

3.3.2 Dynamic Core Reporting

Product is delivered with 300+ core reports organized logically by category and by module of Sales & Productivity or Loss Prevention. Generated all Reports are soft and can be modified by users.

3.3.3 Smart Links

Smart links are informational window that provides key data about store, cashier, employee and other master data.

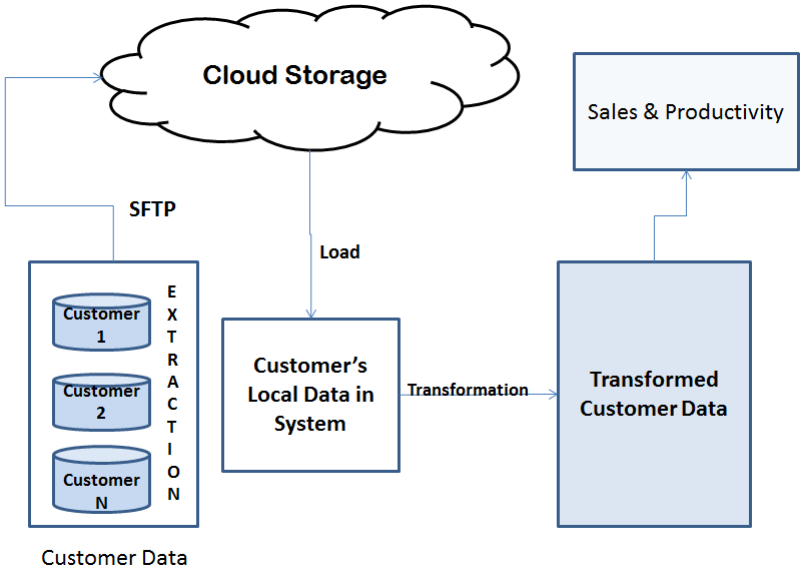

Figure 3.2: ELT Flow [2]

It provides Hover over key data for pop up information window and provides information. we can also able to configure for additional details from master tables.

3.3.4 Interactive dashboards

Dashboard was designed to summarize controls and exceptions in order to provide Loss Prevention (LP) leadership with information on what exceptions are occurring, how they are being addressed, and who may be handling them. The dashboard has a selection of date ranges to summarize on exceptions; looking at total generated, break down by status, type, where they are occurring, and who is assigned to investigate them.
Employee Violations Dashboards (EVDs) are a unique distribution of cashier/employee alerts. When an adhoc report is scheduled with an alert filter in XBR Desktop or Web, an EVD will be generated for each cashier that exceeded the alert threshold. The EVD is a report that is distributed via email in PDF format to each recipient assigned to the scheduled query run. EVD distribution can be enabled at the run or user level. The content of the EVD is dynamic and will be dependent on the KPI used in the alert filter.

- Report header with name of chain, location, and employee
- Summary data for the cashier of the key metrics related to the alert threshold

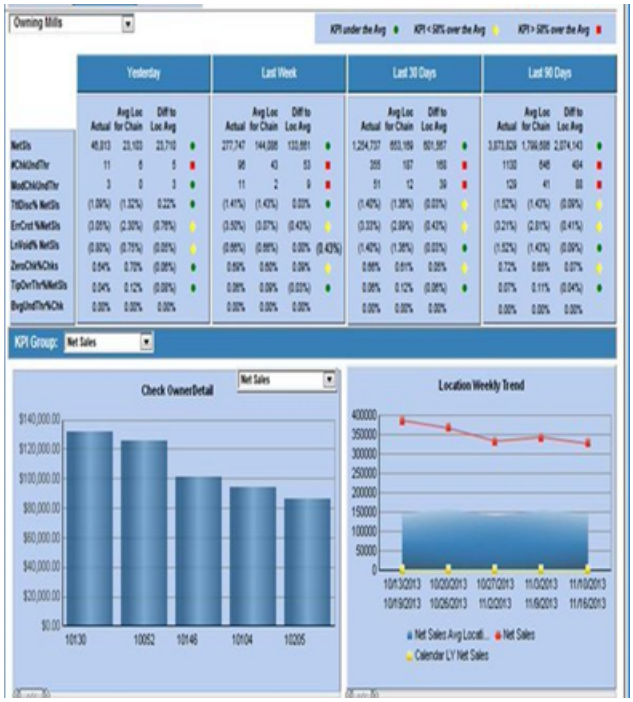

Figure 3.3: Report generation and Linking

- Averages for the other cashiers in the store
- Averages for all cashiers in the chain
- Description for the KPI constraint reached or action steps
- Alert history for the cashier
- watch status and note history for the cashier
- Hyperlink to the Web application
- Customer logo and message on footer

3.3.5 Self Service BI

Self service BI provide business intelligence to the application.

Self service BI enables analytics through self service provides import of data files that summarized and analyzed in dashboards and reports.

Retailer can Connect the data to other data sourced in Excel or Text files. and also users can bring data in and then visually expose it to the team in the field.

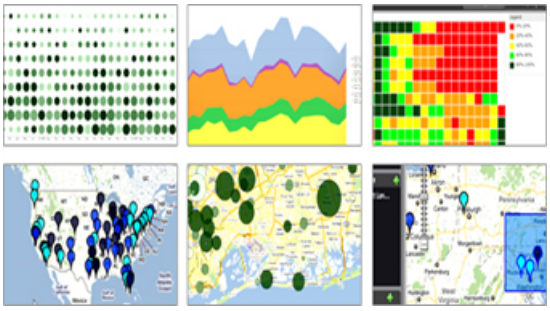

Figure 3.4: Self service BI Feature

3.3.6 Video Integration and Remote Desktop Services

Web Application supports the execution of video links created in the Desktop application. This support includes both Desktop video viewers and URL based video retrieval. Third party remote viewers can now be accessed on a host server as a remote application or on a client machine accessed through a mapped drive.

Figure 3.5: Video Integration

3.4 Architecture

This project follows a layered architecture and the user requests going from layer to layer. The architecture given below depicts the working of the projects in brief and how the data travel from user system to the database and how the reports get generated and communicated back to the desired results.[2]

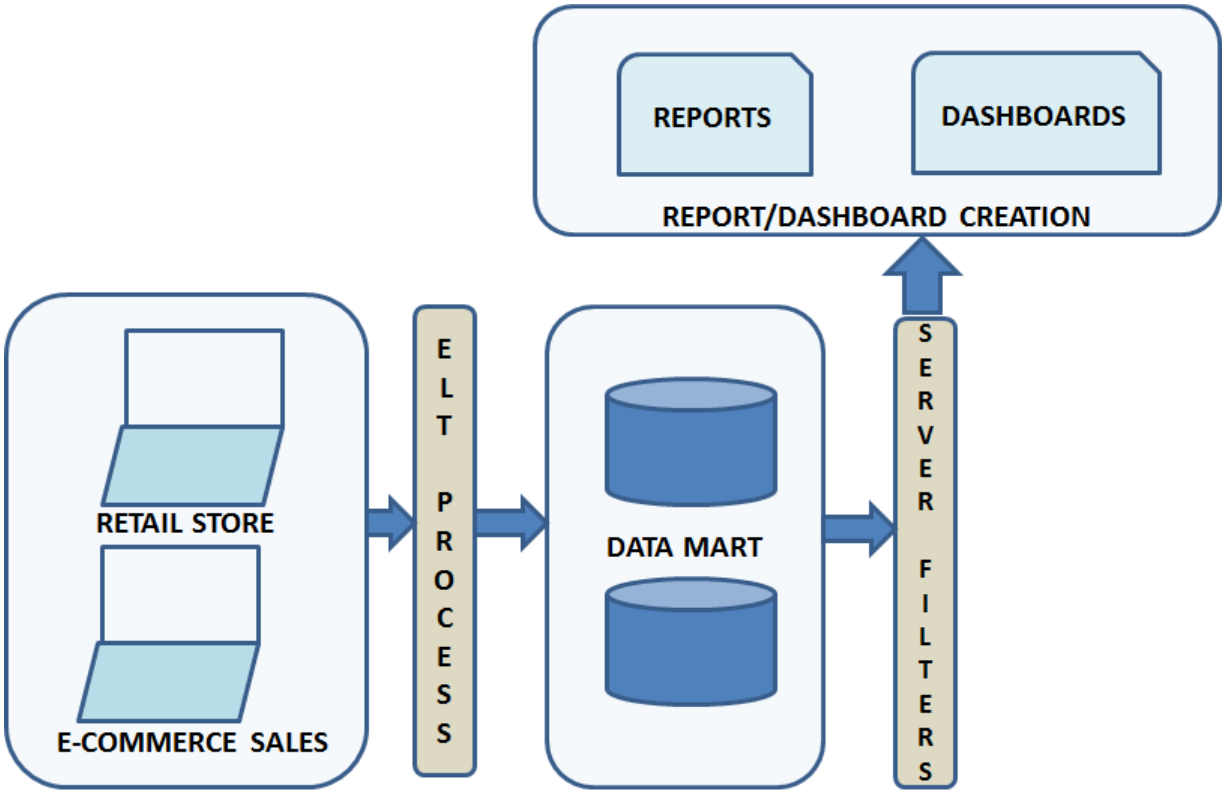

Figure 3.6: Simplified Architecture

Architecture shown in Fig. 3.1 is used in the project and depicts the request and response from the user end. Also, it depicts the layered architecture as discussed above. The following points outline how the application data retrieval is done in the project and shows the work-flow of the application.

- All the customer transaction starts from bottom to up to convert in valuable report.
- On the application architecture, the customer transaction generated at Any Retails Store and E-commerce Inventory Sales.

- All customer related transaction are then transferred to the local database of our system through via path of cloud storage and with the process of ELT
- Transaction then stored into data mart and all related data (data about data) will store into Application Meta Data.
- Then, all the data will transfer to Intelligence Server which leads to take inputs from Data Mart and Application Meta Data.
- Now, all structure getting stored into Schema Object and all values stored into Metrics and then Intelligence Server apply some kind of Filters on data so that valuable information get collected.
- Filtered information then leads to generate reports/dashboards on the basis of given previous data. Dashboard also implemented here as it is nothing, it just collection of various reports.
- Then, generated reports and dashboards are available to display at various devices i.e. Mobile and Web.

Chapter 4

Proposed Solution

4.1 Problem Statement

[6] A retailer need to predict consumer's purchasing pattern to increase the sale. For the analysis, analyst need to gather all transaction data and apply some logic on the data i.e. tedious task. And manually analysis may be erroneous and time consuming.

4.2 Proposed Solution

To reduce time and error in analysis, proposed solution gives easy, graphical and precise prediction of consumer purchasing pattern.

In the proposed solution I am using market basket analysis algorithm to analyze data of retail transactions. After bases of result data analysis, I will apply association mining on data.

4.2.1 Data Mining

Data Mining is technique which processes dataset and identifies particular pattern. Data mining involves methods of intersection of statistics, dataset and machine learning. Data Mining provides business intelligence to the data. Main motive of Data Mining process is to extract information from raw database and create some decision making structure for further use.

Data Mining Process is defined in following stages:[7]

1. Selection - Selecting algorithm according to pattern user want to identify
2. Pre-processing - Data mining is mainly based on datasets. Therefore, before implementing any data mining algorithm we need gather relevant dataset. In all data mining techniques first user need to pre-process dataset.
3. Transformation - In this stage, dataset is transformed in required form of data.
4. Data mining - In this stage, there are 6 common clasess[7]
   (a) Anomaly Detection
   (b) Clustering
   (c) Association Rule Mining
   (d) Classification
   (e) Regression
   (f) Memorization
5. Interpretation/evaluation

4.2.2 Association Rule Mining

Association Rule Mining is a method which is intended to discover visit designs, connections, affiliations, or causal structures from informational collections found in different sorts of databases, for example, social databases, transnational databases, and different types of information vaults[8].

The main applications of association rule mining:

- Basket data analysis - is to analyze the relationship of obtained things in a solitary container or single buy according to the cases given above.
- Cross marketing - is to work with different organizations that supplement your own, not rivals. For instance, vehicle dealerships and producers have cross promoting efforts with oil and gas organizations for evident reasons
- Catalog design - the determination of things in a business' index are frequently intended to supplement each other with the goal that getting one thing will prompt

purchasing of another. So these things are regularly supplements or exceptionally related.

4.2.3 Market Basket Analysis

In retail store, customers purchases different set of products in different time and quantity.One basket says what one customer has purchased at a time.[9]

Market basket Analysis is a technique based on purchasing some set items, you are more likely to purchase another set of items. For example, If you buy butter then there are 95% possibility to buy bread and 70% for JAM and cheese based on store transactions.
Thus, In Market Basket analysis algorithm we gather transaction data to analyze. Rather than focusing to customer, we are more focusing on transaction detail in retail store.
Explanation of used algorithm is as bellow.

1. Gathering Transaction Data from POS - First step is gathering data from Point of sale (POS). POS software gives customer transaction data.

2. Transforming Data - POS gives Data in their format.In this step, We need to transform that data in to JSON or XML format which is acceptable by the system.

3. Cleaning Dataset - In this step, after transformation of data, we need to identify with entity is of our concern.

4. selection of Attributes & measures - In this step, we identify that in the data which attributes are more effective and causes impact on analysis and which are not effecting the pattern. So that we can reduce dimension of analysis and simplify & precise it.

5. Apply Association Mining - we apply this algorithm and Find the pattern.

6. Rule Generation - On the bases of pattern generation, we generate some rule through which we can automate the system. And system has ability to analyze and give output in graphical or textual reports.

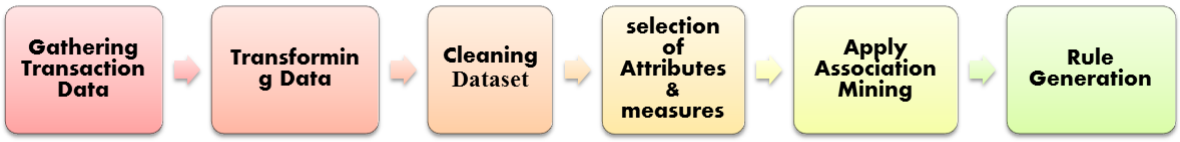

Figure 4.1: Flow For Prediction

Chapter 5

Implementation

5.1 Overview

This chapter consists of how implementation of Market Basket Analysis is done. What is algorithm of Market Basket Analysis. How Dataset is chosen, filtered and clustered based on different parameters. At the based on implementation of algorithm what results are generated. Based on generated result we analyze the results and find the area where we can improve prediction.

5.2 Dataset

For analysis, we have taken online retail transaction dataset in .csv format.
Steps to prepossessing dataset:

1. Gathered unformulated transaction dataset
2. Convert dataset into relational dataset(importing Dataset)
3. Filter data based on departments
4. Clustering dataset based on products
5. Selecting attributes

5.2.1 Gathering Transaction POS dataset

You can collect POS transaction dataset which are formulated by transactions. This raw dataset which is not being processed directly. It looks like bellow figure.

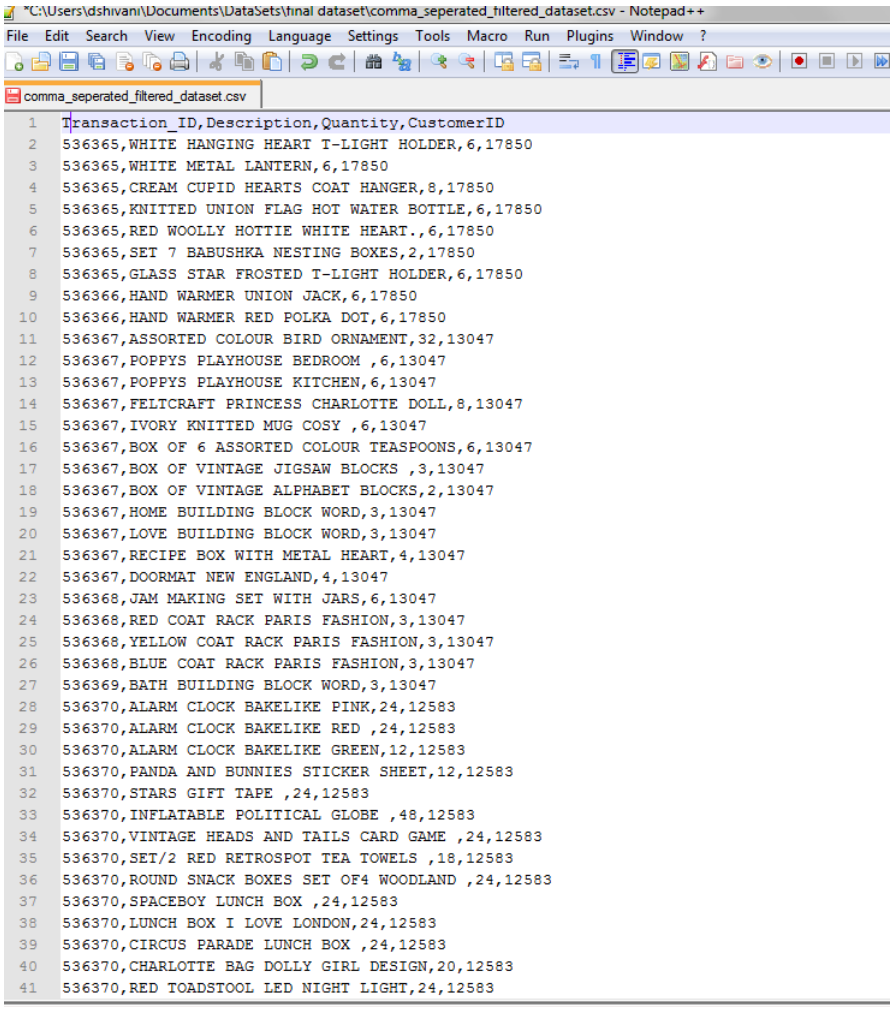

Figure 5.1: Comma separated Raw dataset

5.2.2 Convert dataset into Relational Dataset

It is easy to cluster and manipulated dataset if it is in relational model form.

Here I have used Oracle 11g XE Database for relational data modeling.

- create your instance for database. and Login into the Oracle Database from URL http://127.0.0.1:8080/apex/f?p=4500:1001:637695414893935::NO:::
- Create Relational database tables to filter the attribute and organized them using relation of primary key and foreign key.

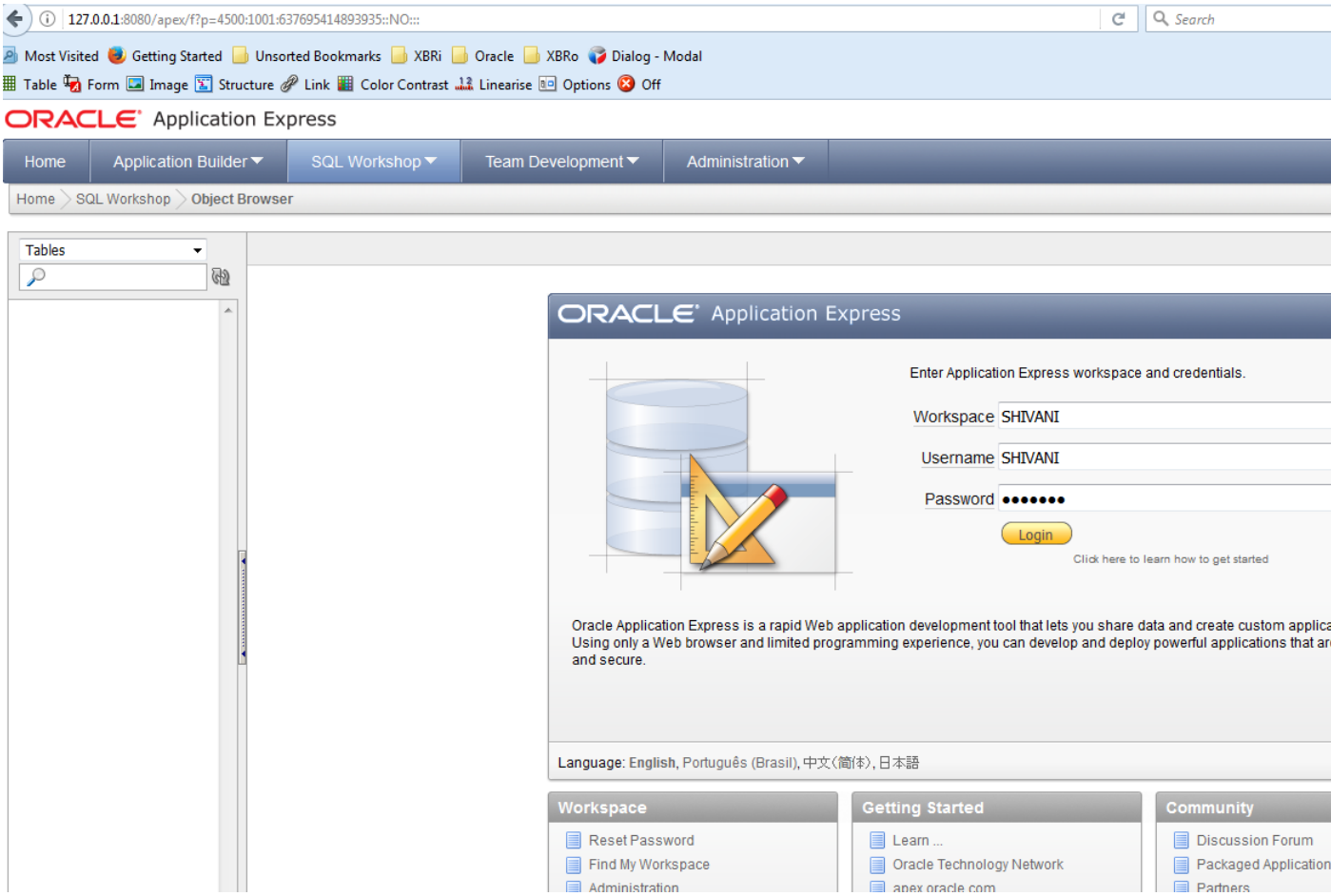

Figure 5.2: Oracle 11g XE Login in own instance

5.3 Algorithm

5.3.1 Terminology

Items are the objects that we are identifying associations between. For an online retailer, each item is a product in the shop. For a publisher, each item might be an article, a blog post, a video etc. A group of items is an item set [7]

$$I = \{i_1, i_2, ..., i_n\} \tag{5.1}$$

Transactions are instances of groups of items co-occuring together. For an online retailer, a transaction is, generally, a, transaction. For a publisher, a transaction might be the group of articles read in a single visit to the website. (It is up to the analyst to define over what period to measure a transaction.) For each transaction, then, we have an item

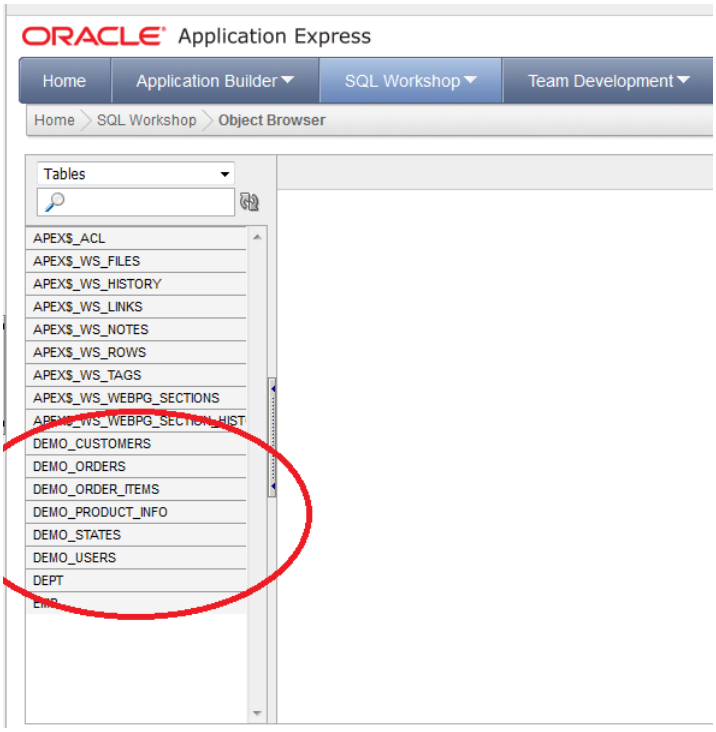

Figure 5.3: Tables for transactions

set.

$$t_n = \{i_i, i_j, ..., i_k\} \tag{5.2}$$

Rules are statements of the form

$$\{i_1, i_2, ...\} \implies \{i_k\} \tag{5.3}$$

i.e. if you have the items in item set (on the left hand side (LHS) of the rule i.e. $\{i_1$, i_2,...$\}$, then it is likely that a visitor will be interested in the item on the right hand side (RHS i.e. $\{i_k\}$. In our example above, our rule would be

$$\{flour, sugar\} \implies \{eggs\} \tag{5.4}$$

The output of a market basket analysis is generally a set of rules, that we can then exploit to make business decisions (related to marketing or product placement, for example).

Column Name	Data Type	Nullable	Default	Primary Key
ORDER_ID	NUMBER	No	-	1
CUSTOMER_ID	NUMBER	No	-	-
ORDER_TOTAL	NUMBER(8,2)	Yes	-	-
ORDER_TIMESTAMP	DATE	Yes	-	-
USER_ID	NUMBER	Yes	-	-
				1 - 5

Figure 5.4: DEMO ORDER

5.3.2 Support

The support of an item or item set is the fraction of transactions in our data set that contain that item or item set. In general, it is nice to identify rules that have a high support, as these will be applicable to a large number of transactions. For super market retailers, this is likely to involve basic products that are popular across an entire user base (e.g. bread, milk). A printer cartridge retailer, for example, may not have products with a high support, because each customer only buys cartridges that are specific to his / her own printer.[8]

$$Support = (Transaction_X \wedge Transaction_Y)/Transaction_{Total} \tag{5.5}$$

5.3.3 Confidence

The confidence of a rule is the likelihood that it is true for a new transaction that contains the items on the LHS of the rule. (I.e. it is the probability that the transaction also contains the item(s) on the RHS.) Formally:[7]

$$Confidence(i_m \implies i_n) = (Transaction_X \wedge Transaction_Y)/Transaction_X \tag{5.6}$$

Column Name	Data Type	Nullable	Default	Primary Key
PRODUCT_ID	NUMBER	No	-	1
PRODUCT_NAME	VARCHAR2(50)	Yes	-	-
PRODUCT_DESCRIPTION	VARCHAR2(2000)	Yes	-	-
CATEGORY	VARCHAR2(30)	Yes	-	-
PRODUCT_AVAIL	VARCHAR2(1)	Yes	-	-
LIST_PRICE	NUMBER(8,2)	Yes	-	-
PRODUCT_IMAGE	BLOB	Yes	-	-
MIMETYPE	VARCHAR2(255)	Yes	-	-
FILENAME	VARCHAR2(400)	Yes	-	-
IMAGE_LAST_UPDATE	DATE	Yes	-	-
				1 - 10

Figure 5.5: DEMO PRODUCT INFO

Column Name	Data Type	Nullable	Default	Primary Key
ORDER_ITEM_ID	NUMBER(3,0)	No	-	1
ORDER_ID	NUMBER	No	-	-
PRODUCT_ID	NUMBER	No	-	-
UNIT_PRICE	NUMBER(8,2)	No	-	-
QUANTITY	NUMBER(8,0)	No	-	-
				1 - 5

Figure 5.6: DEMO ORDER ITEM

5.3.4 Lift

The lift of a rule is the ratio of the support of the items on the LHS of the rule co-occuring with items on the RHS divided by probability that the LHS and RHS co-occur if the two are independent.[8]

$$Lift(i_m \implies i_n) = Transaction_X \vee Transaction_Y / (Transaction_X \wedge Transaction_Y) \tag{5.7}$$

If lift is greater than 1, it suggests that the precense of the items on the LHS has increased the probability that the items on the right hand side will occur on this transaction. If the

lift is below 1, it suggests that the presence of the items on the LHS make the probability that the items on the RHS will be part of the transaction lower. If the lift is 1, it suggests that the presence of items on the LHS and RHS really are independent: knowing that the items on the LHS are present makes no difference to the probability that items will occur on the RHS.

When we perform market basket analysis, then, we are looking for rules with a lift of more than one. Rules with higher confidence are ones where the probability of an item appearing on the RHS is high given the presence of the items on the LHS. It is also preferable (higher value) to action rules that have a high support - as these will be applicable to a larger number of transactions. However, in the case of long-tail retailers, this may not be possible.

5.4 Results

To calculate results, we have taken 50 transactions of clothing department and 10 products of the same department. In market basket analysis, we come to conclusion by finding Lift ratio.

If Lift rate is greater than 0.50 them that two product are mostly bought together. If Lift rate is less than 0.50 they are not co-related. If lift rate is equal to 0.50 then they are related to each other.

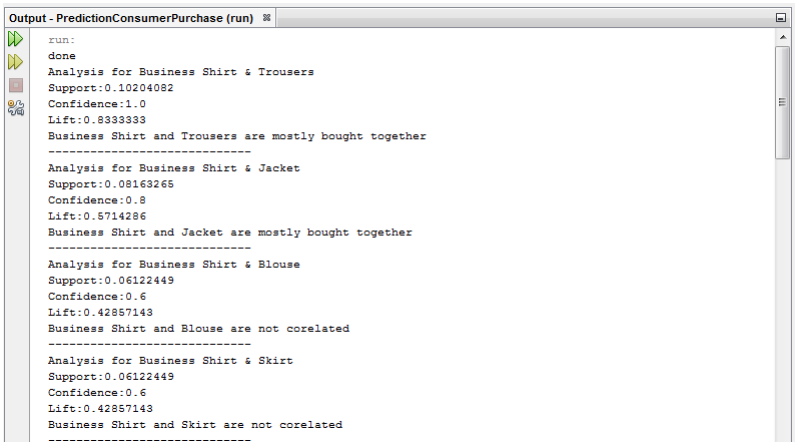

Figure 5.7: Result Screenshot 1

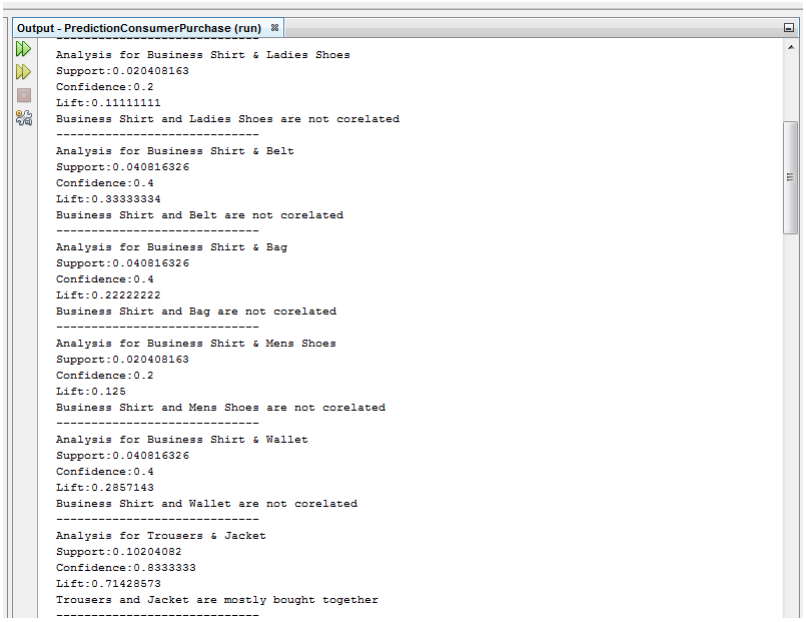

Figure 5.8: Result Screenshot 2

5.5 Result Analysis

For analysis of result, I have taken Business Shirt product for comparison with the other 9 products of clothing department to predict which other products of the same department can be bought together.

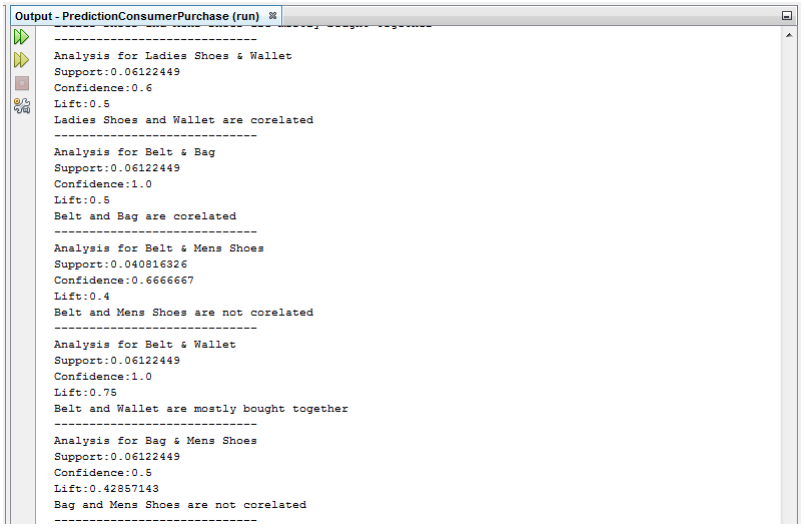

Figure 5.9: Result Screenshot 4

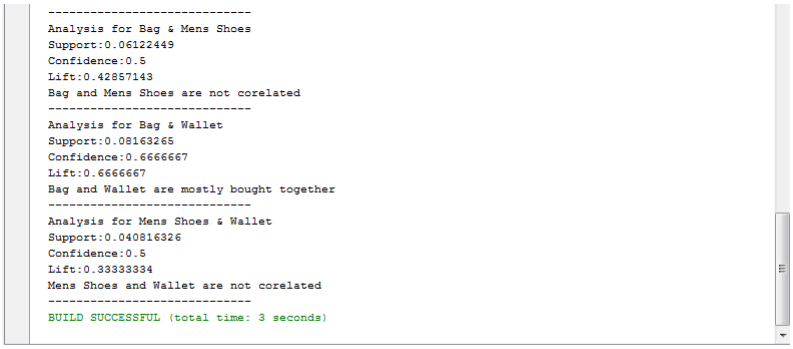

Figure 5.10: Result Screenshot 5

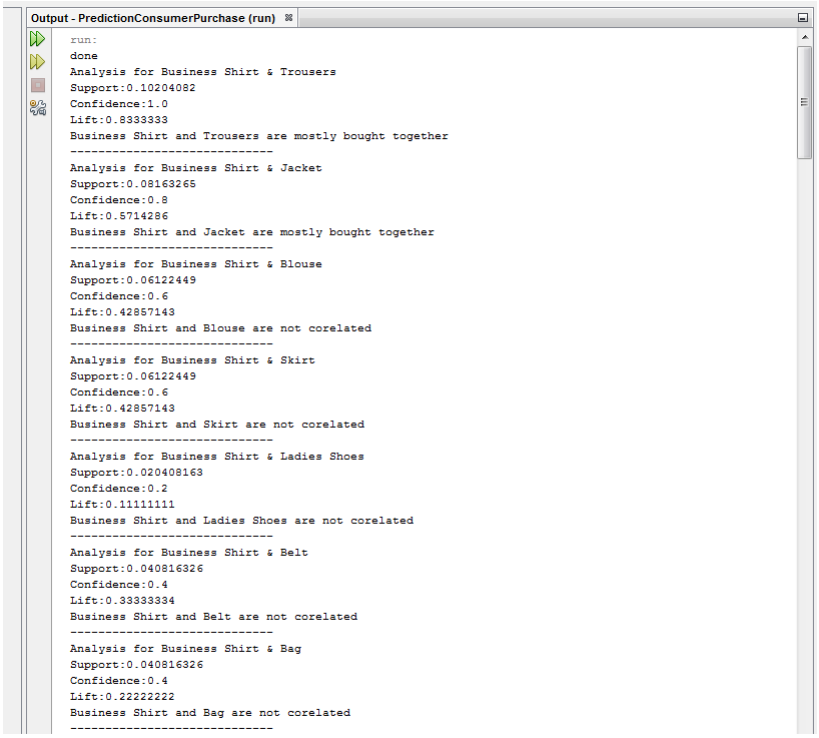

Figure 5.11: Result Screenshot 6

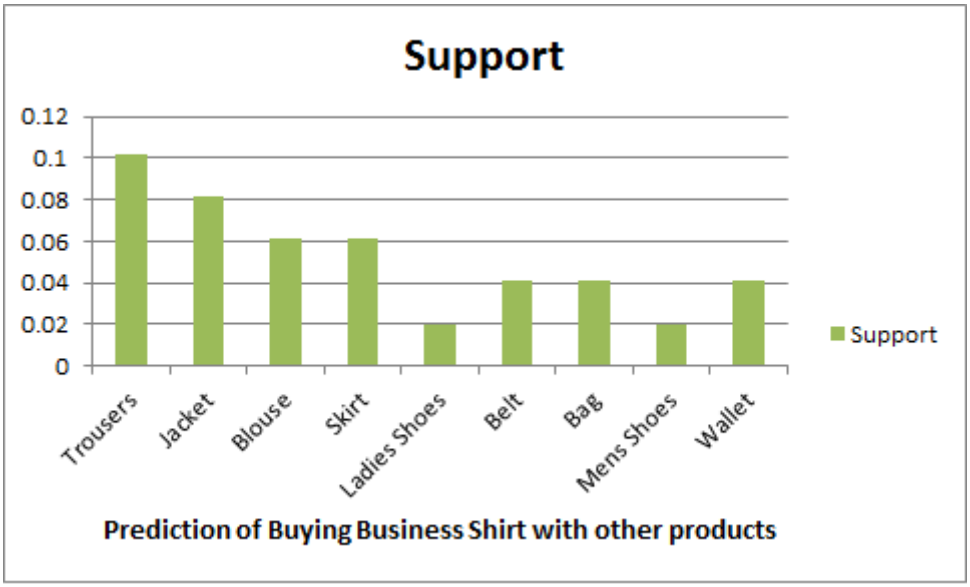

Figure 5.12: Support for Business Shirt $\Longrightarrow$ $Set of Product$

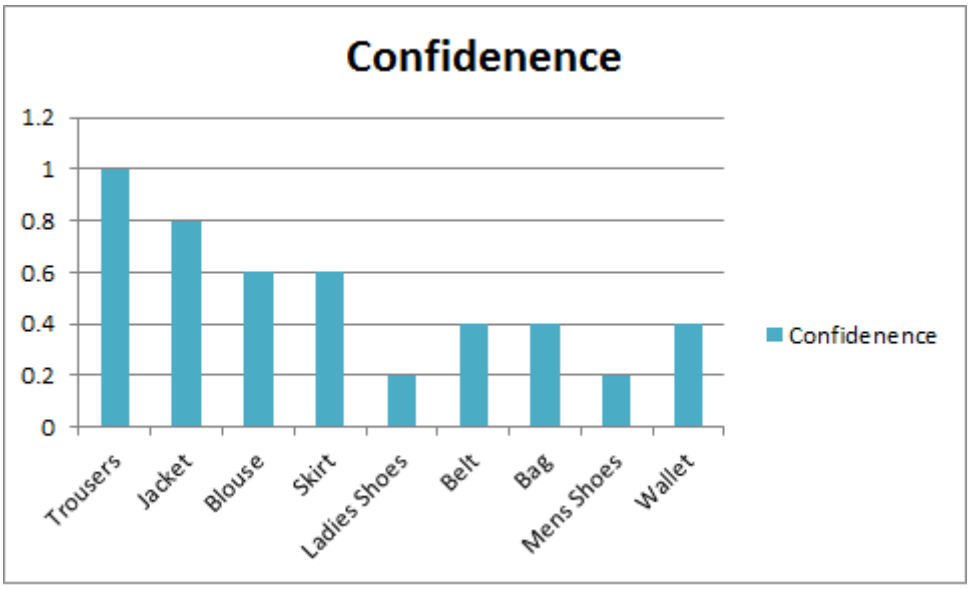

Figure 5.13: Confidence for Business Shirt $\Longrightarrow$ $Setof Product$

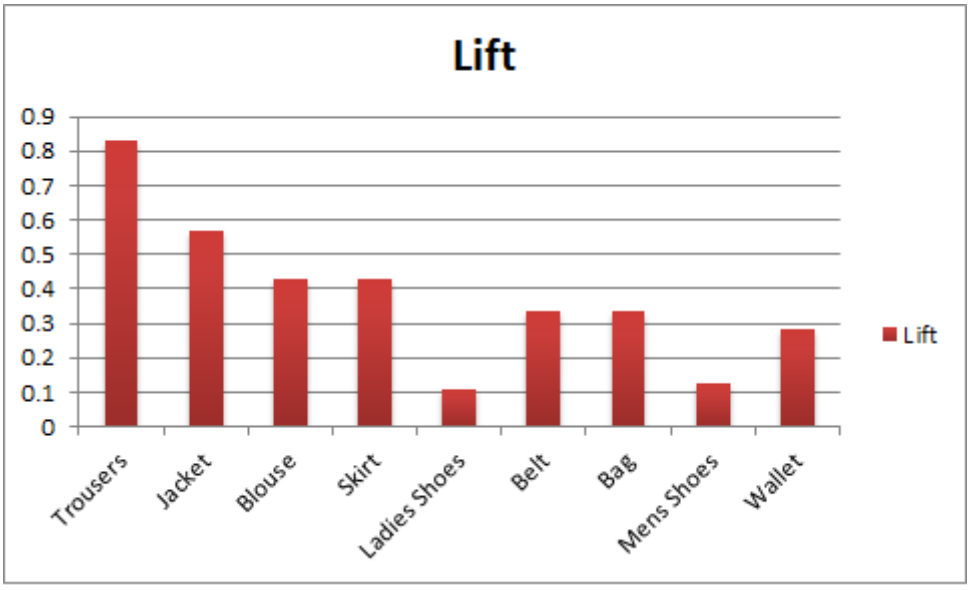

Figure 5.14: Lift for Business Shirt $\Longrightarrow$ $Setof Product$

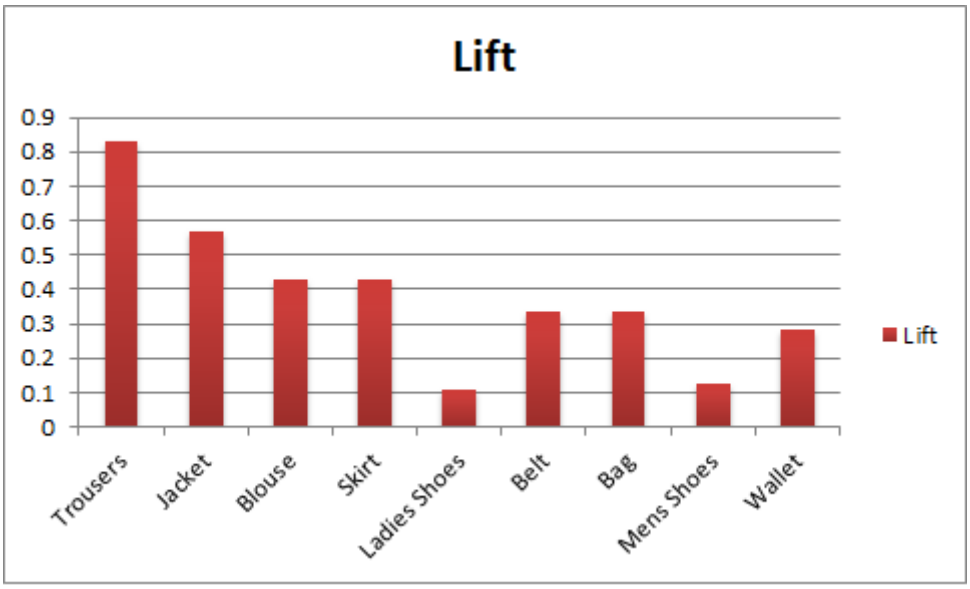

Figure 5.15: Comparison between Support, Confidence, Lift for Business Shirt $\Longrightarrow$ $Set of Product$

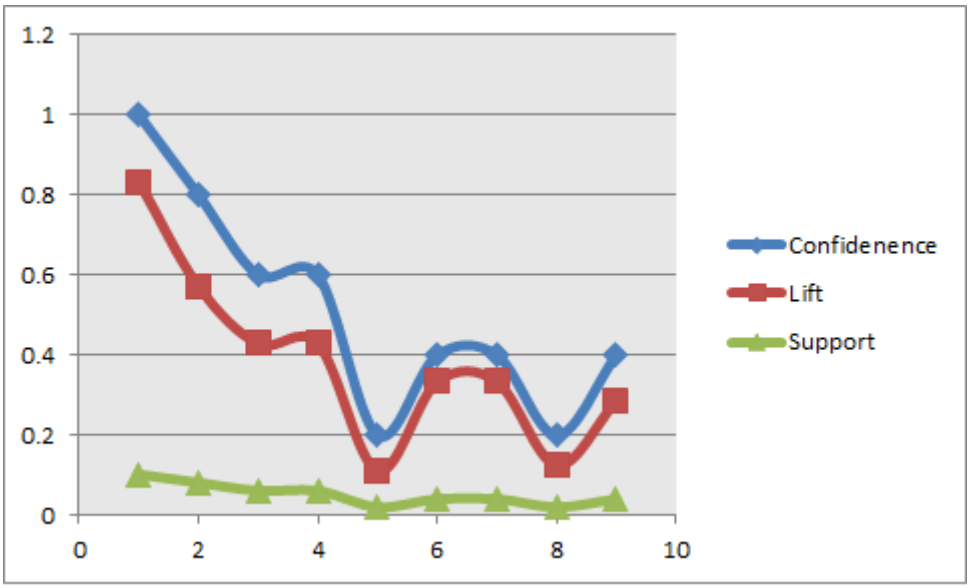

Figure 5.16: Comparison between Support, Confidence, Lift

Chapter 6

Future Work

In future, Based on determined rules in this thesis user can make prediction. To automate this prediction generated rules help. In future, we can do further research by increasing parameters and dimensions of input set. In future, we can enhance implementation by minimizing erroneous resultset.

Bibliography

[1] "Oracle jet cookbook."

[2] B. Goodman, "Loss prevention and sales and productivity cloud services," 2017.

[3] K. Cook, "Method and system for the detection, management and prevention of losses in retail and other environments," Apr. 20 1999. US Patent 5,895,453.

[4] J. Bamfield, "Shrinkage and loss prevention: Evidence from the global retail theft barometer," 2010.

[5] *Oracle Internal Sources*.

[6] K. Y. Y. Zuo and A. B. M. S. Ali, "Prediction of consumer purchasing in a grocery store using machine learning techniques," *Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE), Nadi*, vol. 19, pp. 263–342, 2016.

[7] "Data mining."

[8] "Association rule mining."

[9] "Data mining and its algorithms."