Optimizing Sales and Marketing through Data Mining

[1]Ebin Eldhose,[2] Mr.A.Jeeva
[1]PG Scholar, [2] Assistant Professor,
[1, 2,] Computer Science and Engineering
Salem College of Engineering and Technology, Salem, 636111, India

***Abstract*—The present retail sector uses market rumours and advertisements in mass media and social networking to determine the stock levels and the type of stocks for the market. This causes many unexpected risks due to gap what these channels broadcast and reality. So we propose to find the marketing value of different products by collecting details from the retail outlets themselves and analyze these data programmatically to identify patterns in these sales data. These patterns can be used to create a more accurate picture of real market demand of every product.**

***IndexTerms*—Broadcast,Network,Patterns.**

I. INTRODUCTION (HEADING 1)

The inability of the traditional retail market to effectively analyze the products and determine the required stock level has given rise to heavy inefficiency and loss in the retail sector. Though large technological progress has made its impact on almost every sector, the retail sector has largely remained a sector that only absorbed the negative impacts of these changes.

Therefore the retail sector has become a sector that tries to manage this new rapidly changing world using the old traditional methodologies. So we propose the use of data mining technologies to accurately analyze the modern environment and create a more integrated and smart solutions to the retail sector. Using an integrated network of retailers and an efficient pattern finding algorithm we can easily predict the demand of any widely sold product for the immediate future. To do so, our network should have retail nodes providing data on present sales of every product that are sold through them and a central server that accumulate these data and analyze the rate of flow of products to each region. These analysis reports over an extended period are compared to produce a more or less accurate prediction of demand for each product or product category.

The main objective is to optimize the stock levels of different products in the retail outlet. This avoids the risk of building up of dead stocks and makes sure that enough amount of fast moving stocks are stocked up in the store, increasing the profit margin of retailer's and increasing customer satisfaction.

II. PROPOSED METHOD

A. Proposed System

This system consists of a networked system that has a central system that implements the analysis algorithms to predict the immediate future demand and many retail nodes connected to the central system, spread across many regions that transmits every sales happening through these nodes. For network connectivity we can use presently available networks like internet. This will enable us to have our retail nodes to be anywhere on the globe and also greatly reduce the cost of the total system.

This architecture also greatly reduce work load for the retail nodes. This enable us to turn any modern computers, tablets or smart phones with internet connectivity into a retail node, hence it will be easier to setup retail nodes anywhere in the world with very little cost. Simple web browsers or dedicated applications can be used in retail nodes to send and receive data. In case the retailer is using a connectivity that is not reliable to transfer real time data, the retailer node may store the sale data on the device itself. In such scenario the server will store the date and time of last reliable sale data for each retail node. Once the retailer node notifies the network as clear, the server may send the date and time of the last sale data from the retail node. The retail node then needs to send only those sale data that comes after the given time stamp. Once all data is successfully send to the server, the retail node may clear its local memory to free up memory for new data.

STATISTICAL PATTERN RECOGNITION

An increase or decrease in market demand for one commodity may have its impact felt in the future demand for many other commodities. Some of these relations are most obvious and we can trace a direct connection between the two commodities and can group them together. But for many others the relation is indirect and not easily traceable. Therefore the task of clustering together commodities to increase the accuracy of predictions is essential yet difficult to achieve. Thus it is better to use the established statistical pattern recognition algorithms like square error clustering to classify or cluster the commodity list.

In square error clustering, clustering is achieved by minimizing square error. We initially task the algorithm to find a fixed number of clusters using square error clustering. Then we evaluate the clusters by means of sampling. If the square error is found to be high for our predictions, then we can reduce the number of clusters to be found and reconstruct the clusters. If the clusters the system have found contains only the most obvious or incase the system haven't found some important clusters we can increase the amount of clusters to be found and reevaluate the clusters.

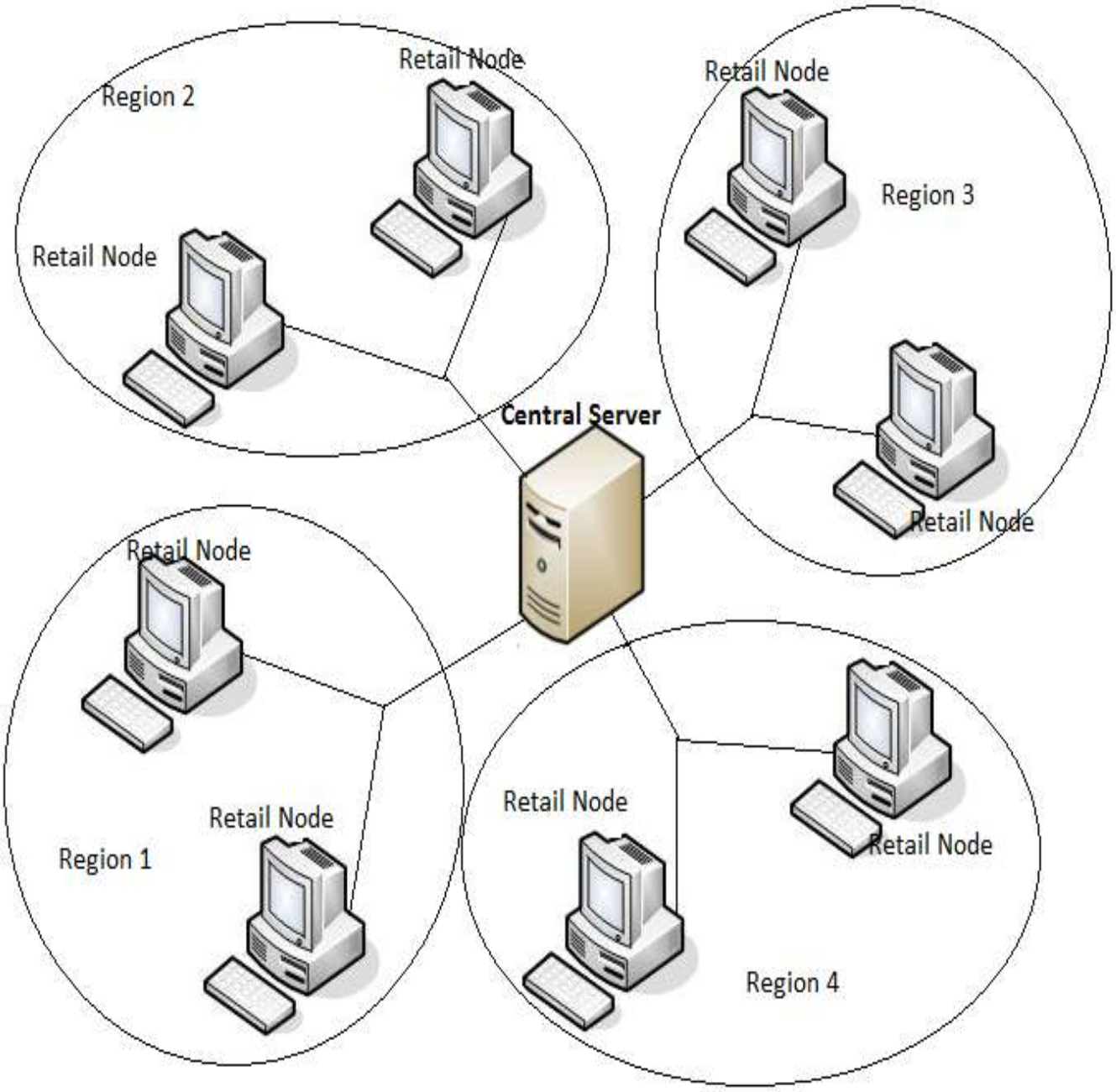

Fig.2.1Central System

CENTRAL SERVER

The central system is the main component of this system. It accepts data from retail nodes, stores those data in the database and produce predictions of future trends on requests from the retail nodes or any other applications that have sufficient privilege to request the report. The main functions of the central server will be:

a) Act as a central storage system that stores the data send by retail nodes all over the globe. The system should have a normalized and systematic database that stores the data send by the retail nodes along with details about the retail node, region from which the data was send along with date and time information.

b) Run the analysis algorithm periodically to generate reports on the rate of flow of products or product category. These reports also need to be stored in the database systematically for each product and product category along with any relevant data about the report like time period for which the report is generated.

c) Generate predictions for products or product category on request. This can be done by the following three steps. 1) Gathering reports generated about the requested product or product category. 2) Filter it further for the relevant region. 3) Identify the pattern of change in the product flow. 4) Extend the identified pattern to the immediate future to predict the future demand for the requested product or product category.

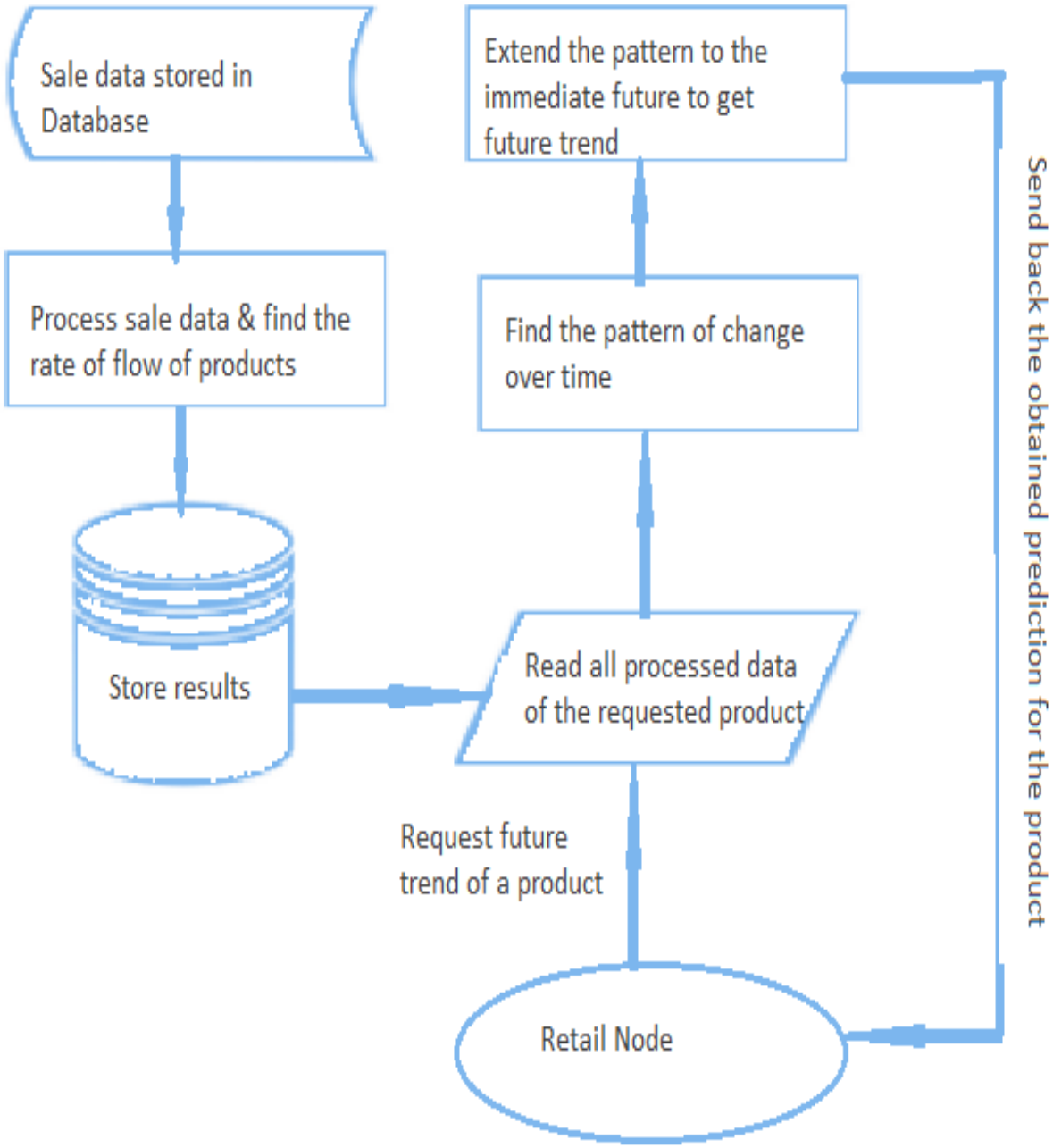

Fig.3.1

RETAIL NODES

Retail nodes are an essential part of the system. These retail nodes may smart phones, tablets or computers which are used by the retailer to connect to the central server to send sale data or to request and receive predictions about a commodity the retailer is interested in stocking up. In order to connect to the central server, these devices may use any available network including the internet. The main functions of the retail nodes will be:

a) Report to the central server, every sales happening through that retail node along with details about the product and quantity of sale.

b) Categorize products that are relevant to that retail node based upon the regional perspective. This decentralized method of product categorizing will produce more relevant categories and greatly reduce the work load of the system management team.

c) Request predictions about any products or product categories the retail outlet is planning to stock up.

d) Use the predictions provided by the central server to adjust the stock level for that product.

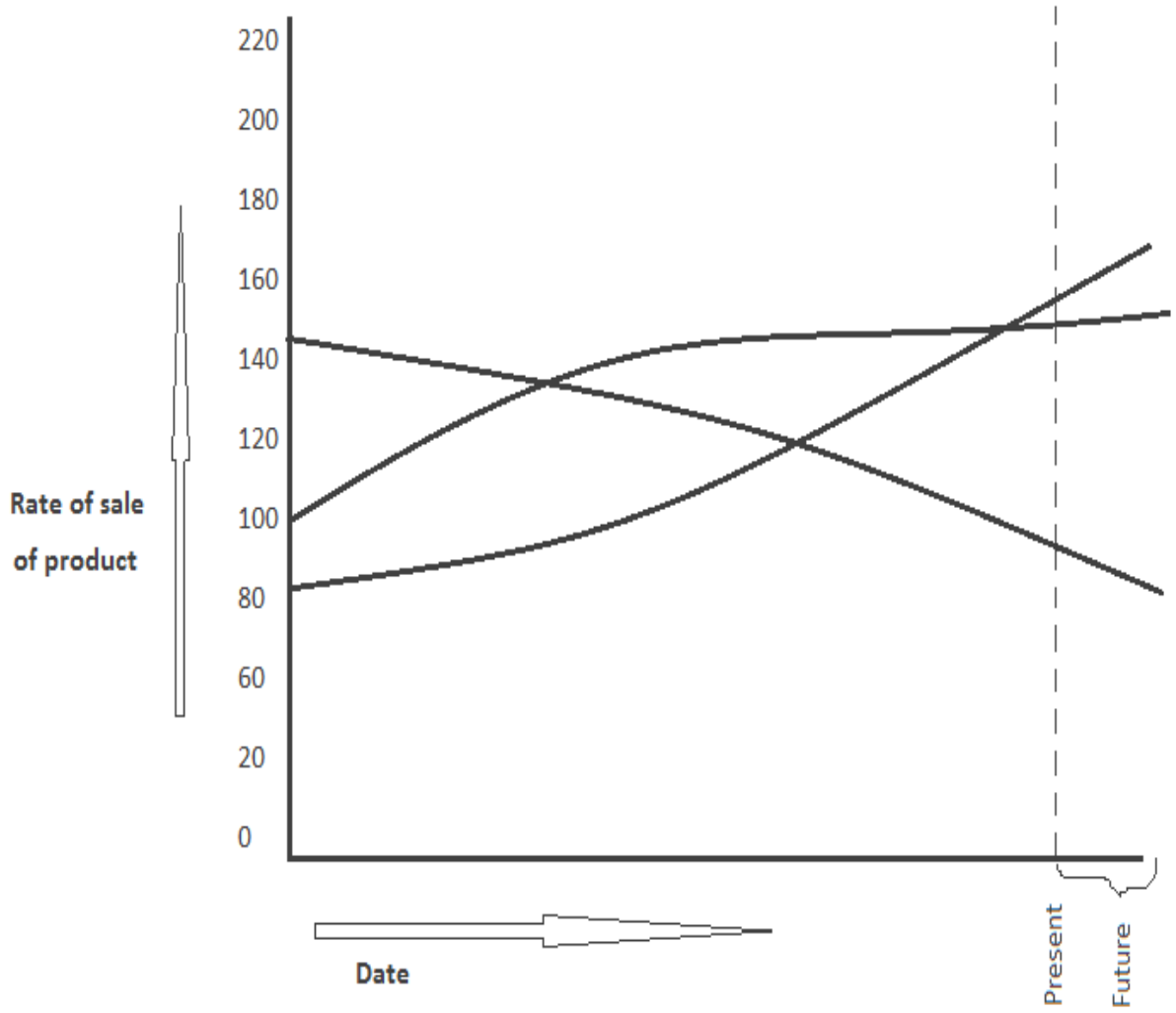

Fig.4.1

III. PERFORMANCE EVALUATION

Performance evaluation of this system can be calculated based upon the accuracy of the predictions created by the system. As everyone expects the system has a short initial period in which predictions made by the system is not reliable simply because of the very small amount of data it has to analyze and predict the trend. Even an increase in number of retail nodes will not significantly improve accuracy of the system to a level that can be considered reliable. This becomes quite obvious if we have reached a local minima or maxima at the period of prediction. This initial phase is due to the fact that the real world commodity demand is not something with a uniform rate of increase or decrease or even something that can be predicted using a complex equation. Many events around the world contribute to the increase or decrease of demand for a commodity.

But once we have acquired data over a length of time we start to get predictions that are somewhat reliable. We can start to consider the predictions to be reliable when they start to be more accurate than the average guess estimates made by the retailers no (who rely on these guesses stock up the commodities). As we move further on time we will see the system getting more accurate except on occasions of sudden upward or downward trend in commodities due to some local or global events. This increase in accuracy is due to the fact that the local maxima and minima that occur quite frequently in commodity market more or less cancels out each other to give a more accurate picture of commodity demand extended over a period of time.

IV. CONCLUSION

Incorporation of data mining technologies in the retail sector can have profound effect on the retail sector. Efficiency of the present day retail sector can be greatly improved if we can provide a centralized structure that is able to collect real time sales data and produce future predictions about the demand of different products. This will significantly reduce the risk of stocking up unnecessary dead stock that are not in demand or not having enough products in the retail nodes when the demand rises. Though this method cannot be used for long term predictions about a product, it can give reasonably accurate picture of demand in the immediate future.

REFERENCES

[1] Aurangazeb Khan, KhairulllahKhan,behramB.Baharuddin "Mining Frequent Patterns MinningOf Stock Data Using Hybrid Clustering Association Algorithm" International Conference 2009.

[2] Anil K. Jain, Fellow, IEEE, Robert P.W. Duin, and Jianchang Mao, Senior Member, IEEE"Statistical Pattern Recognition: A Review".

[3] Rachita Sony Krotha, SatishMuppidi, "Performance Evaluation of K-Means Algorithm andEnhanced Mid-point based K-Means Algorithm on Mining Frequent Patterns", 2013..

[4] NehaAggarwal,Kirti Agarwal " A Mid-point based K-mean clustering algorithm for DataMining", IJCSE 2012.

[5] http://www.roselladb.com/sales-trend- forecast.htm.

[6] Matt Hartely "Using Data Mining to predict inventory levels" IEEE, 2005.