A SURVEY OF TIME SERIES DATA PREDICTION ON SHOPPING MALL

Mohammed Ali. Shaik*,
Assistant professor in CSE department at Aurora's Research and Technological Institute (ARTI), Warangal, Andhra Pradesh, India.
niharali@gmail.com
http://niharali.hpage.com

S.Narasimha Rao,
Associate professor in CSE department at Aurora's Research and Technological Institute (ARTI), Warangal, Andhra Pradesh, India,
getnrao@gmail.com

Abdul Rahim,
Associate professor in Dept of ECE at Aurora's Scientific, Technological and Research Academy (ASTRA), Hyderabad, Andhra Pradesh, India,
rahim.astra@gmail.com

Abstract

Tremendous amount of data streams are often generated by dynamic environments such as stock's and bond's price indices, telecommunications data, audio and video data, Network traffic and data related to various Shopping malls. Mining regular patterns is one of the most important task in data mining. A time series database consists of various sequences of values that are obtained over a stipulated period of time. The values are typically measured at equal time stamps (eg., hourly, daily, weekly) which are sequence of ordered events, with or without concrete notations of time. The function is to mine all the transactional data which describes the behavior of various transactions. In an online business or in a shopping mall, the customers can purchase more than one item at a time. Frequent patterns are those that appear most often in a data set as a collection of various item sets or its subsequences. The algorithms like Apriori and FP Growth are used to mine the frequent patterns of a item set. The Apriori algorithm generates candidate set during its each iteration. It reduces the dataset by removing all the irregular itemsets which does not meet the minimum threshold values from the candidate sets. The most expensive phase of FP Growth algorithm is to generate a candidate set and to mine the database [1].

***Keywords*: Data mining; Time series; Frequent patterns;**

1. Introduction

A survey is a data collection tool used to gather the information about individuals or a organization or a firm or a enterprise, surveys are commonly used in research's to collect self-report data from study participants more over a survey may focus on factual information about individuals, or it might aim to collect the opinions of the survey takers.

Retail shopping behavior in India has changed spectacularly over the past century prior to the industrial revolution retail purchases were most often made at the shopping center neighboring to an individual's residence (i.e., the general store) a customer may travel to more distant shopping centers for a lower price or better selection was unaffordable due to shipping costs but with the popularization of the automobile in the mid-l920s, however, travel costs were reduced and the consumer became more mobile and with the larger use of the automobile and the use of mass transit systems, central cities became the main focal point of many retail purchases, in particular purchases of high-order goods.

A mall is said to be "a spreader of culture, a shopper's heaven to escape the chaos of daily life in a postmodern world and a mall is a space, which the individual 're-appropriates' in an effort to construct a self, and as a place where a woman might find the chance and space to resist the imposition of male values".

Shopping involves much more than eyesight and feet. "shopping is the kind of activity that involves experience that portion of the world or space that has been deemed for sale, using all our senses such as sight, touch, smell, taste, hearing as the means for choosing this or rejecting that; virtually all unplanned purchases and many planned ones, too come as a result of the shopper seeing, touch, smell or taste something that promise pleasure, if not total fulfillment.

There are many types of shopping behaviors' and shopper types in evolving shopping malls some of them are broadly classified into two categories based on their objectives which are obviously different firstly, Serviceable shoppers are those according to whom shopping is a form of work or a task which is to be accomplished, until they make a purchase, sometimes these kind of shoppers may make it as a goal to spend some sum of money and the other category constitutes of hedonistic shoppers who give importance to the time being enjoyment and pleasure that they experience during the shopping trip who also consider shopping as a leisure activity and may derive pleasure through it, along with the purchase of products, many researches have also stated that majority of shoppers combine both Serviceable and experiential values during their shopping activities.

To get good results one must acquire the ability to forecast the future based on the previous data that can push a personal or the organization forward. Time Series forecasting, which forecasts based on time controlled variable, is an important tool under this scenario, where the research aim is to predict the behavior of complex systems solely by analyzing the past data. With an immeasurable enhancement of computer speed and ability to process huge data , a new data model-the stream data has been emerged, where the stream data is coming continuously, quickly, changing with time, and may be irregular and infinite in the way.

Stream data mining is a process of finding and extracting probable information and knowledge hidden in the stream of data, which is useful but people do not know in advance as the stream data is on the nonstop arrival of large or even infinite data, which cannot be all stored and processed at once, hence many traditional data mining algorithms are not right and proper for the stream data mining. This paper propose a model of stream data time-series pattern in a dynamic shopping mall which aims to improve the prediction method by employing techniques from the statistical and data mining areas to improve the prediction accuracy.

2. Literature Review

Background of the study of shopping mall is a organized system where a group of people work collectively for accomplishment of a common purpose that is selling of various types and brands of products, In a shopping mall the roles and responsibilities are stated very openly without any uncertainty such as the positions occupied by different individuals are displayed in the form of organization chart.

A shopping mall is typically, a shopping complex that is associated by many walkways as well it provides huge entertainment options to the target consumers, which often contains one anchor store that consumes twenty five percent of its retail space. In addition to it a mall contains specialty stores for clothes, accessories, home needs, books, as well as food court, multiplexes and entertainment hubs.

Why shopping mall prices are less than the retail markets? Like any intelligent consumer, you ask how I can save up to 30% just by shopping at shopping mall or online the major dissimilarity is in the distribution steps a product takes from the manufacturer to the retail shelves where a distribution process which the consumers are familiar with, what we like to called as the "Walmart Way".

2.1. *Definition of data mining*

Data mining is one of the most cutting-edge researches in and off the field of the existing international databases and information decision making. From a practical point of view, it refers to extracting the previously unknown, the probable useful patterns or knowledge discovery from huge databases including association rules, time series, artificial intelligence, statistics, databases, multimedia databases, text stream databases, etc. The phrase data mining has frequently been used by statisticians, data analysts, and the management information systems (MIS) communities. Knowledge discovery in databases (KDD) refers to the overall process of discovering useful knowledge from data, and data mining refers to a precise step in this process (Fayyad et.al., 1996). In this research, data mining techniques will be applied to the data on shopping mall in order to predict the next day's or months or years, shopping mall scenario.

2.2. *Approaches to forecasting/ prediction*

- Trending or forecasting a continuation based on the shape of the line drawn by plotting previous outcomes which is a very basic approach to generating a future forecast. Where past data represents changes which are plotted to generate a proximity curve which is unfortunately a very inaccurate method when the framework is dynamic and changes can be easily made by external factors.

- Another more advanced way of prediction is to create a predictive model of the forecasted environment where the created model must comprise and evaluate all the variables showing effect on the results, the goal of the model is to predict the results in relation to a future pattern, Data mining technology is used to understand the connection between the inputs and the outputs of each model which is a more enhanced, more accurate and realistic way of forecasting or predicting for invention of a result set from a data set.

2.3. *Mining Time-Series Data*

A time-series database persists of sequences of values or events segregated over repeated capacity of time [1]. The values are typically calculated at identical time intervals (e.g., hourly, daily, weekly), Time-series databases are popular in many applications, such as shopping mall analysis, economic and sales forecasting, budgetary analysis, utility studies, inventory studies, yield projections, workload projections, process and quality control, execution of natural phenomena (such as atmosphere, temperature, wind, earthquake), scientific and engineering experiments, medical treatments, education and research areas. A time-series database is also a sequence database, where a sequence or successive database is any database that consists of sequences of ordered events, with or without tangible notions of time. For example, Web page traversal sequences and purchaser shopping transaction sequences are sequence data, but they may not be time-series data. With the developing deployment of a large number of sensors, telemetry devices, and other on-line data collection tools, the amount of time-series data is growing rapidly, often in the order of gigabytes per day (such as in stock trading) or even per minute (such as from NASA space programs). How can we find correlation relationships within time-series data? How can we analyze such vast numbers of time series to find alike or regular patterns, trends, bursts (such as sudden sharp changes), and outliers, with fast or even on-line real-time response? This has become more and more important and challenging problem. In this section, we examine many aspects of mining time-series databases, with a motivation on trend analysis and similarity search.

2.4. *The necessity of data mining in shopping mall*

Major aspects that are likely to affect the prices of shopping malls include: 1) What people expect its future discounts will be. 2) When the discounts are expected to be given based on the depreciation value calculated. 3) Which products are to be purchased such that the stocks will be made available and all the products will be sold out. 4) The quantity of risk involved. Data Mining is a process of abstracting unawared, probable and functional information and knowledge from plentiful, incomplete, noisy, fuzzy and stochastic data, this information and knowledge can't be achieved relying on a simple data search which consists of mainly three parts: data, information and domain knowledge. Data is the majorly valuable only when mobilize or converted into useful information where accessing the data is not the decisive goal of data mining in fact, the main aim of data mining is using that information to enhance business decision-making efficiency and to develop more suitable decisions. The shopping mall data is stream data, at the same time, the shopping mall data shares sequential nature, which can be used to analyze stream data based on the time-series pattern mining methods. In enhancement, there are many disturbing factors of product prices, making the price data show non-linear features, trends, exceptions, which transports new challenges to the traditional data mining algorithms.

3. Related Work

In this section we briefly discuss some key approaches for performing similarity search in time series data based on dimension reduction in a stream of data.

Faloutsos, Agrawal, and Swami (1993) used the Discrete Fourier Transform (DFT) to perform dimension reduction. The DFT was used to map the time sequences to the frequency domain and the index so the built was called as the Findex. For most sequences of practical interest, the low frequency coefficients are strong hence the first few Fourier coefficients are used to represent the time sequence in frequency domain these coefficients were indexed using the R*-tree (Beckmann, Kriegel, Schneider and Seeger 1990) for fast retrieval. The basis for this indexing technique is Parseval's theorem.

The F-index may raise false alarms but does not introduce false dismissals when pruning the data. The actual matches are obtained in a post-processing step wherein the distance between the sequences are calculated in the time domain and those sequences which are within Î distance are retained and the others are pruned. The F-index typically handles comparisons and whole matching of queries.

The F-index method was generalized by Faloutsos, Ranganathan and Lopoulos (1994) and called the STindex. In this technique, subsequence queries are handled by mapping data sequences into a small set of multidimensional rectangles in feature scope. These rectangles are indexed using spatial access methods like the R*-tree (Beckmann, Kriegel, Schneider and Seeger 1990) which is brought a significant change in the research area.

A sliding window is used to extract features from the data sequence resulting in a trail in the problem space. These trails are divided into sub-trails which can be represented by their (MBR) Minimum Bounding Rectangles, in place of storing all the points in a trail, only a few MBRs are stored as and when a query is presented to the database, all the MBRs intersecting the query region are retrieved.

Chan and Fu (1999) proposed to use the DWT in place of DFT for performing dimension reduction in time series data. Unlike the DFT which misses the time localization of sequences, their sub sequences, the DWT

allows time as well as frequency localization concurrently. The DWT thus bears more information of signals in contrast to DFT in which only frequencies are considered. The approach used by Chan and Fu (1999) employed the Haar Wavelet Transform for mapping high-dimensional time series data to lower dimensions.

A data dependent indexing scheme was proposed by Yi and Faloutsos (2000) and is known as the SVD method for dimension reduction. The database consists of n dimensional points. We map them on a *k*-dimensional subspace, where $k < n$, maximizing the variations in the chosen dimensions. An important drawback of this approach is the deterioration of performance upon incremental update of the index. Therefore the new projection matrix should be calculated and the index tree has to be reorganized periodically to keep up the search performance.

In PAA (Faloutsos, Ranganathan, and Lopoulos 1994) each time sequence say of length k is segmented into m equal length segments such that m is a multiple of k. The averages of segments together form the new feature vector for the sequence. The correct selection of m is very important because if m is very large, the approximation becomes very rough but if m is very small, the performance deteriorates.

3.1. *Stream Data Time Series Pattern*

The shopping mall price stream data shares the characteristics of realtime, continuous, orderly, large amount, rapid arrival, nonlinear, and timing, as well as on-line analysis of the attention requirements, which bring many challenges to the traditional data mining algorithms.

3.1.1. *Shopping mall Data Mining Algorithm Chosen Principles*

1. Single extended linear scanning: The algorithm has to read the stream data once in agreement with the data order.

2. Little space and time complexity: as It is an on-line main memory algorithm; the processing time of each data must not be too long, space complexity must not be unlimitedly raising as data amount in order to meet the data streams velocity and limited space.

3. Adapting to the dynamic changes and velocity of data approach, the results estimates is to a higher degree.

4. The aptitude of processing nonlinear problems: discount price data is non-linear, which requires the algorithm to be a influential algorithm of processing nonlinear problems to avoid losing or changing the data patterns based on the depreciation of the products.

5. Effectively dealing with cleaning of noise and null data: The discount price data system must be robust because of the non-linear requirements. The algorithm can deal with noise and null effectively by replacing with mean values or standard deviation or the most frequent values or the constant values given by a user.

6. Anytime online reaction or response to the request of the user or the administrators.

7. Summary data structure will not only support the goal of the current calculation method, but also support other calculations such as aggregate functions or other statistical functions.

3.1.2. *Key Technologies of the Shopping Mall Data Mining Algorithm*

Based on the analysis of the stream data mining the algorithms selected principles of the shopping mall analysis, the following will be introduce a number of key technologies universally used in these algorithms which are incorporated in summary data structure, window technology, attenuation factor, approximation technology, and so on.

1. The Summary Data Structure: As the stream data is far superior than the amount of data stored in the available memory or the capacity of the memory, the system cannot conserve all the scanned data in the memory, while the query and mining of stream data frequently require analysis the data. In order to dodge costly disk access time and cost, it is need to establish a summary data structure in the memory, to preserve the scanned information.

2. The Window Technology: Window is a restricted region of the stream data, that is, the pageant of the unlimited text stream data, so that all qualified operations are limited within the range of the window. As only part of the stream data in the window for processing, the query consequences are not accurate but approximate to the data given. Therefore, window technology can be regarded as alike query technology and can be used well. Sliding window is described in the following is also one type of windows.

3. The Attenuation Factor: In addition to window technology, an alternative kind process to abolish the historical data which influence on the existing calculation results is called as the attenuation factor. It gives an attenuation factor which diminishes over time incessantly to each item. The data multiply attenuation factor

prior to be involved in computation. Therefore, the data which shows influence to the results decrease gradually over time.

4. The Approximation Technology: Because of the strict time and space prototypes of data stream processing, there are relatively uncommon and defined stream data mining algorithms. For most algorithms, they diminish the complexity of space and time at the cost of reducing accuracy of the results. Thus, Window technology and attenuation factor can be viewed as approximation technologies.

3.2. *Evaluation of forecasting accuracy*

Let us assume that, for $i \geq 1$, the conditional allocation of Y_i given the earlier observations $Y^{i-1} = \{Y_j\}^{i-1}_{j=1}$ has the conditional mean mi and variance v_i . That is, $Y_i = m_i + e_i$, where e_i is the arbitrary error that displays the conditional uncertainty in Y at time i. Note that $E(e_i \mid Y^{i-1}) = 0$ with probability one for $i \geq 1$. Let $\hat{Y}_i$ be a predicted value of Y_i based on Y^{i-1}. "Eq. (1)" the one step in advance mean square prediction error can be shown or displayed using below mentioned equation:

$$E(Y_i - \hat{Y}_i)^2, \qquad (1)$$

Very uncertainly, it can be worn as a performance measure for predicting Y_i. Note that the conditional one step ahead. "Eq.(2)"forecasting mean square error can be divided into squared conditional bias and conditional variance as shown below:

$$E(Y_i - \hat{Y}_i)^2 = (m_i - \hat{Y}_i)^2 + v_i^2 , \qquad (2)$$

Here E_i represents the conditional anticipation given Y^{i-1}. The latter part is not in one's manage and is constantly present regardless of which method is used for prediction. Since v_i is the same for all predicts, it is rational to disconnect it in measuring the performance of a predicting procedure. "Eq. (3)" Accordingly, for predicting Y_i, we may describe the net loss using:

$$L(m_i - \hat{Y}_i) = (m_i - \hat{Y}_i)^2 , \qquad (3)$$

Let δ be a forecasting procedure that yields forecasts $\hat{Y}_1$, $\hat{Y}_2$, ….. $\hat{Y}_n$ times 1, 2 and so on. "Eq. (4)" We take into account the average net mean square error in prediction

$$= \frac{1}{n - n_0 + 1} \sum_{i=n_0+1}^{n+1} E(m_i - \hat{Y}_i)^2 \qquad (4)$$

From forecasting Y_{n0+1} up to Y_{n+1} . The observations $\{Y_i\}_{i=1}^{n0}$ are used for preliminary estimation and for each I = $n_0 + 1 \ldots n + 1$, one step ahead forecasts are gained with all the available observations. "Eq. (5)" For a particular case of $n_0 = n$, it is basically the net mean square error in prediction at time n , and will be represented as mentioned above formula.

For the purpose of correlating the forecasting procedures depending on real data stream, given a time series of size n, we will judge the (sequential) average square error in prediction as:

$$= \frac{1}{n - n_0} \sum_{i=n_0+1}^{n} (Y_i - \hat{Y}_i)^2 \qquad (5)$$

The performance measure of a forecasting procedure with n_0 essence a fraction (but not too small) of n. Clearly and unlike the theoretical measure as shown in above formula, it can be evaluated based on the data deservedly. In the past few years, there has been motivating work on assessing or comparing predictive truthfulness among competing forecasts under general loss and or possibly non-Gaussian errors; see, for example, Diebold and Mariano (1995), Swanson and White (1995), White (2000), Corradi, Swanson, and Olivetti (2001), and references therein. The proposed approaches can also grant constructive information for model comparison.

4. Proposed System

The traditional stream data time-series pattern mining algorithms are extra suitable for linear data, on the other hand, the object of shopping mall mining product price data, it shares non-linear features, and the non-linear data that cannot be hastily processed, because they may indicate a model or structure altering. Therefore, it is needed to sub-straight linear the product price data first, converting the continuous space into discrete space to do time-series pattern matching.

4.1. *Sub-straight-linear Method*

Sub-timing real-time or online product price data series extracting sub-sequence do advantageous to do verifications and forecasts of temporal patterns timely the way to remove patterns from the uninterrupted time series data stream is, splitting the original time series, converting the subsequence into a kind of symbol data such as string or a series of featured space point), then classifying such symbol data, producing similar pattern or patterns collection the major dilemma is how to split the time-series data stream

In the data stream time series mining, sub straight linear approach is one well known process to do partition on the time series data while this method is more suitable to people's visual occurrence, and research areas, the number of the index dimension is low, so it calculates much faster each segment of the data stream is converted into a straight line after sub- straight-linear processing using a kind of symbol data such as the slope of the line to show each segment. In this way, penetrating interval is converted into discrete space from continuous space.

4.2. *Pattern Extraction Algorithm*

The fundamental thought of the algorithm is to generate Sub-straight-linear the non-linear product data stream with the algorithm to reduce the nonlinear features of the product price data and convert the continuous space into discrete space for timing model searching; defining time series patterns to display the oscillation characteristics of product price data visually; by means of sliding window technology of stream data mining to sort out the outdated pattern timely and add novel patterns to be processed, thus it can reduce the memory burden to expand the algorithm efficiency. In short, the algorithm comprehends the conventional time-series models of discrete data mining similarity searching methods with stream data mining techniques, it can attain the equivalent and extraction to the product market time-series pattern similarity effectively.

Next is the anti-propagation of error is carried out by contrasting the tangible output with preferred output where every weight is revised and back propagated layer by layer from output layer to the hidden layer and input layer this process will be sustained until the output error of network is reduced to an satisfactory level or the predestined time of learning is achieved the processing outcome of information are exported by output layers to the outside.

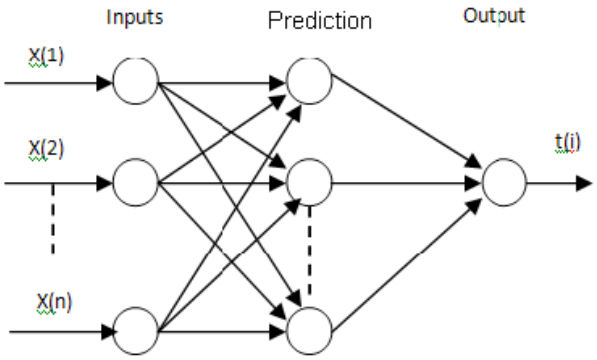

Fig. 1. Process of predicting sales.

Shopping mall prediction has been an area of concentrated significance due to the potential of pertaining a very high return on the invested money in a very short time. However, according to the well organized market hypothesis, all such attempts at prediction are pointless as all the information that could affect the actions of product worth or the market index must have been already integrated into the current market quotation.

A lot of studies have been organized, for example, in, which question the proficient market hypothesis by screening that it is, in fact, potential to predict the feature sales, with some degree of correctness, the future behavior of the shopping malls. Technical analysis has been used ever since a very long time but have had restricted triumph.

The algorithm used is Apriori which is a determining algorithm proposed by R. Agrawal and R. Srikant in 1994 for mining frequent itemsets for Boolean association rules the name of the algorithm is based on the fact that the algorithm uses former knowledge of frequent itemset properties, as we shall see following this algorithm employs an cyclic method known as a level-wise search, where k-itemsets are used to explore (k+1) item sets, firstly, the set of frequent 1-itemsets is originate by scanning the database to mount up the count for each item, and collecting those items that satisfy minimum support the ensuing set is representated by L1, next, L1 is used to find L2, the set of frequent 2-itemsets, which is used to discover L3, and so on, until no more frequent k-itemsets can be discovered. The resulting of each Lk requires one full scan of the database.

To improve the efficiency of the level-wise generation of frequent itemsets, a significant characteristics called the Apriori property, is used to reduce the search space [1].

Apriori property which is based on the following observation, if an itemset I does not satisfy the minimum support threshold, min sup, then I is not frequent; that is, P(I) < min sup. If an item A is added to the itemset I, then the ensuing itemset (i.e., I UA) occurs infrequently than I. Therefore, I U A is not frequent either; that is, P(I U A) < min sup.

This property belongs to a special group of properties called anti monotone in the logic that if a set does not overstep a test, all of its supersets will fail the same test as well. It is called anti monotone because the property is monotonic in the framework of failing a test.

One of the most trendy data mining method is to determine frequent itemsets from a transaction dataset and obtain the various association rules, discovering the frequent itemsets with frequency greater than or equal to a user given minimum support is not insignificant because of its combinatorial explosion, once frequent itemsets are generated, it is easy to generate association rules with confidence superior than or equal to a user indicated minimum confidence.

5. Implementation

The concepts described in this paper are implemented in Microsoft.Net 2010. All processes are operated user-interactively. The simulation of shopping mall prediction is shown below. Time-series data sets can be read and smoothed by moving average. The minimum support and minimum confidence values are given.

Sl.No	Product Id	Product Name	Type	Shop Name	Purchased times
1)	108	Wyo Just Did It Designer T Shirt	Clothing	Home	20
2)	262	Toshiba Hayabusa 8GB Pen Drive (White)	Memory Devices	Computers	20
3)	271	Whirlpool Ace 68i Semi-Automatic 6.8 kg	Washing Machine	Home Appliances	8
4)	109	Designer Stoll - 5	Clothing	Home	6
5)	122	FLORENCE 29A02074 SILVER WOMEN PARTY WEAR	Footwears	Home	6

Associations Generated are

Sl.No	Product Id's	Product Name's	Purchased times
1)	108, 262	Wyo Just Did It Designer T Shirt, Toshiba Hayabusa 8GB Pen Drive (White)	13
2)	262, 122	Toshiba Hayabusa 8GB Pen Drive (White), FLORENCE 29A02074 SILVER WOMEN PARTY WEAR	6
3)	265, 135	Panasonic CS-ZC15NKY 1.25 Tons, Samsung Galaxy Pocket S5300	5
4)	108, 271	Wyo Just Did It Designer T Shirt, Whirlpool Ace 68i Semi-Automatic 6.8 kg	4
5)	271, 109	Whirlpool Ace 68i Semi-Automatic 6.8 kg, Designer Stoll - 5	3
6)	102, 100	LOTTO DRUM BAG, Sennheiser PC 163D Headphones	3
7)	102, 101	LOTTO DRUM BAG, Maple MW4021 Brown Gents Wallet	3

Fig. 2. Prediction Results based on minimum support and confidence.

By taking the shopping mall data as data set the possible cluster generated based on the time series data is depicted as show in below figure:

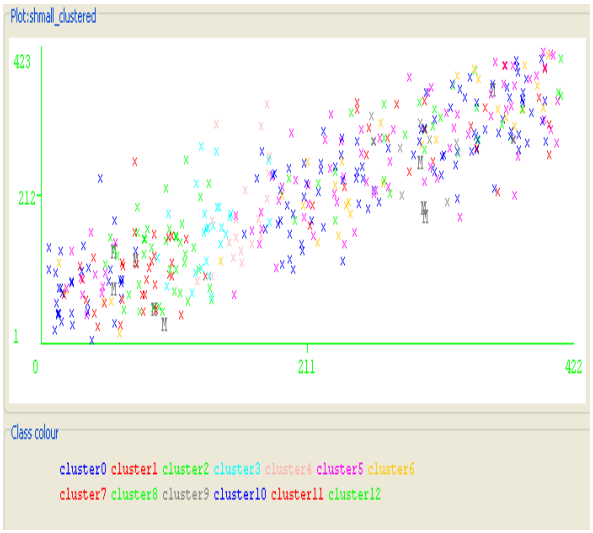

Fig. 3. Clusters generated on Time series data.

While cleaning the data the classifier errors generated in the shopping mall data set are:

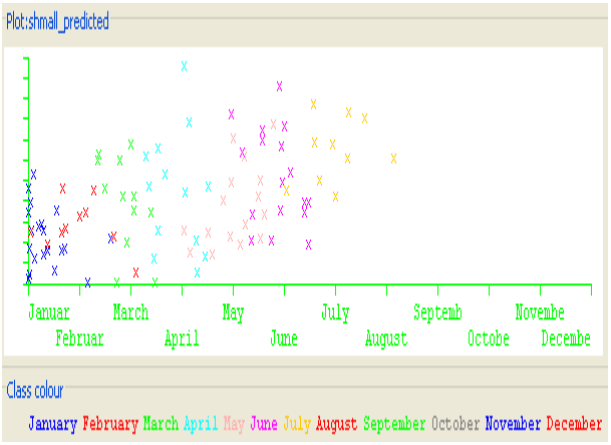

Fig. 4. Classifier Errors generated.

The Margin Curve generated is:

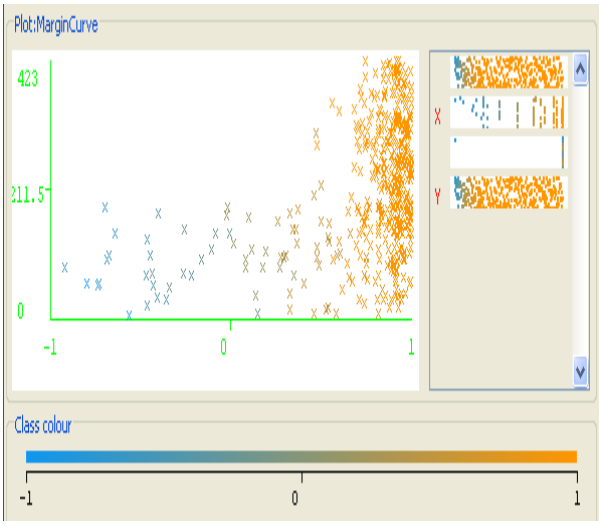

Fig. 5. Margin Curve.

By considering the training set the R* tree generated is:

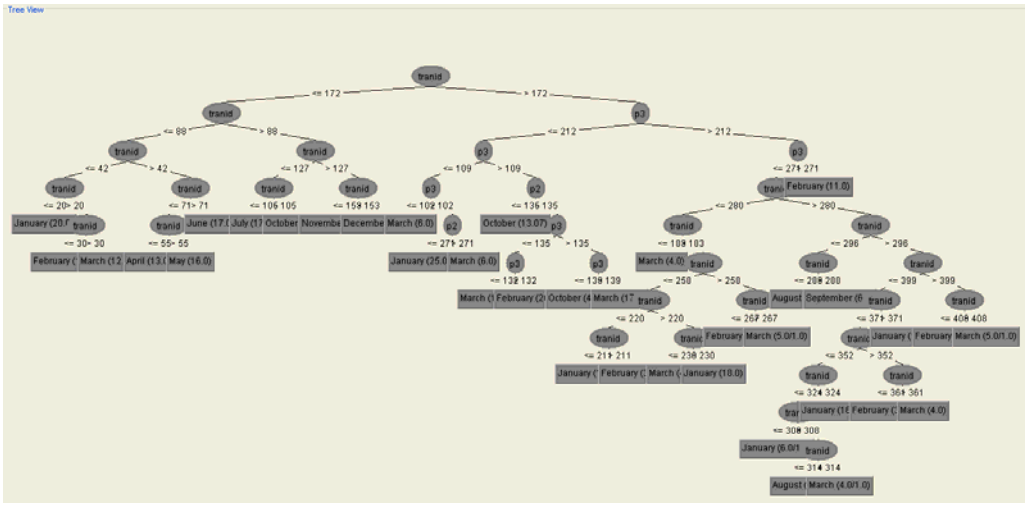

Fig. 6. R* Tree Generated based on data.

Cost curve generated based on time series, that is monthly basis is

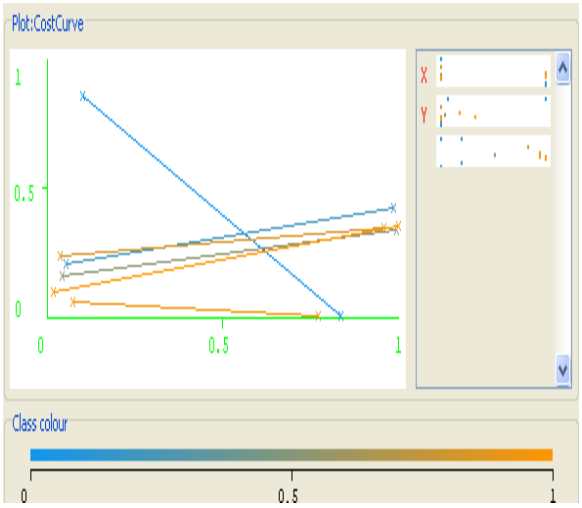

Fig.7. Cost curve generated for the month of January

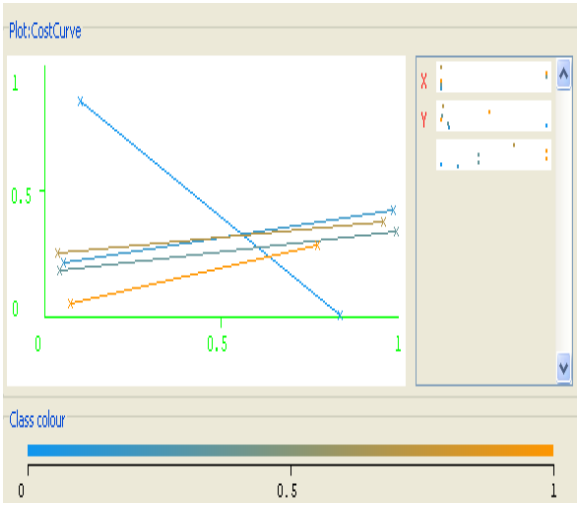

Fig.8. Cost curve generated for the month of February

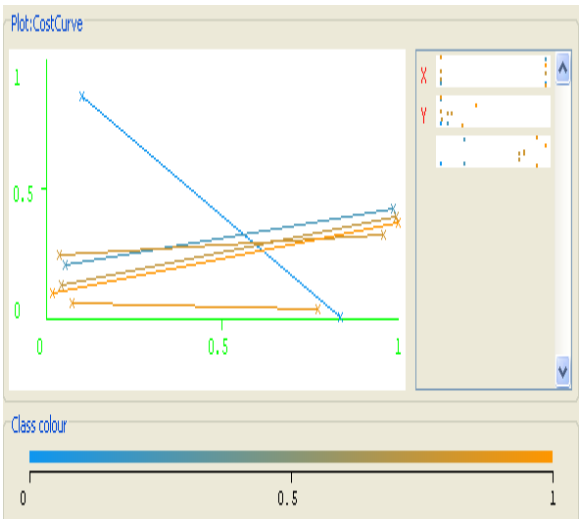

Fig.9. Cost curve generated for the month of March

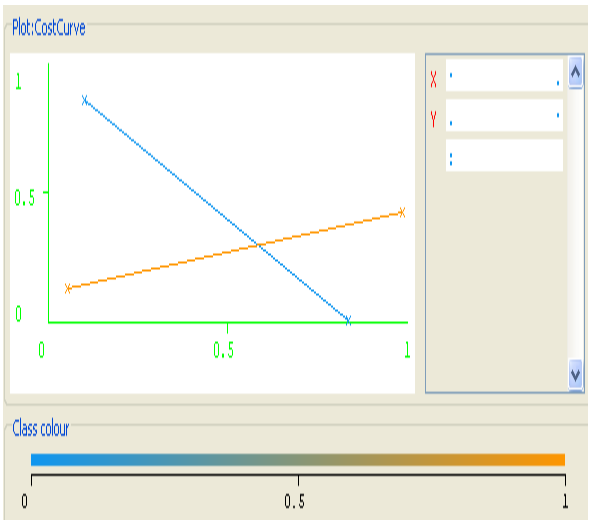

Fig.10. Cost curve generated for the month of April

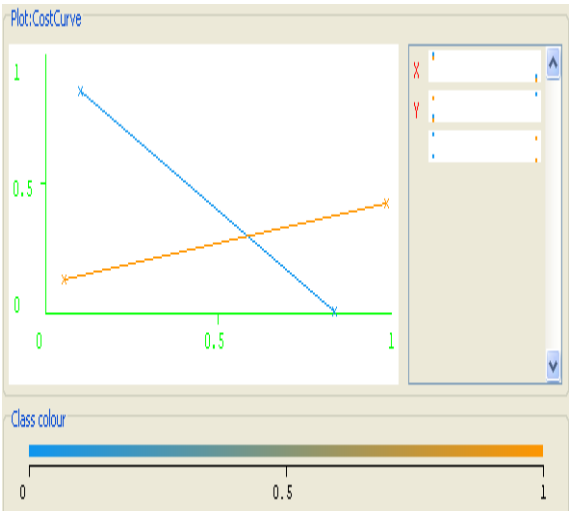

Fig.11. Cost curve generated for the month of May

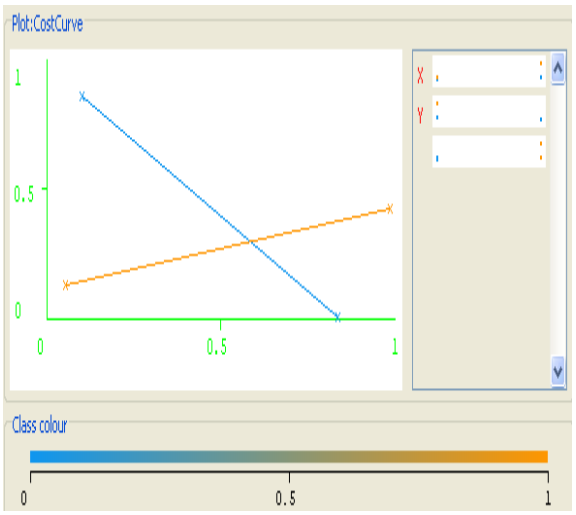

Fig.12. Cost curve generated for the month of June

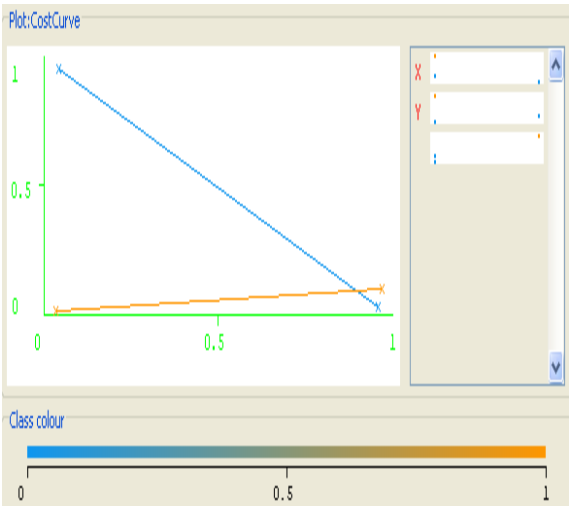

Fig.13. Cost curve generated for the month of July

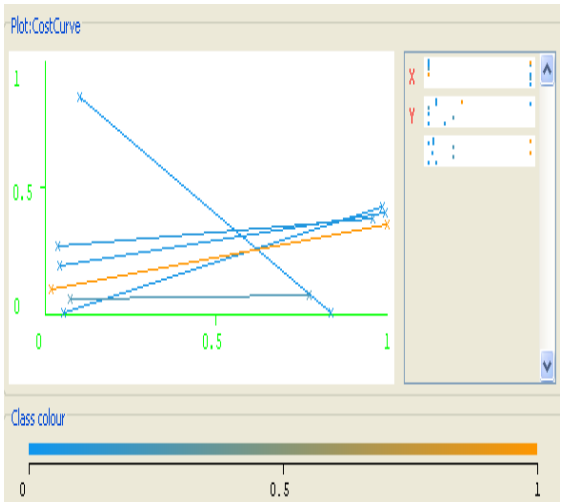

Fig.14. Cost curve generated for the month of August

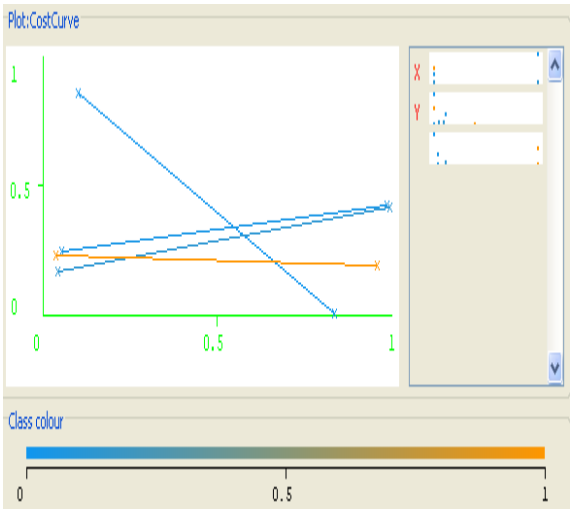

Fig.15. Cost curve generated for the month of September

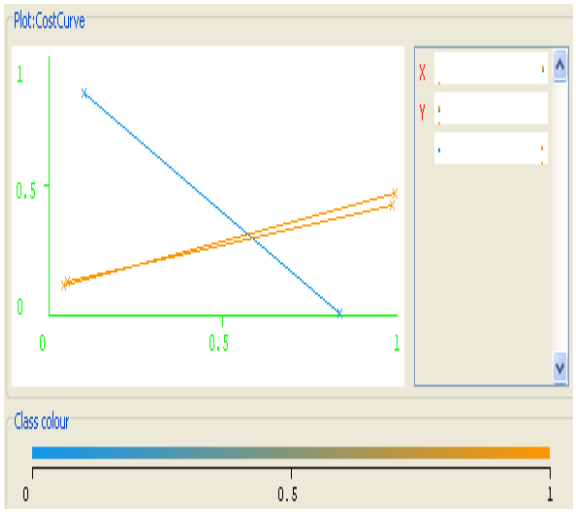

Fig.16. Cost curve generated for the month of October

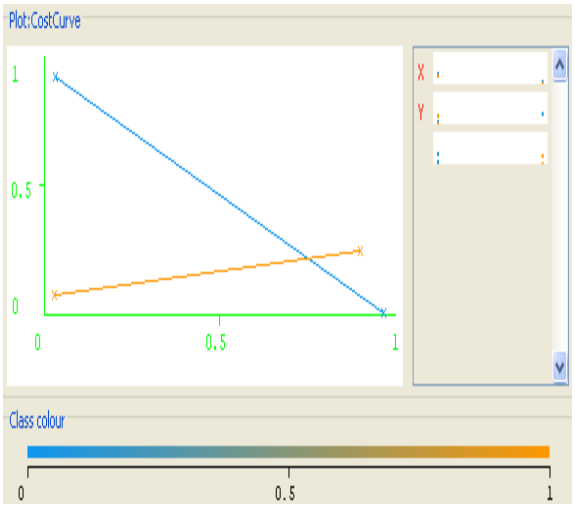

Fig.17. Cost curve generated for the month of November

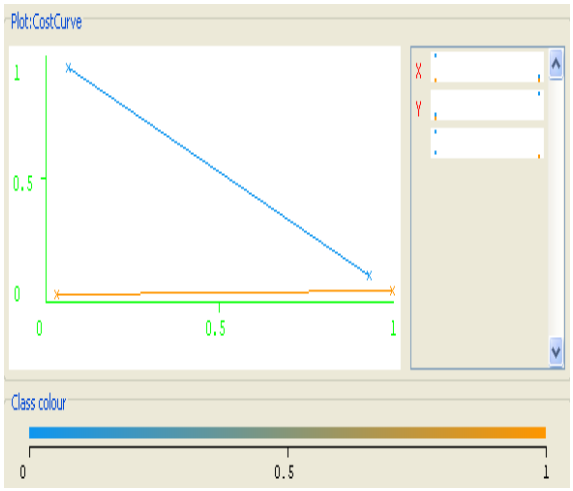

Fig.18. Cost curve generated for the month of December

6. Conclusion

We have conducted survey in many shopping malls, some of them are Reliance super, Moore, Megamart, etc., and gathered data such as the types of shoppers, types of products their sales based on the time and distinctive, manipulating the depreciation of the products and giving discounts.

Despite the time-series pattern mining can carry out reasonable forecast for the next phase of the product price trends, and the correctness of the results increases along if the correlation increases, because the subject of shopping mall analyzing system is subjective investors, who makes a significant error of the analyzing system as the complexity of the business investment and the affect on future events, and the error will increase radically as time increases, hence here in this paper we propose a tree based data mining algorithm that treats market's behavior and interest as input & filter the desired output efficiently & a mining model of stream data time-series pattern in a dynamic shopping mall, here we use the combination of decision tree for prediction of next day's or next month's sales prediction or forecasting the decision tree and classification are similar and have equivalent property thereby a decision tree can be used to provide a systematic design method for an effective decision making.

References

[1] Jiawei Han and Micheline Kamber, Data Mining Concepts and Techniques, by 2nd edition.
[2] Shweta Tiwari, Alka Gulati, "Prediction of Stock Market from Stream Data Time Series Pattern using Neural Network and Decision Tree", International Journal of Electronics & Communication Technology, 2(4).
[3] Khadar Babu SK, Karthikeyan K, Ramanaiah K and Ramanah D, "Prediction of Rain-fall flow Time Series using Auto-Regressive Models", Advances in Applied Science Research, (2011), pp. 128-133
[4] Shaminder Singh, Pankaj Bhambri and Jasmeen Gill, "Time Series based Temperature Prediction using Back Propagation with Genetic Algorithm Technique", IJCSI International Journal of Computer Science Issues, 8(5),(2011), pp. 1694-0814
[5] Weiwei Deng, Hong Peng, Jinsong Hong., "Data Stream Pattern Reliance Mining and Application", Computer Engineering, (2007).
[6] Jade M, Jayaraman V K and Kulkarni B D, "Improved time series prediction with a new method for selection of model parameters", Journal of physics a: Mathematical and general, (2006), pp.483–491.