ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

SMARkeT: Sales and Inventory Management System Incorporating Data Mining Techniques

Ganesh Sangale[1], Pratik Jain[2], Shrinath Apake[3], Chirag Soni[4], Prof. Manjitsing Valvi[5]

B. E. Student , Dept. of I.T., K. J. Somaiya College of Engineering, Mumbai, India[1,2,3,4]

Assistant Professor, Dept. of I.T., K. J. Somaiya College of Engineering, Mumbai, India[5]

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 algorithms,data mining techniques,sales forecasting, Inventory.

I. 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. The system demands historic dataset consisting of detail information in order to perform data mining techniques. The system uses Inventory control technique[4] to predict the optimum value of order to be placed for the next month. Our system basically makes use of clustering algorithm to categorize the products into best-fit, good-fit and low-fit category depending on their demands. The system also provides flexibility to add and determine category of new product as per the owner's requirement. It predicts the appropriate value of percentage of stock sold for any particular category by aggregation of all the products of that category with the help of linear regression method. The owner of the system can also eliminate the products from the inventory by setting a threshold value of stock percentage sold of any category of the products. Finally, data of sales of the products can be visualized easily with the help of various graphs generated.

II. LITERATURE SURVEY

A. *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.

B. *A KNN RESEARCH PAPER CLASSIFICATION METHOD BASED ON SHARED NEAREST NEIGHBOR.*

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

1. Authors:Yun-lei,CaiDuo Ji,Dong-feng Cai.
2. Publication:Proceedings of NTCIR-8 Workshop Meeting, June 1518, 2010, Tokyo.

C. A PROGRAMME IMPLEMENTATION OF SEVERAL INVENTORYCONTROL ALGORITHMS.
1. Authors:Vladimir Monov,Tasho Tashev.
2. Publication:Bulgarian Academy of Sciences,Cybernetics and Information Technologies,Volume 11,No1,Sofia

III. ALGORITHMS USED IN THE SYSTEM

A. *K-MEANS*

K-Means is one among the simplest unsupervised learning algorithms that solve the well acknowledged clustering problem. The procedure follows an easy manner to classify a given data set through an explicit number of clusters.[2][1]

Input:
- k: the number of clusters,
- D: a data set containing n objects.

Output:
- 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.

Advantages:
1. Quick, robust and very easy to understand.
2. Gives best result when data set are distinct or well separated from each other.

B. *KNN*

KNN algorithm is generally used to solve problems of classification and regression, but gives more efficient results when use for classification of training dataset. It is one of the instance-based learning, where the function is only approximated locally and all computational work is conceded until classification. [3][5]

Input:
- Object for which the corresponding class to be found.

Output:
- Class corresponding to the object.

Method:
1. Input: $D = \{(x_1, c_1), \ldots, (x_n, c_n)\}$.
2. $x = (x_1, \ldots, x_n)$ new instance to be classified.
3. For each labelled instance (x_i, c_n) calculate $d(x_i, x)$.
4. Order $d(x_i, x)$ in ascending order, $(i = 1, \ldots, n)$.
5. Select the K nearest instances to x: D^K_x.
6. Assign to x the most frequent class in D^K_x.

Advantages:
1. Robust to noisy data.
2. Can effectively handle large training dataset.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

IV. Implementation

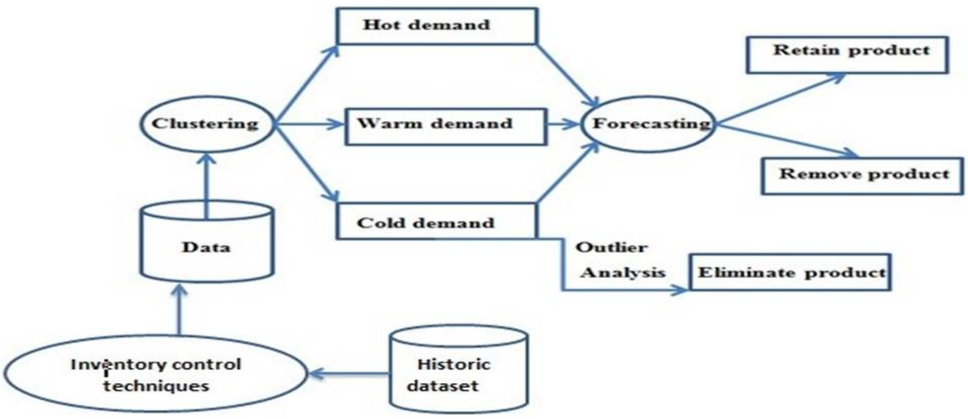

Figure 1: Overview of the system.

Working of the system is described in the following steps:

Step 1: **Admin Validation**

The Admin needs to enter the username and password to get access to the system. The credentials are compared with the data stored in the inventory and then the valid users are allowed to access the system.

Step 2: **Check inventory**

The system provides tables of monthly-sales and order. The monthly-sales table provides complete information of the sales done of a particular product in the current month and the order table contains the details of orders to be placed for the next month.

Step 3: **Order Stock**

Our System provides facility to Admin to order the product in two ways, firstly by considering the amount of stock left in the inventory and secondly to order the product which is currently not present in the inventory.

- The system makes use of inventory control technique [4] which is implemented with the help of following formula:

$R=\lambda/Q\ A+IC(Q/2+s)$ … (1.1)

Where, λ is rate of demand,

Q is optimal quantity of an order,

A is fixed cost of an order,

R is optimal average annual cost in the system,

I is inventory holding cost,

C is cost of a unit stock.

This equation provides us optimal quantity of an order (Q) which is used by our system.

- In the first way, admin needs to enter the category of product along with stock left percentage of the same, using which the system shows the list of products of that category which are below the stock left percentage and also provides the option of re-order of the product. Our system also returns the optimum value of stock, by using equation (1.1), which can be used as a reference value to place the order of adequate quantity for any particular product for the next month.
- In the alternative way, admin needs to provide all the details of product and then system provides the optimum value of stock by using equation (1.1) that can be used by the admin to place appropriate order of the respective product.

Step 4: **Categorization of products.**

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

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 (as shown in Figure 2).
It makes use of K-Means algorithms to perform categorization of the products present in the training dataset.[1]

- 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.

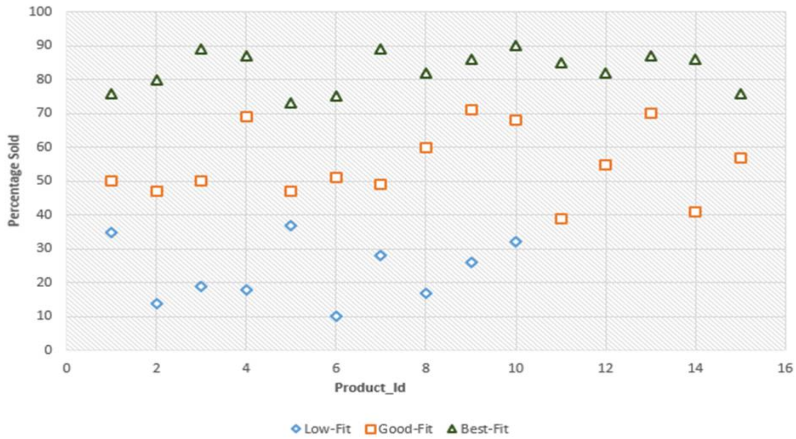

Figure 2: Categorization of products.

Step 5: **Classification of products.**
Our system provides facility to the admin to know the class of any particular product in two ways, firstly to know the class of product which is present in the inventory and secondly to know the class of product which is currently not present in the inventory.

- In the first way, admin needs to enter the category to get the list of products of that category, with the option of check class. It returns the class of the product by using KNN algorithm[5] which can be used as a reference to place the order for next month.
- In the second way, admin needs to enter all the details of that product and then the system provides the class to which the product belongs, by using KNN algorithm [5], which can be used as a reference to place the order for next month.

Step 6: **Forecasting**.
The system, SMARkeT allows us to perform predictive analysis on data present in the inventory and also provides flexibility to take decision to retain/remove the in/from inventory.

- In case Predictive analysis, admin needs to enter the category and then system will provide the percentage sold of that category by using linear regression method.

The Formula used in Linear Regression method is as follows [1]:
$y = a + b x$ …..(1.2)
Where, a& b are constants and are calculated by using equations 1.3, 1.4:
$a = \bar{Y} - b\bar{X}$ …..(1.3)
$b = (\sum X Y - n\bar{X}\bar{Y}) / (\sum X^2 - n\bar{X}^2)$ …..(1.4)
$\bar{X} = (\sum X)/n$ …..(1.5)
Where, X = Stock available
$\bar{Y} = (\sum Y)/n$, …..(1.6)
Where, Y = Percentage stock sold

- While taking decision to retain/remove the product in/from inventory, admin needs to provide the category and the threshold percentage of stock sold of the same. After processing this data, the system returns the list of products of that category, which are below the threshold value from the low-fit table, with the option to remove the product from the orders table.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

Step 7: **Statistics**.

Graphical representation in three different forms is done to get the overall idea about various products along with their selling percentage.

- In type 1, Bar graph for each month with categories on X -axis and average sold percentage on Y-axis.
- In type 2, Bar graph for various categories with various brands on X -axis and average sold percentage for each brand on Y-axis.
- In type 3, Linear graph for a year for various products with months on X -axis and sold percentage on Y-axis.

V. ALGORITHM USED FOR COMPARISON OF RESULTS OF THE SYSTEM.

A. *Divisive Hierarchical Clustering Algorithm.*

Steps:[7]

1. Consider one single large cluster with all objects (data points) in the same.
2. Decide initial value of means m1 & 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:
   - Check the sums of distance between items in a given sub-cluster and choose the one with the largest value.
6. Split the chosen cluster into sub-clusters by repeating steps 2 to 5.
7. Stop after obtaining desired number of clusters. (Our system requires 3 clusters- Best-fit, Good-fit, Low-fit).

B. *Comparisonbetween K- Means and Divisive Hierarchical Clustering*:

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 [1] 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 [8] of O $(m*n^2)$

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 Table 1. When Divisive hierarchical clustering [6] was used to categorize products in the inventory, we obtained clusters with a range of percentage sold as shown in Table 2. Figure 3 represents the graphical analysis of categorization of products in the inventory using K-Means algorithm whereas Figure 4 represents the graphical analysis of categorization of products in the inventory using Divisive Hierarchical Clustering algorithm. On comparing Figure 3 and Figure 4,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% share of range as compared to Divisive hierarchical Clustering where Low-fit Category was dominant with 50 % 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

Table 1: Categorization of products using K-Means clustering.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

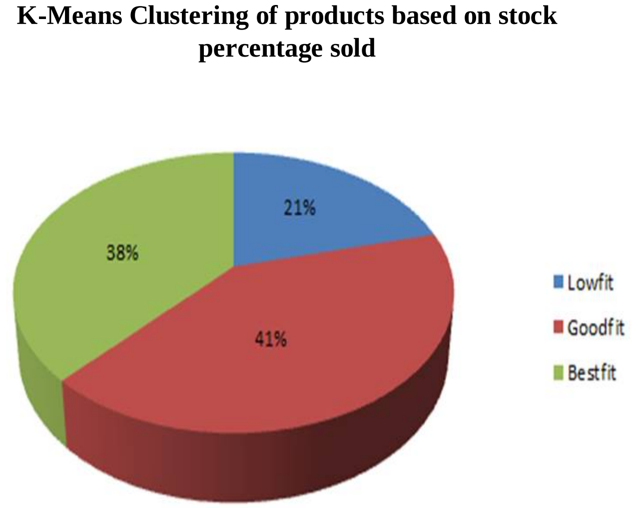

Figure 3: K-Means clustering of products based on stock percentage sold.

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

Table 2: Categorization of products using Divisive hierarchical clustering.

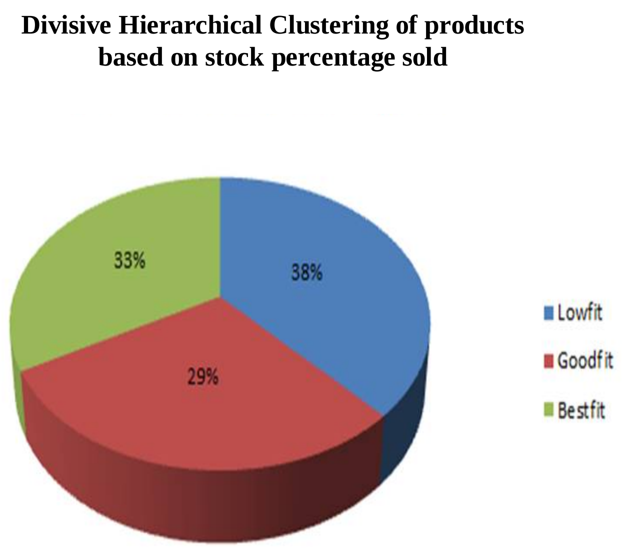

Figure 4: Divisive Hierarchical Clustering of products based on stock percentage sold.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

VI. FUTURE SCOPE

1. 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.

2. Our current system is performing most of the analysis on profit and stock sold of the products, but this analysis can be drilled down by adding intelligent analysis factor to each product in the inventory. This factor will be depend on the different priorities assigned to multiple attributes available for products in the system, like profit attribute will have highest priority (say 7), then percentage of stock sold with priority of 6, average customer rating with priority 5, and so on. Finally Intelligent Analytical Factor (IAF) will be calculated by the formula,

IAF(for product x)= (Priority value of attribute1)*(actual value of attribute1 of product X)+ (Priority value of attribute2)*(actual value of attribute2 of product X)+ ………...+(Priority value of attribute M)*(actual value of attribute M of product X)

The analysis that will be performing on the basis of this IAF of products will me much precise and will bolster to optimize business profit.

3. 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).

4. Along with this system can also be improved by integrating Market Basket Analysis on the products and giving special offers on the same (Example: 10% discount on total MRP of products if you buy ONE PLUS X mobile + SONY Power Bank + Skullcandy earphones)

5. The system, SMARkeT can have special module of CRM (Customer Relationship Management) to receive feedbacks, suggestions from customers and subsequently act on it.

6. 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.

VII. CONCLUSION

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-Means algorithm.We found 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.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 includes 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 quantity 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. Low-fit 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 maximize the sales and minimize the chances for loss in the future.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

REFERENCES

1. Jiawei Han, Jian Pei, Micheline Kamber, "Data Mining: Concepts and Techniques", The Morgan Kaufmann Series in Data Management Systems, Third Edition.
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 15–18, 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. http://biocomp.cnb.csic.es/~coss/Docencia/ADAM/Sample/Sample_Classification.pdf
6. https://www.youtube.com/watch?v=Fm01pqWLqzU.
7. http://www.cs.princeton.edu/courses/archive/spr08/cos435/Class_notes/clustering4.pdf.
8. https://en.wikipedia.org/wiki/Hierarchical_clustering.