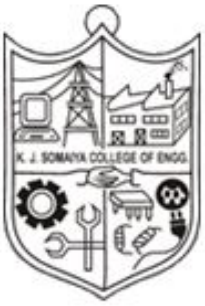
SMARkeT

Submitted in partial fulfilment of requirements
For the degree of

BACHELOR OF ENGINEERING
BY

Chirag Soni-1314055
Shrinath Apake-1314053
Ganesh Sangale-1314040
Pratik Jain-1314038

Under the guidance of
Prof. Manjitsing Valvi



DEPARTMENT OF INORMATION TECHNOLOGY
K.J.Somaiya College of Engineering, Mumbai-400077
(Autonomous College Affiliated to University of Mumbai)

2013-2017

K.J.Somaiya College of Engineering, Mumbai-400077
(Autonomous College Affiliated to University of Mumbai)

Certificate

This is to certify that the Dissertation entitled "SMARkeT" is bona fide record of the dissertation work done by

1. Chirag Soni
2. Shrinath Apake
3. Ganesh Sangale
4. Pratik Jain

in the year 2016-17 under the guidance of Prof. Manjitsing Valvi of Department of Information Technology in partial fulfillment of requirement for the Bachelor Degree in Information Technology of University of Mumbai.

_____-

Guide/Co-Guide

_____

Head of Department

_____

Principal

Date :
Place: Mumbai-77

K.J.Somaiya College of Engineering, Mumbai-400077
(Autonomous College Affiliated to University of Mumbai)

Certificate of Approval of Examiner

We certify that this report entitled "SMARkeT" is
bona fide record of the project work done by

1. Chirag Soni
2. Shrinath Apake
3. Ganesh Sangale
4. Pratik Jain

This is approved for the award of Bachelor Degree in Information Technology of University of Mumbai.

External Examiner

Internal Examiner

Date :
Place: Mumbai-77

Abstract

In the days of the corner market, shopkeepers had no trouble understanding their customers and responding quickly to their needs. The shopkeepers would simply keep track of all of their customers in their heads, and would know what to do when a customer walked into the store. But today's shopkeepers face a much more complex situation. More customers, more products, more competitors, and less time to react, means that understanding your customers is now much harder to do. Along with this the problem of inventory control in a store system involves a variety of activities aimed at a successful management of stocks of goods which are held in the system for future use or sale. This problem arises in many areas of the business including companies in production and commercial branches such as manufacturers, wholesalers and retailers. This system uses advanced data mining techniques for recognizing and tracking patterns within data so as to not only react but also anticipate the customer needs in advance. The System would implement several inventory control algorithms developed on the basis of widespread mathematical models of inventory systems. It also aims to categorize the products according to their demands which will help to identify the products which are more/less in demand. The system will help in making the decision regarding the product to retain/remove from the shells of supermarket store by performing sales forecasting for products.
Keywords: Inventory control,data mining techniques,supermarket,forecasting.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter discusses about the main concept the project is based on.
It identifies what the actual project is meant to accomplish.

1.1 Problem Definition

In the days of the corner market, shopkeepers had no trouble understanding their customers and responding quickly to their needs. The shopkeepers would simply keep track of all of their customers in their heads, and would know what to do when a customer walked into the store.But today's shopkeepers face a much more complex situation. More customers, more products, more competitors, and less time to react, means that understanding your customers is now much harder to do.This system uses advanced data mining techniques for recognizing and tracking patterns within data so as to not only react but also anticipate the customer needs in advance.
The system includes following modules:

1. Clustering of Products
2. Classification of Products
3. Outlier Analysis and Sales Forecasting

1.2 Motivation of the thesis

It has been observed that problem of inventory control in a store system involves a variety of activities aimed at a successful management of stocks of goods which are held in the system for future use or sale.This problem arises in many areas of the business including companies in production and commercial branches such as manufacturers, wholesalers and retailers.The main reason for taking up this project is to help and guide owners of shop market regarding predictives analysis of products in their inventory so as to fulfill customers requirement.

1.3 Scope of the thesis

The project aims to categorize the products in supermarket according to their demands which will help management to identify the products which are more/less in demand using clustering algorithms. The system will also help in making the decision regarding the product to be retain/remove from the shells of supermarket store. The system will also perform the sales forecasting and implement inventory control algorithms to perform inventory management.

1.4 Organisation of the thesis

1. Chapter 1 majorly talks about overview of the project selected and define the Motivation behind the project selected and Scope of the project selected.

2. Chapter 2 provides literature survey done to understand the topic.

3. Chapter 3 provides the Software Project Management Plan (SPMP) for this project which defines project management goals of the project and includes a description of deliverables and deadlines.

4. Chapter 4 provides the Software Requirement Specification which contains the functional and non-functional requirements of the project. This document presents an initial description of the various functionalities and services provided by the software. The document will also serve the basis for acceptance testing by the user.

5. Chapter 5 provides the Software Design Document which includes a comprehensive architectural overview of the system, using a number of different architectural views to depict different aspects of the system. It is intended to capture and convey the significant architectural decisions which have been made on the system.

6. Chapter 6 provides the System Test Document. The goal of this document is to develop a test plan for the SMARkeT i.e A sales and Inventory Management System. This document defines all the procedures and activities required to prepare for testing of the functionalities of the system. The objective of the test plan is to define the activities to perform testing, define the test deliverable documents and to identify the various risks and contingencies involved in testing.

7. Chapter 7 provides the conclusion and scope for the future work of the project.

Chapter 2

Literature Survey

This chapter discusses about the papers referenced in preparation for undertaking this project.
These papers serve as a benchmark to enable this project to be undertaken

2.1 Modified K-Means Clustering Algorithm for Data Mining in Education Domain.

1. Authors:Anurag Bhardwaj,Ashutosh Bhardwaj.

2. Publication:International Journal of Advanced Research in Computer Science and Software Engineering,Volume 3, Issue 11, November 2013.

3. Abstract:In data mining, K-Means clustering is a method of cluster analysis which aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean. In this paper we concentrated on the application of data mining in an engineering education environment. The relationship between students university internal examination results and their success was studied using cluster analysis and modified K-Means algorithm techniques. We proposed an improved K-Means algorithm using noise data filter. The algorithm developed density-based detection methods based on characteristics of noise data where the discovery and processing steps of the noise data are added to the original algorithm.

2.2 A KNN Research Paper Classification Method Based on Shared Nearest Neighbor.

1. Authors:Yun-lei,CaiDuo Ji,Dong-feng Cai.

2. Publication:Proceedings of NTCIR-8 Workshop Meeting, June 1518, 2010, Tokyo.

3. Abstract:The patents cover almost all the latest, the most active innovative technical information in technical fields, therefore patent classification has great application value in the patent research domain. This paper presents a KNN text categorization method based on shared nearest neighbor, effectively combining the BM25 similarity calculation method and the Neighborhood Information of samples. The effectiveness of this method has been fully verified in the NTCIR-8 Patent Classification evaluation.

2.3 Comparisions Between Data Clustering Algorithms.

1. Authors:Osama Abu Abbas.

2. Publication:The International Arab Journal of Information Technology,Vol.5,No. 3,July 2008

3. Abstract:Clustering is a division of data into groups of similar objects. Each group, called a cluster, consists of objects that are similar between themselves and dissimilar compared to objects of other groups. This paper is intended to study and compare different data clustering algorithms. The algorithms under investigation are: k-means algorithm, hierarchical clustering algorithm, self-organizing maps algorithm, and expectation maximization clustering algorithm. All these algorithms are compared according to the following factors: size of dataset, number of clusters, type of dataset and type of software used. Some conclusions that are extracted belong to the performance, quality, and accuracy of the clustering algorithms.

2.4 A Programme Implementation of Several Inventory Control Algorithms.

1. Authors:Vladimir Monov,Tasho Tashev.

2. Publication:Bulgarian Academy of Sciences,Cybernetics and Information Technologies,Volume 11,No 1,Sofia.

3. Abstract: The paper presents several inventory control algorithms developed on the basis of widespread mathematical models of inventory systems. The algorithms have been implemented as software modules in the programme environment of the MATLAB system. Their functionality and basic characteristics are described and illustrated by numerical examples.

Chapter 3

Software Project Management Plan

This document is the fully developed Software Project Management Plan for SMARkeT
This document highlights the deliverables, roles, tasks and probablistic schedule for the team and this project.

3.1 Introduction

When a business carries more inventory than it can sell in the short term, it loses money on warehousing and other costs. When a business carries too little inventory to meet demand, it can lose customers who don't want to wait for the product to come in. The purpose of a sales and inventory system is to make sure the company always has the right amount of inventory.

3.1.1 Project Overview

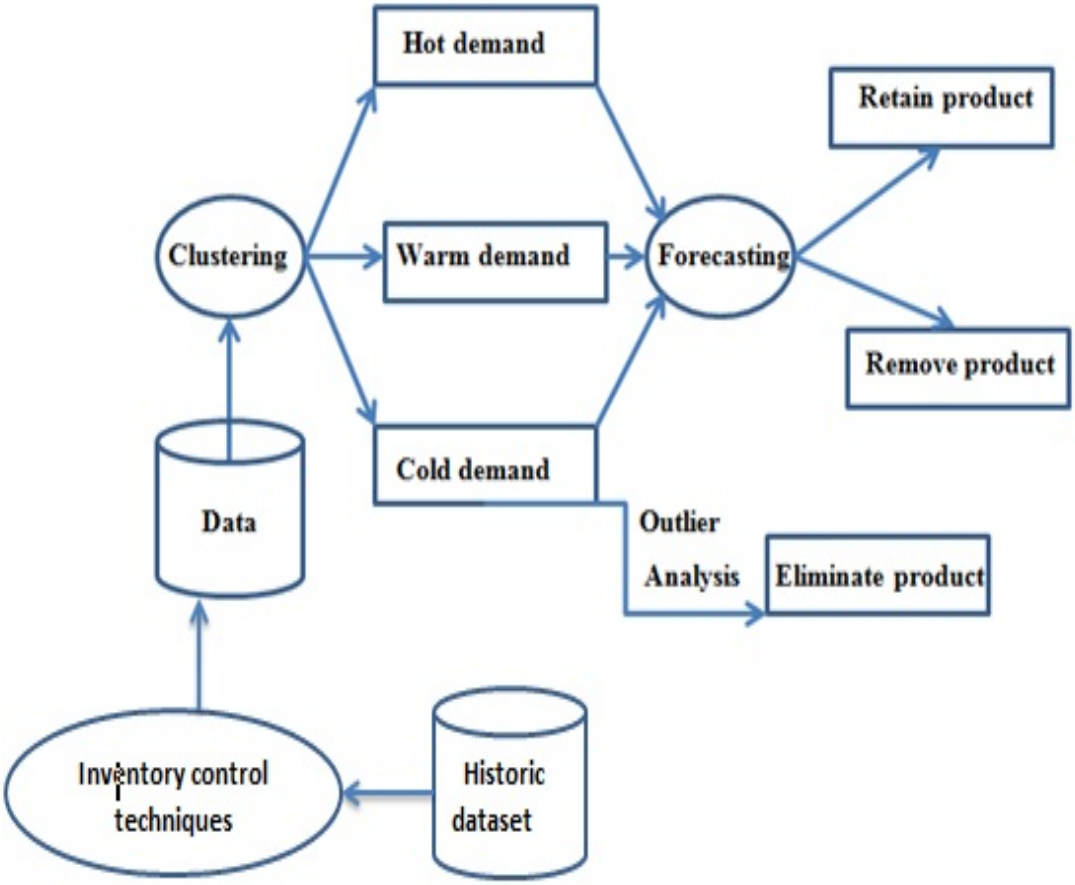


Figure 3.1: System Outline

3.1.2 Project Deliverables

Table 3.1: Project Deliverables

Deliverable	Structure	Standards	Approval Needed By	Resources Required
Software Requirements Specification	Document	As defined in project methodology	Project Guide	Latex
Software Design Document	Document	As defined in project methodology	Project Guide	Latex
System Test Document	Document	As defined in project methodology	Project Guide	Latex
Data Entries	Document	As defined in project methodology	Project Guide	WampServer
Clustering and Classification Modules	Data and Algorithm	As defined in project methodology	Project Guide	Eclipse IDE, WampServer and PHP MyAdmin
Classified Result	System	As defined in project methodology	Project Guide	Eclipse IDE,PhpMyadmin
OutlierAnalysis, Regression and Sales Forecasting Modules	System	As defined in project methodology	Project Guide	Eclipse IDE,PhpMyadmin
User Interface	Prototype	As defined in project methodology	Project Guide	Programmer,Eclipse IDE, BootStrap libraries and Dreamweaver 8

3.2 Project Organization

3.2.1 Software Process Model

The project will follow an incremental and iterative development model for its deliverables. The development will be done in several phases and each phase will represent a complete development cycle, with certain functionality of the system delivered at the end of each phase. The phased approach to delivery provides flexibility in what the team will deliver, gives an opportunity to reassess the effort for each phase and allows the team to change any of the phases content.

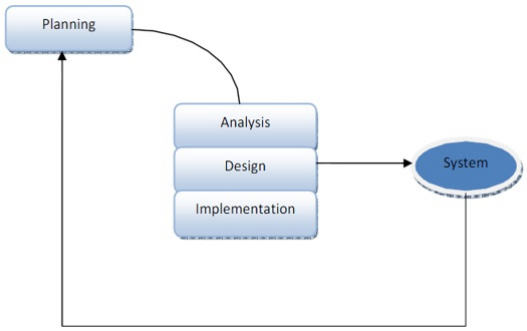


Figure 3.2: System Outline

3.2.2 Advantages Of Iterative Model

1. Some working functionality can be developed quickly and early in the life cycle.
2. Results are obtained early and periodically.
3. Testing and debugging during smaller iteration is easy.
4. Risks are identified and resolved during iteration; and each iteration is an easily managed milestone.
5. During life cycle software is produced early which facilitates customer evaluation and feedback.

3.2.3 Advantages Of Iterative Model

1. Real projects rarely follow the sequential flow and iterations in this model are handled indirectly.
2. It is often difficult to get customer requirements explicitly. Thus specifications cant be frozen.
3. In this model we freeze software and hardware. But as technology changes at a rapid pace, such freezing is not advisable especially in long-term projects.
4. This method is especially bad in case client is not IT-literate as getting specifications from such a person is tough.
5. Even a small change in any previous stage can cause big problem for subsequent phases as all phases are dependent on each-other.
6. Going back a phase or two can be a costly affair.

3.2.4 Roles and responsibilities

Table 3.2: List of Students

Name of Student	Roles	Responsibility
Chirag Soni	Project Team Member	Documentation, Database Repository, GUI and Algorithm Implementation
Shrinath Apake	Project Team Member	Algorithm Implementation, Design and Testing, GUI and Documentation
Ganesh Sangale	Project Team Member	Database Repository, Design and Testing, GUI and Documentation
Pratik Jain	Project Team Member	Algorithm Implementation,Design and Testing and GUI

3.2.5 Tools and techniques

3.2.6 Development Methodology

The project aims execution in an incremental prototyping scheme. The initial phases would consist of requirements gathering, wherein the task of the project team would be to understand the requirements. The successive task will be to then determine the data collection strategy. The system will then be developed in prototypes, each of them accomplishing incremental functionalities in temporal succession.

Hardware Requirements

1. Minimum2 GB RAM required
2. Core2Duo or above.

Software Requirements

1. Eclipse IDE, Wamp Server Software
2. BootStrap Templates,Dreamweaver 8 software and Libraries.

Language to be used

1. Data ETL (Extract,Transform,Load) process: Java,JSP/JSTL,Javascript
2. GUI: HTML 5, CSS 3,JSP/JSTL,Javascript
3. Back-end: PhpMyadmin
4. Web Server: WAMPP,Apache Tomcat

3.3 Project Management Plan

3.3.1 Tasks

Task-1 : SRS(Software Requirement Specification) - T1

Description This task deals with determining the scope to be fixed for the project and also deals with the determination of the functional and non-functional requirements.

Deliverables and Milestones A well defined scope and requirement statement

1. Milestone 1: Identifying Scope
2. Milestone 2: Determining Functional Requirements
3. Milestone 3: Determining Non Functional Requirements

Resources Needed The SRS document will be generated by the team members using Latex based on the Literature Review and research done by the team.

Dependencies and constraints The project must be approved by the assigned Panel

Risks and contingencies The SRS may go under Change Control Processes in case there is a requirement.

Task-2 : Software Design Document - T2

Description Implementation of Inventory Control algorithms for management of stocks of goods in the system. This document will be used to specify the design for implementation of the project in the form of flow charts, class diagrams etc. to unambiguously specify the proposed method of working of application.

Deliverables and Milestones Formulate a mathematical model describing the behaviour of the inventory system.The team members will generate a Software Design Document which will then be verified and will act as the deliverable.

Resources Needed Microsoft Office

Dependencies and constraints Task-1 :Software Requirement Specification Documentation

Risks and contingencies During the progress of the project, it may be required to revise the document.

Task-3 : Developing the clustering and classification algorithm - T3

Description This task deals with the development of the core algorithm for data categorization, wherein the algorithm takes input values from the database.

Deliverables and Milestones Compare and determine best efficient clustering algorithm on the basis of performance.

1. Milestone 1: Identifying input for the algorithm.
2. Milestone 2: Development of the core logic for the algorithm.
3. Milestone 3: Development of the overall algorithm with input output specifications.

Resources Needed Procedural Resources:

1. Meetings with the project guide.
2. Technological Resources
3. IDE like Eclipse.

Dependencies and constraints Task-2 :Software Requirement specification document

Risks and contingencies The initial prototypes of the algorithm shall focus on static development, while the final stages shall focus on developing an algorithm which works in real time.

Task-4 : Implementing OutlierAnalysis,Regression and Sales Forecasting along with Inventory Control Technique for a particularproduct - T4

Description Development of the module implementing outlier analysis and regression techniques to design a decision support system for sales forecasting.Implementation of an Inventory Control Technique for the calculation of Optimum Quantity to be ordered for a Particular Product in the next month.

Deliverables and Milestones A system which provides efficient results pertaining to the retainment and removal of the product.

Resources Needed IDE like Eclipse, WampServer.

Dependencies and constraints Task-2 : Software Design Document and This project depends upon the underlying that the data shall be well categorized and classified.

Risks and contingencies The initial prototypes of the algorithm shall focus on static development, while the final stages shall focus on developing an algorithm which works in real time.

Task-5 : Test Data with algorithms- T5

Description The aim of this module it to display the classified result of the input data of a product.

Deliverables and Milestones Prompt in the UI stating the classified result

Resources Needed IDE like Eclipse, WampServer

Dependencies and constraints Task-3 : Developing the clustering and classification algorithm - T3

Risks and contingencies Improperly implementation of clustering and classification algorithm may result in an incorrect classification of the input data of a product.

Task-6 : User Interface - T6

Description This task aims at maintaining an attractive, easy to use User Interface for the Integrated System

Deliverables and Milestones A fully complete Sales and Inventory System with an attractive User Interface ready to be tested

Resources Needed IDE like Eclispe, Wamp Server, BootStrap Libraries ,Templates and Dreamweaver 8 Software

Risks and contingencies The admin(user) compatibility might vary the design of the complete system.

3.3.2 Timetable

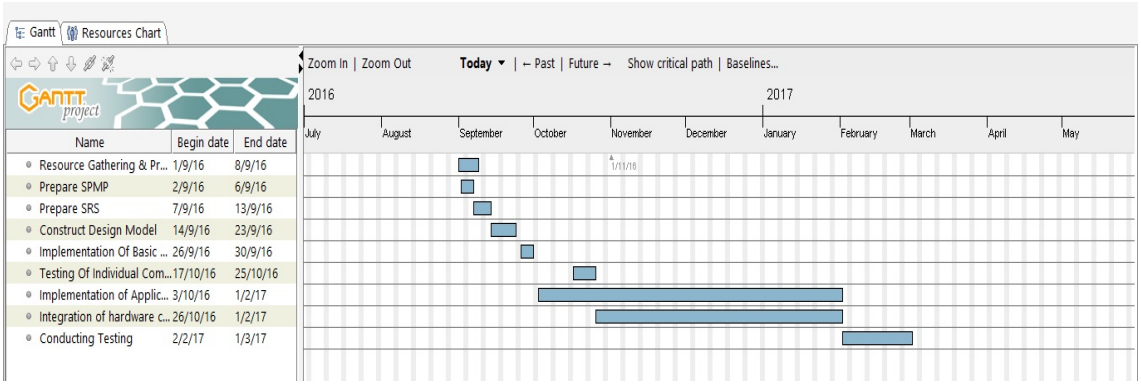


Figure 3.3: Schedule GanttChart

Sr. no	Task	Start Date	End Date
1	Resources Gathering & Procurement	1/9/16	6/9/16
2	Prepare SPMP	2/9/16	6/9/16
3	Prepare SRS	7/9/16	13/9/16
4	Construct design Model	14/9/16	23/9/16
5	Implementation of basic GUI	26/9/16	30/9/16
6	Testing of individual components	17/10/16	25/10/16
7	Implementation of application	3/10/16	1/2/17
8	Conduct Testing	2/2/17	1/3/17

Figure 3.4: Schedule Timetable

Chapter 4

Software Requirement Specification

This SRS attempts to cover in detail the requirement specication of the system. It includes all functional requirements ranging from, and inclusive of the inputs to the system till the output and its analysis.
The non-functional requirements enlist all relevant performance criteria.

4.1 Introduction

4.1.1 Product Overview

The purpose of this system is when a business carries more inventory than it can sell in the short term, it loses money on warehousing and other costs. When a business carries too little inventory to meet demand, it can lose customers who don't want to wait for the product to come in. The purpose of a sales and inventory system is to make sure the company always has the right amount of inventory.

4.1.2 Target Audience

The intended readers of this document are the developers of the site, testers, owners and coordinators. This document is also intended for any individual who is supervising the project. Any suggested changes on the requirements listed on this document should be included in the last version of it so it can be a reference to developing and validating teams.

4.2 Specific Requirements

4.2.1 External Interface Requirements

User Interfaces

1. The admin(user) will utilize the GUI to input required information.

2. The user will be able to click on-screen buttons to maneuver around the program.

3. The user will also be able to click on fields to enter product data into the fields.

4. Overall Statistics of the inventory will be available(displayed) on the dashboard.

Hardware Interfaces

1. Processor Required: Core2Duo or above.

2. RAM Required: 2GB RAM or more.

Software Interfaces

1. Database system
   Name: WampServer(PhpMyadmin)
   This database system will be used to store the data entries of product along with other attributes for faster CRUD (create, read, update, delete) operations.

2. Data Processing System
   Name: Eclipse IDE
   Eclipse IDE will be required to standardize, pre-process and segment the large inventory of product data given as input to use them for clustering,classification and other statistcal prediction analysis.

3. Graphical User Interface
   Name: Dreamweaver 8 softwarre, Bootstrap GUI Libraries and Templates

Performance Requirements

The system should run at such a speed that the admin can move at his/her own pace, without noticing interruption due to processing.

4.2.2 Design Constarints

The program is designed for and will only operate under the Windows OS.

4.2.3 Software Product Features(Functional Requirements)

1. The System would implement several inventory control algorithms developed on the basis of widespread mathematical models of inventory systems.

2. The System aims to categorize the products according to their demands which will help to identify the products which are more/less in demand.

3. The system will help in making the decision regarding the product to retain/remove from the shells of supermarket store.

4. The system will also perform the sales forecasting and enable an admin to have an optimum quanity of a particular product to be ordered using inventory control technique.

4.2.4 Software System Attributes(Non-Functional Requirements)

Reliability

The system should be reliable enough to correctly cluster,classify into clustered groups and predict the sales of the relevant product for the upcoming month using an inventory control technique.

Security

The system should not be susceptible to outside interference and the output should be solely dependent on the input given by the admin user.

Maintainability

The code should be written in a way that it favors implementation of new functions in the future. The development of application is modular so the components of the application shall be reused when adding additional functionality as per the user requirement to the application.

Capacity

The system should able to handle large datasets.

Performance

The client machine should be able to handle minimal computation with respect to the processing requirements with the quick response time.

Chapter 5

Software Design Development

This chapter discusses about the various UML diagramatical representations used to describe the system to be implemented for SMARkeT.

5.1 Introduction

5.1.1 Introduction

The problem of inventory control in a store system involves a variety of activities aimed at a successful management of stocks of goods which are held in the system for future use or sale. This problem arises in many areas of the business including companies in production and commercial branches such as manufacturers, wholesalers and retailers. Inventory control problems also play a central role in the modern supply chain management where the main objective is the overall control of material flows from suppliers of raw materials to final customers. Hence,we propose to design a system that uses advanced data mining techniques for recognizing and tracking patterns within data so as to not only react but also anticipate the customer needs in advance.The various modules to be integrated are as follows,

1. Project Documentation
2. GUI
3. Deployment and testing of the entire system

5.1.2 Requirement Traceability Matrix

A Requirements Traceability Matrix is created by associating requirements with the work products that satisfy them. The configuration management plan is how changes will be tracked and controlled. Traceability is a key part of managing change. Traceability ensures completeness, that all lower level requirements come from higher level requirements, and that all higher level requirements are allocated to lower level ones. Traceability also provides the basis for test planning.

Table 5.1: Requirement Traceability Matrix

	GUI	K-Mean	K-NN	Outlier Analysis and Inventory Control Algo	Regression Technique	Database
Inventory Control Process	x					x
Characterization of Products using Clustering		x				x
Classification of new product	x		x			x
Sale Forecasting						x

5.2 System Architectural Design

System design is the process to define the system architecture, interface, and data for the system in order to satisfy the specified requirements. In this phase, several things are taken into account, which include the interactions between the system and its environment, and the dependencies with other things in order to solve the problem statements. The outcome of this design phase is fully developed.

5.2.1 Chosen System Architecture

The chosen architecture for the following application is a client server model. There application comprises of the user interface, hardware interface and a software interface. The request for a operation and the handling of the request is carried out in the same system.

5.2.2 System Interface Description

The system is confined to single operating system that is Windows.

5.3 Detailed Description of Components

5.3.1 User Interfaces

1. The user interface will be clean and contain simple arrangements which will contain the application window with login screen.
2. Admin(User) will be able to cluster and classify the products with one click in the system along with the other convenient modules to be used.
3. The classification result will be displayed once the user submits the relevant details of the new product to be added in the inventory.

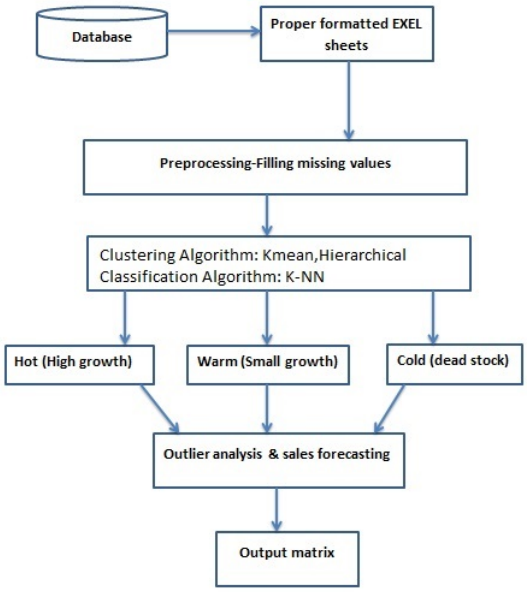


Figure 5.1: Client Server Architecture

4. The user will be allowed to navigate back and forth.

5.3.2 Hardware Interfaces

The application can run on the minimum system configuration that is mentioned.

5.3.3 Software Interfaces

Operating System Interface

This system is compatible to work on windows operating system. So, required Windows OS Version is Windows 7 onwards or higher.

5.3.4 UML Diagrams

Use-Case Diagram

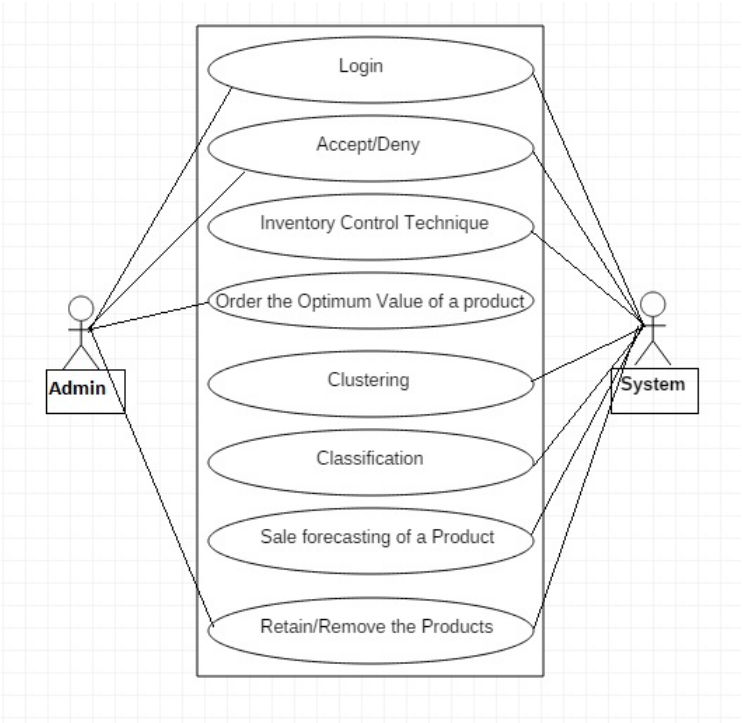


Figure 5.2: Use-Case Diagram

5.3.5 DataFlow Diagrams

Level-0 DFD Level-1 DFD

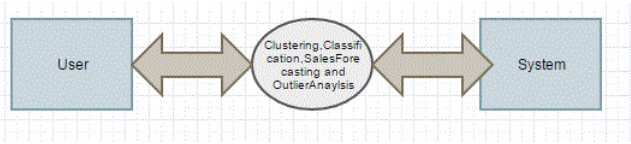


Figure 5.3: DataFlow Diagram Level-0

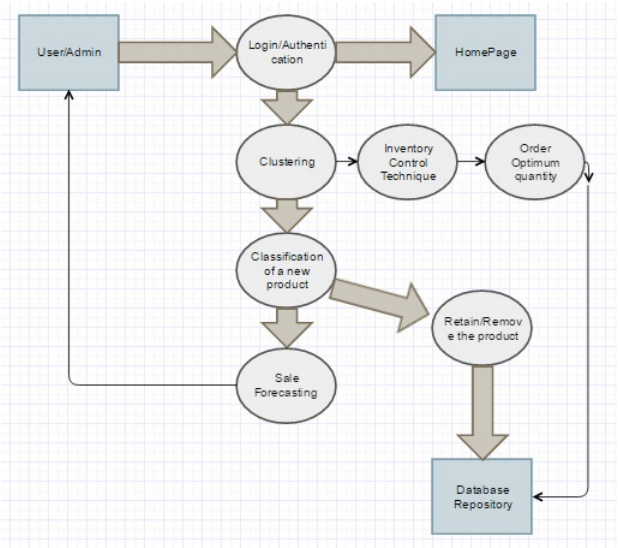


Figure 5.4: DataFlow Diagram Level-1

5.3.6 Sequence Diagram

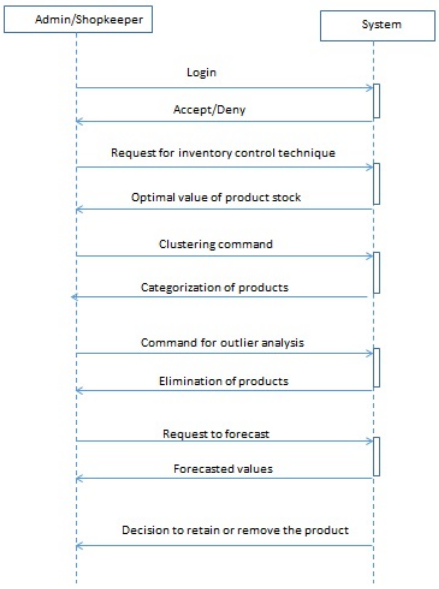


Figure 5.5: Sequence Diagram

Chapter 6

Implementation

6.1 Pre-Processing

The dataset for the products were added and changed to a predefined format as per the requirement of the system.
The project implementation approach was divided into following modules:

1. Admin Validation.
2. Check inventory.
3. Order Stock.
4. Categorization of products.
5. Classification of products.
6. Forecasting.
7. Statistics.

6.2 Tools and Technologies

Software Requirements

1. Eclipse IDE, Wamp Server Software
2. BootStrap Templates,Dreamweaver 8 software and Libraries.

Language to be used

1. Data ETL (Extract,Transform,Load) process: Java,JSP/JSTL,Javascript
2. GUI: HTML 5, CSS 3,JSP/JSTL,Javascript
3. Back-end: PhpMyadmin
4. Web Server: WAMPP,Apache Tomcat

6.3 Algorithms

6.3.1 K-MEANS - Clustering

Inputs:

1. k: the number of clusters
2. D: a data set containing n objects

Output:

1. A set of k clusters

Method:

1. Arbitrarily choose k objects from D as the initial cluster centres.
2. (Re)assign each object to the cluster to which the object is the most similar, based on the mean value of the objects in the cluster.
3. Update the cluster means, that is, calculate the mean value of the objects for each cluster.
4. Repeat until no change.

6.3.2 KNN - Classification

Inputs:

1. Object for which the corresponding class to be found

Output:

1. Class corresponding to the object.

Method:

1. Input: D = (x1, c1), . . . , (xn , cn).
2. x = (x1, . . . , xn) new instance to be classified.
3. For each labelled instance (xi, cn) calculate d(xi, x).
4. Order d (xi, x) in ascending order, (i = 1, . . . , n).
5. Select the K nearest instances to x: DKx.
6. Assign to x the most frequent class in DKx.

6.3.3 Divisive Hierarchical Clustering Algorithm

Steps:

1. Consider one single large cluster with all objects (data points) in the same.
2. Decide initial value of means m1 and m2.
3. Find the Manhattan distance of each object to the means.
4. Based on minimum distance group the objects into clusters which give two clusters.
5. Choose the cluster to split into the sub-clusters:
6. Check the sums of distance between items in a given sub-cluster and choose the one with the largest value.
7. Split the chosen cluster into sub-clusters by repeating steps 2 to 5.
8. Stop after obtaining desired number of clusters. (Our system requires 3 clusters- Best-fit, Good-fit, Low-fit).

Chapter 7

System Test Document

The Software Testing Document (STD) gives an overview of the testing approach for the given project.
It describes the basic tests approach undertaken along with the Test Plan which will include the following: features to be tested, testing tools and environment.
It also includes the test specications.

7.1 Introduction

7.1.1 System Overview

SMARKet is basically a system from administrative point of view. Over here we are checking proper validation of user by taking his username and password and comparing it with the values stored in the database. After successful validation of user, the system shows complete details of various products by classifying it into Hot Demand, Cold Demand and Warm Demand. The products are classified based on their profit percentage which is calculated by using attributes such as stock in the market and the amount sold. Admin will be given authority to classify every new product added to the system.

7.1.2 Test Approach

We will be using a dynamic test approach for this system.We shall be testing the functionality of every module as we encounter new errors.The testing of the system is done through a process of different tests to ensure the correct working of the system software and measure its capabilities and limitations. A brief explanation of the proposed type of tests to be conducted are given below:

Unit Testing

Unit Testing will check each separate unit to ensure it is working as per its requirement. This allows the entire project to be split and tested in replacable modules to enable easy modication and corrective measures to be done.

System Testing

The entire Software will be executed and tested after integrating the different modules to ensure they work in collaboration together correctly.

Performance Testing

Performance test will be conducted to ensure that the softwares performance meets the user expectations and does not exceed the specified performance criteria. During the tests, response time will be measured.

Documentation Test

The user documentation will be tested to ensure its accuracy and to make sure all features are described correctly and in an understandable manner.

7.2 Test Plan Objectives

1. Testing is the process of executing the program the program with the intention of finding an error.
2. A good test is one that has high probability of finding an as yet undiscovered.
3. A successful test is that which uncovers as-yet-undiscovered error.

7.3 Test Plan

7.3.1 Features to be Tested

1. Check whether the credentials of an admin are authenticated correctly in a Login Page or not.
2. Check whether the tables containing product details are saved in and retrieved from a database accurately.
3. Check whether the products of an inventory are clustered into distinct categories viz.Cold(Underfit),Hot(Overfit) and a Warm(Goodfit) category.
4. Check whether a new product is classified correctly into distinct category using existing clusters viz.Cold(Underfit),Hot(Overfit) and a Warm(Goodfit).
5. Check whether the products from an existing inventory are classified correctly into distinct category using existing clusters i.e. Cold(underfit),Hot(overfit) or a Warm(goodfit) category.
6. Check whether the product is removed according to the threshold value of percentage sold entered by the admin.

7. To predict the value of a stock percentage that may be sold out for selected category of a product forthe next month using Regression technique.
8. To calculate an optimum value of the stock for a new product to be ordered for the next month.
9. To calculate an optimum value of the stock for a product in an existing inventory to be re-ordered for the next month.
10. To check whether an interface is user friendly.

7.3.2 Features not to be Tested

1. Whether the data is preprocessed/cleaned.

7.3.3 Testing Tools and Environment

Hardware Configuration

1. Minimum 2GB RAM.
2. Core2Duo or above

Software Configuration

1. Eclipse,WampServer,Tomcat Server

7.4 Test Cases

7.4.1 TC-1

Purpose

Check whether the credentials of an admin are authenticated correctly in a Login Page.

Input

Feed the username and password.

Expected Outputs and Pass/Fail Criteria

Error prompt if the admin is not authenticated and invalid, hence the test case with required credentials is passed. If the user is valid, no prompt appears.

Test Procedure

The procedure for the following test involves only feeding the user credentials to the system with the help of the interface.

7.4.2 TC-2

Purpose

Check whether the tables containing product details are saved in and retrieved from a database accurately.

Input

All the appropriate fields related to product entered properly in the standardized format.

Expected Outputs and Pass/Fail Criteria

Entries should be saved in the database successfully.If the format of any data entry mismatches the predefined datatype then,query is fired prompting an error in saving the data.

Test Procedure

Different product details like category,subcategory,brand,cost price per unit,storage price,quantity to be ordered(stockavailable) are manually entered in a text file which is the latest orders table.Furthermore,the data entries are converted into a single SQL query file to be fired in the PhpMyadmin.

7.4.3 TC-3

Purpose

Check whether the products of an inventory are clustered into distinct categories viz: Hot(Bestfit),Warm(Goodfit),Cold(Lowfit).

Input

Current months product table containing various categories of products and their details.

Expected Outputs and Pass/Fail Criteria

Categorisation of products in Hot(Bestfit),Warm(Goodfit) and Cold(Lowfit) categories.If the products with the similar details(like sold percentage) after a monthly cycle are not categorized in the same class then it fails the test case.

Test Procedure

Clicking on Categorize command will form the three different clusters which would be reflected in the tables present in the database as well to an admin upon requesting the details.

7.4.4 TC-4

Purpose

Check whether a new product is classified correctly into a distinct category using existing clusters viz.Cold(Underfit),Hot(Overfit) and Warm(Goodfit).

Input

New Products details like stock available and stock which may be sold out in the next month,if ordered.

Expected Outputs and Pass/Fail Criteria

Following test case is passed if an appropriate category of the product is displayed.

Test Procedure

For conducting the following test, a new product is selected to be orderd and and its details are fed into the system(Details relevant to K-NN algorithm). Following data of the product is already cleaned and undergoes classification algorithm in the system. After processing, the classified result is displayed.The admin is then given the option to order that product with his/her desired quantity.

7.4.5 TC-5

Purpose

Check whether the products from an existing inventory are classified correctly into distinct category using existing clusters i.e. Cold(underfit),Hot(overfit) or a Warm(goodfit) category.

Input

Details of the product table of the current month from the Data repository are fed as an input to the K-NN algorithm.

Expected Outputs and Pass/Fail Criteria

Following test case is passed if an appropriate category of the product is displayed.The displayed category can be compared with the already existing category(By looking into Hot,Cold,and Warm clusters) of the selected product.

Test Procedure

For conducting the following test, an existing product is selected to be re-orderd and its details are fed into the system(Details relevant to K-NN algorithm). Following data of the product is already cleaned and undergoes classification algorithm in the system. After processing, the classified result is displayed.This approach is undertaken to determine the class of each product individually as desired.The product can then be re-ordered for the next month with the desired quantity.

7.4.6 TC-6

Purpose

Check whether the product is removed according to the threshold value of percentage sold entered by the admin.

Input

Dataset of products categorized as Lowfit(Cold) category.

Expected Outputs and Pass/Fail Criteria

The test case is passed if the percentage of stock sold of a selected product lies below the threshold entered by an admin and it is removed from the next months Orders List successfully.

Test Procedure

The admin is given the option to select any category of a product and its threshold value below which all the products of that category should be displayed.The desired product is then removed as per the admin's choice.Outlier analysis is implemented on the Lowfit products table since,it is the table with the least percentage of stock sold.

7.4.7 TC-7

To predict the value of a stock percentage that may be sold out for a selected category of a product for the next month using Regression technique.

Input

Data such as stock availability and stock sold of a selected category is retrieved and passed on to Regression algorithm.

Expected Outputs and Pass/Fail Criteria

The test case is passed if the the accurate percentage value for the next months sale for a selected category should be displayed.

Test Procedure

Data such as stock availability and stock sold of a selected category is retrieved and passed on to Linear Regression algorithm.Further,the predicted value of the selected category is displayed.

7.4.8 TC-8

To calculate an optimum value of the stock for a new product to be ordered for the next month.

Input

Product details like Cost Price per unit, Storage cost and the assumed value of the stock which may be sold out are manually entered by the admin.

Expected Outputs and Pass/Fail Criteria

The test case is passed if the the system returns the optimum value of the product to be ordered successfully.

Test Procedure

Feeding all product details after accepting from a form to the algorithm and then the optimum value is displayed.The admin can then order that product with the optimum quanitity or the desired quantity,thereby replacing the details of that product in the orders table for next month.

7.4.9 TC-9

To calculate an optimum value of the stock for a product in an existing inventory to be re-ordered for the next month.

Input

Data such as Cost Price per unit, Storage cost and stock sold of a selected product is retrieved and passed on to the algorithm.Apart from this,admin is given an option to set the desired buffer value which acts as a percentage of stcok left.

Expected Outputs and Pass/Fail Criteria

The test case is passed if the Optimum quantity of the selected product is displayed.

7.4.10 Test Procedure

Feeding all product details to the algorithm and then the optimum value is displayed for the desired product.The admin can then re-order that product with the optimum quanitity or the desired quantity,thereby replacing the details of that product in the orders table for next month.

7.4.11 TC-10

Purpose

To check whether an interface is user friendly.

Input

Datasets from an inventory along with the data that an admin enters using forms

Expected Outputs and Pass/Fail Criteria

The test case is passed if the desired results are displayed with proper navigation among the modules while the system is in use.

Test Procedure

Feeding large amount of data to the system and reflecting the desired output tables in the User Interface.

7.5 Test Cases

Test Case No	Test Case	Description	Input	Expected Output	Actual Output	Result
TC-1	Login Page	This test case is used for Admin authentication	Username and Password	Admin is identified	Admin is successfully authenticated	Pass
TC-2	Adding/Retrieving product details to/from Database	Adding the values in the tables	Fields related to the products	Entries should be saved in and retrieved from the database	Product entries are successfully added with Dynamic update	Pass
TC-3	Clustering	This test case is used for clustering products into Bestfit, Goodfit and Lowfit categories	Current month's product table containing various categories of products	Categorization of products into Bestfit, Goodfit and Lowfit categories	Products are distinctly categorized	Pass
TC-4	Predictive classification of a new product on the basis of manual input given by the user	Check whether the new product is classified correctly into distinct category using existing clusters	New Product's details like stock available and stock sold. Etc.	Appropriate category of the product should be displayed	An appropriate category of the product is determined and displayed successfully	Pass
TC-5	Classification of the existing products into appropriate categories	Check whether the product from an existing inventory are classified correctly into distinct category using existing clusters	Details of the product table of the current month from the Data repository	Selected product should be classified appropriately	An appropriate class of the selected products are displayed successfully.	Pass
TC-6	Outlier Analysis of products categorized as Lowfit(Cold) category	Check whether the product is removed according to the threshold value of percentage sold entered by the admin	Dataset of products categorized as Lowfit(Cold) category	Selected product should be removed from the next month's Orders List.	Selected product is removed from the next month's Order List which is reflected from the Orders Table	Pass

Figure 7.1: Test Cases

Test Case No	Test Case	Description	Input	Expected Output	Actual Output	Result
TC-7	Sale Forecasting of a selected category of a product for the next month	To predict the value of a stock percentage that may be sold out for selected category of a product for the next month using Regression technique	Data such as stock availability and stock sold of a selected category is retrieved and passed on to Regression algorithm	Accurate percentage value for the next month's sale for a selected category should be displayed	Accurate percentage value for the next month's sale for a selected category is determined and displayed successfully	Pass
TC-8	Inventory Control Technique to order Optimum Quantity for a new product	To calculate an optimum value of the stock for a new product to be ordered for the next month	Product details like Cost Price per unit, Storage cost and the assumed value of the stock which may be sold out are manually entered by the admin	Optimum Quantity of the new product should be displayed.	Optimum Quantity of the new product is displayed successfully and an option to place an order is made available.	Pass
TC-9	Inventory Control Technique to re-order Optimum quantity for the existing products	To calculate an optimum value of the stock for a product in an existing inventory to be re-ordered for the next month	Data such as Cost Price per unit, Storage cost and stock sold of a selected product is retrieved and passed on to the algorithm	Optimum quantity of the selected product should be displayed.	Optimum quantity of the existing product is displayed successfully and an option to re-order is made available.	Pass
TC-10	User Interface	To check whether an interface is user friendly or not	Datasets from an inventory along with the data that an admin enters using forms	Desired results should be displayed with proper navigation among the modules while the system is in use	Desired results are displayed with proper navigation among the modules while the system is in use.	Pass

Figure 7.2: Test Cases

Chapter 8

Results and Discussion

This chapter discusses about the results of the project.

8.1 Results and Discussion

Our system provide the authority to categorize the product table into three classes (Best-fit, Good-fit, Low-fit) depending on the sales done in current month.It makes use of K-Means algorithms to perform categorization of the products present in the training dataset. Best-fit table shows the list of product with percentage sold in the range of 72-100.Good-fit shows the list of product with percentage sold in the range of 39-71.Low-fit shows the list of product with percentage sold in the range of 0-38.

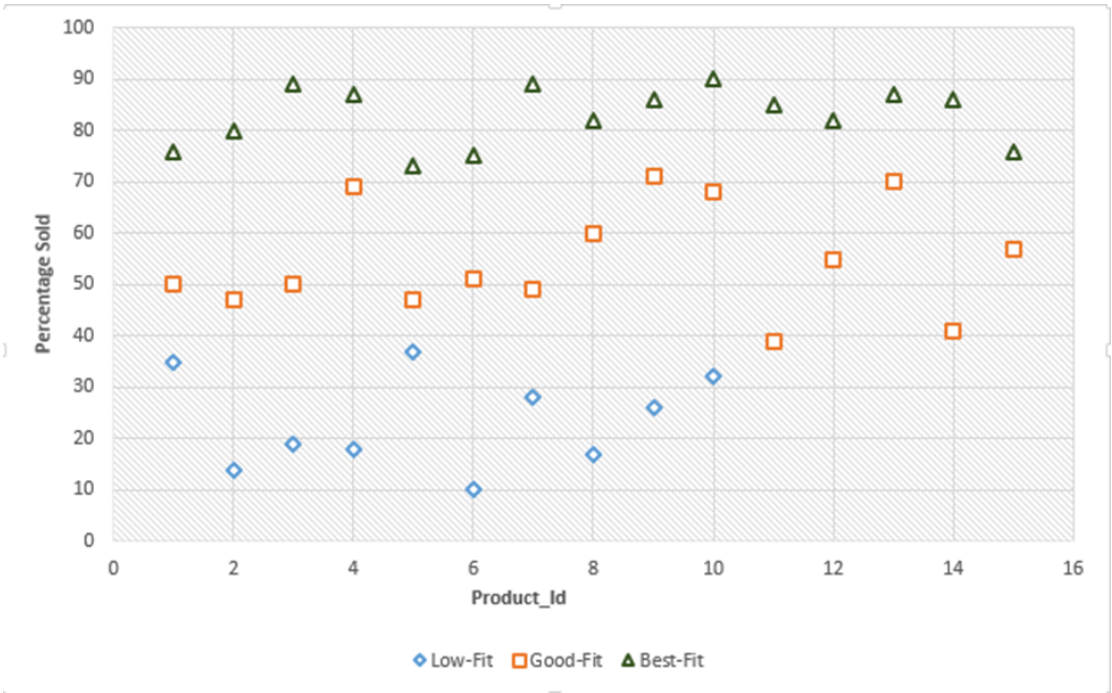


Figure 8.1: Categorization of products by K-Means

In this section, we compared the two clustering algorithms i.e. K- Means and Divisive hierarchical Clustering, which were used in our system to categorize products in the inventory into categories such as Best-fit, Good-fit and Low-fit.K-Means has a time complexity of O (i*k*n).Where, i is number of iterations, k is number of clusters, n is number of objects.Divisive Hierarchical Clustering has a time complexity of O (m*n2).Where, m is number of clusters, n is number of objects. When K- Means was used for categorizing the products in the inventory, we obtained clusters with a range of percentage sold as shown in Figure 8.2. When Divisive hierarchical clustering was used to categorize products in the inventory, we obtained clusters with a range of percentage sold as shown in Figure 8.4. On comparing two algorithms, we can say that K-Means algorithm distributes the percentage sold range more evenly into Best-fit, Good-fit and Low-fit category each getting about 30 percent share of range as compared to Divisive hierarchical Clustering where Low-fit Category was dominant with 50 percent share of range. So from these observations, we can say that K-Means performs better in evenly distributing the products into Best-fit, Good-fit and Low-fit category based on percentage sold, then divisive hierarchical clustering algorithm.

Category	No of Products	Percentage Sold Range
Low-fit	877	0-38
Good-fit	1701	39-71
Best-fit	1588	72-100

Figure 8.2: Categorization of products using K-Means clustering

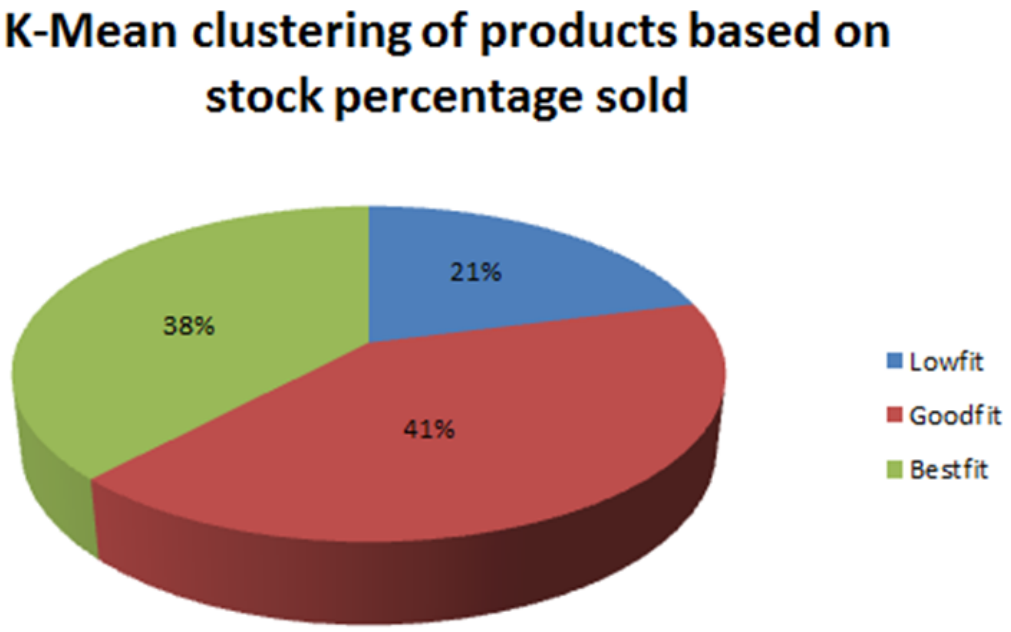


Figure 8.3: K-Means clustering of products based on stock percentage sold

Category	No of Products	Percentage Sold Range
Low-fit	1590	0-50
Good-fit	1200	51-75
Best-fit	1376	76-100

Figure 8.4: Categorization of products using Divisive hierarchical clustering

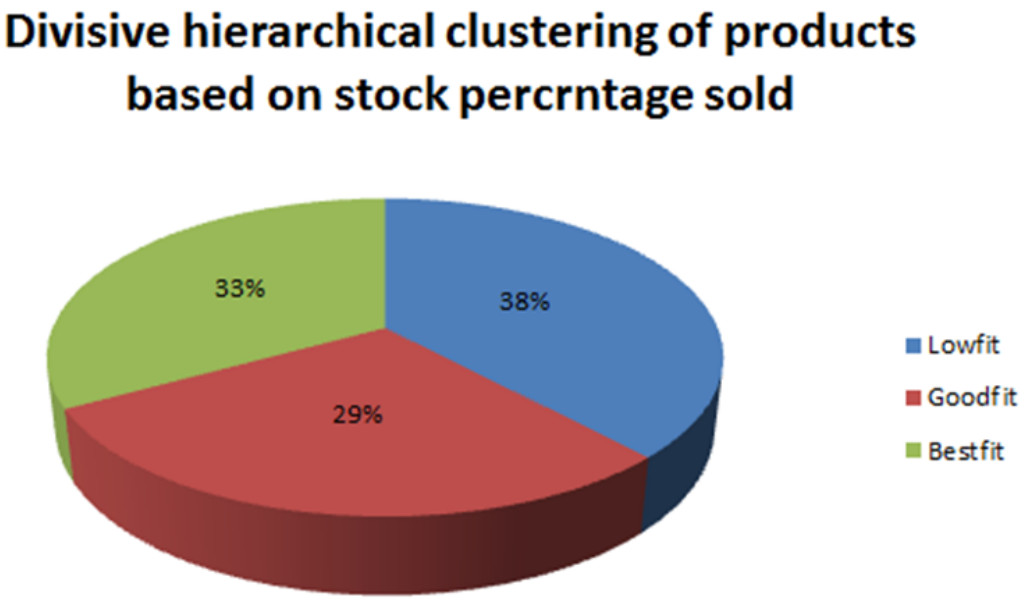


Figure 8.5: Divisive Hierarchical Clustering of products based on stock percentage sold

Chapter 9

Conclusion and Future Enhancements

This chapter discusses about lessons learned and knowledge gained after the completion of every module of our project and the possible future scope and expansion of our project.

9.1 Conclusions

In the first phase of our project we have acquired all the needed product data for the dataset. After acquiring it,the products were categorized successfully using K-mean algorithm.Any new entry in the product inventory is always classified using K-NN algorithm on the basis of clusters formed before.The second phase of our project inludes the implementation of the modules where the admin/user is been given the feature in a system that uses the algorithm of Inventory Control Technique to order an optimum quanitity of the stock for an existing or a new product for the next month.The user can have the predictive analysis of any product he/she wants regarding the sale forecasting for the next month.Furthermore,if any product which is not yielding the desired profit for a while or is not matching the desired sale percentage value(i.e Cold Category Product) set by user then it can be removed from the next month's order table or retained according to the admin's choice.Overall,the system is mainly designed to reduce the manual work of updating,tracking and also make it easier for the admin/user to handle the stock details of every product in an inventory along with providing the features that maximises the sales and minimises the chances for loss in the future.

9.2 Future Enhancements

There is always a room for improvement in any system so, however good and efficient it may be.Considering this important factor, the system is designed in such a way that provisions can be given for further enhancement without affecting the system presently developed.The system, SMARkeT is widely open for the future improvements which can include forecasting of particular products rather

than forecasting of complete category of the products. This forecasting can be done with polynomial regression method to improve the accuracy of predicted forecasted value.Further, this system can be improved by adding flexibility to offer various discounts on the products to increase business profit. These discounts can be given on various products to clear the long time retained stock. Also, in different seasons these discounts can be given on the basis of the variability of rate of demands of the products (say, during summer season products like cold drink, ice-creams will have high demand rate).The system, SMARkeT can have special module of CRM (Customer Relationship Management) to receive feedbacks, suggestions from customers and subsequently act on it.In addition to all this features, AI component can also be integrated to automate decision making process and to place optimum orders for the products without less contribution of the owner of the system.

Chapter 10

References

10.1 References:

1. Data Mining: Concepts and Techniques (The Morgan Kaufmann Series in Data Management Systems), Jiawei Han (Author), Jian Pei (Author), Micheline Kamber (Author).

2. Anurag Bhardwaj, Ashutosh Bhardwaj, Modified K-Means Clustering Algorithm for Data Mining in Education Domain, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 11, November 2013.

3. Yun-lei, CaiDuo Ji, Dong-feng Cai, A KNN Research Paper Classification Method Based on Shared Nearest Neighbor, Proceedings of NTCIR-8 Workshop Meeting, June 1518, 2010, Tokyo.

4. Vladimir Monov, Tasho Tashev, A Programme Implementation of Several Inventory Control Algorithms, Bulgarian Academy of Sciences, Cybernetics and Information Technologies, Volume 11, No 1, Sofia.

5. https://grofers.com/

6. https://www.bigbasket.com/choose-city/?next=/cl/grocery-staples/

7. http://www.baniyasupermarket.com/

8. http://www.bigbazaar.com/

9. http://www.dmartindia.com/

10. http://www.indiastudychannel.com/projects/5169-K-Means-Clustering-Algorithm-in-Java.aspx

11. http://www.softwareadvice.com/inventory-management/spotlight/user-interfaces-2014/

12. https://www.youtube.com/watch?v=Fm01pqWLqzU

13. Cecil Bozarth and Robert Handfield, Introduction to Operations and Supply Chain Management,Prentice-Hall, 2006. ISBN 0139446206.

Chapter 11

Acknowledgement

We have taken efforts in this project. However, it would not have been possible without the kind support and help of many individuals. We would like to extend our sincere thanks to all of them.We are highly indebted to Mr.Manjitsing Valvi for his guidance and constant supervision as well as for providing necessary information regarding the project and also for his support in completing the project.We would like to express our special gratitude and thanks to our project panel member Prof. Nilkamal More.Her thoughts and opinions have contributed tremendously in making this project an immaculate contrast to what we had imagined.Furthermore, we would like to express our gratitude towards our parents for their kind co-operation and encouragement which help us in completion of this project.Our thanks and appreciations also go to all those people who have willingly helped us out with their abilities.