

## TIME SERIES DATA PREDICTION ON SHOPPING MALL

Dr. B. Srinivasan<sup>1</sup>, P. Mekala<sup>2</sup>

<sup>1</sup>Associate Professor of Comp. Science, PG & Research Dept. of Comp. Science  
Gobi Arts & Science College (Autonomous), Gobichettipalayam – 638453  
srinivasan\_gasc@yahoo.com

<sup>2</sup>M.Phil Research Scholar, PG & Research Dept. of Comp. Science  
Gobi Arts & Science College (Autonomous), Gobichettipalayam – 638453  
gpmekala@gmail.com

---

### Abstract

Tremendous quantity of information streams are typically generated by dynamic environments like stock's and bond's worth indices, telecommunications information, audio and video information, Network traffic and information associated with numerous searching malls. Mining regular patterns is one among the foremost vital tasks in data processing. A time series database consists of sequences of values that are obtained over a stipulated amount of your time. The values are generally measured at equal time stamps (e.g., hourly, daily, weekly) that are sequence of ordered events, with or while not concrete notations of time. The perform is to mine all the transactional information that describes the behaviour of assorted transactions. In a web business or during a shopping mall, the shoppers should purchase over one item at a time. Frequent patterns are people who seem most frequently during information set as a group of assorted item sets or its sub sequences. The algorithms like Apriori and FP Growth are accustomed mine the frequent patterns of an item set. The Apriori algorithmic rule generates candidate set throughout its every iteration. It reduces the dataset by removing all the irregular item sets that doesn't meet the minimum threshold values from the candidate sets. The costly part of FP Growth algorithmic rule is to come up with a candidate set and to mine the information [1].

**Keywords:** Data mining, Time series, Frequent Patterns, Apriori Algorithm

---

### 1. Introduction:

Retail searching behaviour in India has modified hugely over the past century. Searching involves rather more than seeing and feet. "shopping is that the reasonably activity that involves expertise that portion of the planet or house that has been deemed available, victimization all our senses like sight, touch, smell, taste, hearing because the suggests that for selecting this or rejecting that; nearly all unplanned purchases and lots of planned ones, too come back as a results of the customer, touch, smell or style one thing that promise pleasure, if not total fulfilment.

There are varieties of searching behaviours' and shopper types in evolving shopping malls a number of them are loosely classified into 2 classes supported their objectives first off, Serviceable shoppers are those per

whom searching could be a sort of work or a task that is to be accomplished, till they create a sale, typically these shoppers could build it as a goal to pay some add of cash and also the different class constitutes of epicurean shoppers, provide importance to the present enjoyment and pleasure that they expertise throughout the searching trip who also think searching as a leisure activity and will fancy through it, together with the acquisition of product, several researchers have conjointly expressed that majority of shoppers mix each Serviceable and experiential values throughout their searching activities.

To get smart results one should acquire the flexibility to forecast the long run supported the previous information which will push a private or the organization forward. Time Series forecasting, that forecasts supported time controlled variable, is a vital tool underneath this situation, wherever the analysis aim is to predict the behaviour of complicated systems only by analyzing the past information. With associate degree immeasurable improvement of computer speed and capacity to method enormous data , a replacement information model-the stream information has been emerged, wherever the stream information is returning incessantly, quickly, dynamic with time, and will be irregular and infinite within the method.

Stream data mining could be a process of finding and extracting probable data and information hidden within the stream of knowledge, that is beneficial however persons don't recognize before because the stream information is on the nonstop arrival of enormous or perhaps infinite information, that cannot be all keep and processed right away, thus several data mining algorithms don't seem to be right and correct for the stream data processing. This paper propose a model of stream information statistic pattern during a dynamic shopping mall that aims to enhance the prediction technique by using techniques from the applied math and data processing areas to enhance the prediction accuracy.

The rest of this paper is organized as follows. Section 2 explains the review of literature. In section 3 is a brief review of related work. In section 4 describes the proposed approach. Section 5 explains the implementation of this research. The last section draws the conclusions and point out future directions of research.

## **2. Literature Review**

Background of the study of store could be a organized system wherever a set of individuals work jointly for accomplishment of a typical purpose that's commerce of varied sorts and makes of product, in a store the roles and responsibilities are fixed very frankly with none uncertainty like the positions occupied by different people area unit displayed within the kind of chart.

A shopping mall is often, a looking advanced that is associated by several walkways further it provides large diversion choices to the target shoppers, which frequently contains one anchor store that consumes 25 % of its retail house. Additionally a mall contains specialty stores for garments, accessories, home needs, and books, further as food court, multiplexes and diversion hubs.

### **2.1 Definition of Data Mining**

Data mining is one amongst the foremost stylish researches in and off the area of the prevailing international databases and knowledge higher cognitive process. From a sensible purpose of read, it refers to extracting the antecedently unknown, the probable helpful patterns or data discovery from vast databases as well as association rules, statistic, computer science, statistics, databases, multimedia system databases, text stream databases, etc. During this analysis, data processing techniques are going to be applied to the information on mall so as to predict ensuing days or months or years, mall state of affairs.

### **2.2 Approaches to forecasting/ prediction**

Trending or prognostication a continuation supported the form of the road drawn by plotting previous outcomes that may be a terribly basic approach to generating a future forecast. Wherever past information represents changes that are planned to get a proximity curve that is sadly a really inaccurate technique once the framework is dynamic and changes are often simply created by external factors.

Another additional advanced method of prediction is to make a prophetic model of the forecasted surroundings where the created model should comprise and assess all the variables showing impact on the results, the goal of the model is to predict the leads to respect to a future pattern, data processing technology is employed to know the association between the inputs and therefore the outputs of every model that may be a

additional increased, additional correct and realistic means of foretelling or predicting for invention of a result set from an information set.

### **2.3 Mining Time-Series Data**

Time-series information persists of sequences of values or events white over recurrent capability of your time [1]. The values area unit usually calculated at identical time intervals (e.g., hourly, daily, weekly), Time-series databases area unit fashionable in several applications, like store analysis, economic and sales prediction, monetary fund analysis, utility studies, inventory studies, yield projections, work projections, method and internal control, execution of natural phenomena (such as atmosphere, temperature, wind, earthquake), scientific and engineering experiments, medical treatments, education and analysis areas. A time-series information is additionally a sequence info, wherever a sequence or serial info is any info that consists of sequences of ordered events, with or while not tangible notions of your time. During this section, tend to examine several aspects of mining time-series databases, with a motivation on analysis and similarity search.

### **2.4 The necessity of data mining in Shopping mall**

Major aspects that area unit probably to have an effect on the costs of looking malls include: 1) what individuals expect its future discounts are. 2) Once the discounts area unit expected to tend supported the depreciation price calculated. 3) That merchandise area unit to be purchased specified the stocks are created accessible and every one the merchandise are sold-out out. 4) The amount of risk concerned. Data mining could be a process of abstracting unawared, probable and practical information and data from plentiful, incomplete, noisy, fuzzy and random knowledge, this info and data can't be achieved looking forward to an easy knowledge search that consists of primarily 3 parts: knowledge, information and domain data. Knowledge is that the majorly valuable only if mobilize or reborn into helpful information wherever accessing the information is not the decisive goal of knowledge mining in reality, the most aim of mining is victimization that information to reinforce business decision-making potency and to develop a lot of appropriate choices. The shopping mall data is stream knowledge, at a similar time, the store knowledge shares ordered nature, which might be wont to analyze stream knowledge supported the time-series pattern mining ways. In improvement, there are a unit several troubling factors of product costs, creating the value knowledge show non-linear options, trends, exceptions that transports new challenges to the normal data processing algorithms.

## **3. Related Work**

### **3.1. Stream Data Time Series Pattern**

The mall value stream knowledge shares the characteristics of realtime, continuous, orderly, great amount, fast arrival, nonlinear, and timing, yet as on-line analysis of the eye needs that bring several challenges to the normal data processing algorithms.

#### **3.1.1. Shopping mall Data Mining Algorithm Chosen Principles**

1. Single extended linear scanning: The algorithmic rule needs to scan the stream information once in agreement with the information order.
2. Very little area and time complexity: because it is an on-line main memory algorithm; the time interval of every knowledge should not be too long, area complexness should not be unlimitedly raising as knowledge quantity so as to satisfy the info streams rate and restricted area.
3. Adapting to the dynamic changes and speed of information approach, the results estimates is to a better degree.
4. The power of process nonlinear problems: discount value knowledge is non-linear, which needs the rule to be a potent rule of process nonlinear issues to avoid losing or dynamical the info patterns supported the depreciation of the merchandise.
5. Effectively managing improvement of noise and null data: The discount value system should be strong owing to the non-linear necessities. The formula will trot out noise and null effectively by substitution with mean values or variance or the foremost frequent values or the constant values given by a user.
6. Anytime on-line reaction or response to the request of the user or the directors.
7. Outline organisation won't solely support the goal of this calculation methodology, however conjointly support different calculations like mixture functions or different applied math functions.

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

Based on the analysis of the stream data processing the algorithms elect principles of the shopping center analysis, the subsequent are introduce variety of key technologies universally utilized in these algorithms that are incorporated in outline organisation, window technology, attenuation issue, approximation technology, and so on.

1. The outline knowledge Structure: because the stream knowledge is way superior to the number of knowledge hold on within the on the market memory or the capability of the memory, the system cannot conserve all the scanned knowledge within the memory, whereas the question and mining of stream knowledge oftentimes need analysis the info. So as to dodge expensive access time and price, it is have to be compelled to establish a outline organisation within the memory, to preserve the scanned data.
2. The Window Technology: Window may be a restricted region of the stream knowledge, that is, the pageant of the unlimited text stream knowledge, in order that all qualified operations area unit restricted inside the vary of the window. As solely a part of the stream knowledge within the window for process, the question consequences don't seem to be correct however approximate to the information given. Therefore, window technology is often thought to be alike question technology and might be used well. Window is represented within the following is additionally one form of windows.
3. The Attenuation Factor: additionally to window technology, an alternate kind method to get rid of the historical knowledge that influences on the present calculation results is named because the attenuation issue. It provides AN attenuation issue that diminishes over time ceaselessly to every item. The information multiply attenuation issue before be concerned in computation. Therefore, the info that shows influence to the results decrease bit by bit over time.
4. The Approximation Technology: thanks to the strict time and area prototypes of knowledge stream process, there square measure comparatively uncommon and outlined stream data processing algorithms. For many algorithms, they diminish the complexness of area and time at the value of reducing accuracy of the results. Thus, Window technology and attenuation issue will be viewed as approximation technologies.

## 4. Proposed Work

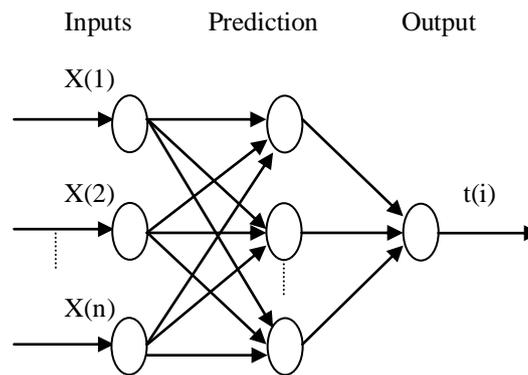
The traditional stream knowledge time-series pattern mining algorithms are further appropriate for linear knowledge, on the opposite hand, the article of store mining product value knowledge, it shares non-linear options, and therefore the non-linear knowledge that cannot be in haste processed, as a result of they will indicate a model or structure varying. Therefore, it is required to sub-straight linear the products rate knowledge initial, changing the continual house into distinct house to try to time-series pattern matching.

### 4.1. Sub-straight-linear Method

In the knowledge stream statistic mining, sub straight linear approach is one accepted method to try and do partition on the statistic knowledge whereas this technique is additional appropriate to people's visual incidence, and analysis areas, the amount of the index dimension is low, thus it calculates a lot of quicker every phase of the information stream is regenerate into a line when sub- straight-linear process employing a reasonably image data like the slope of the road to indicate every phase. During this method, penetrating interval is regenerate into distinct area from continuous area.

### 4.2. Pattern Extraction Algorithm

Fundamental thought of the rule is to come up with Sub-straight-linear the non-linear product information stream with the rule to cut back the nonlinear options of the products worth information and convert the continual house into separate house for temporal order model searching; shaping statistic patterns to show the oscillation characteristics of product worth information visually; by means that of window technology of stream data processing to delineated the obsolete pattern timely and add novel patterns to be processed, so it will scale back the memory burden to expand the rule potency.



**Figure 1:** Process of predicting sales.

Shopping mall prediction has been an area of focused significance attributable to the potential of pertaining a really high come on the invested with cash in a very short time. A lot of studies are organized, as an example, in that question the skilful market hypothesis by screening that it is, in fact, potential to predict the feature sales, with a point of correctness, the longer term behaviour of the searching malls.

The algorithmic program used is Apriori that is planned by R. Agrawal and R. Srikant in 1994 for mining frequent item sets that the algorithmic program uses former data of frequent item set properties, this algorithmic program employs Associate in Nursing cyclic methodology called a level-wise search, wherever k-item sets square measure wont to explore (k+1) item sets, firstly, the set of frequent 1-itemsets is originate by scanning the information, and grouping those things that satisfy minimum support the following set is represented by L1, next, L1 is employed to seek out L2, the set of frequent 2- item sets, that is employed to find L3, and so on, till no additional frequent k item sets is discovered. The ensuing of every Lk needs one full scan of the information.

To improve the potency of the level-wise generation of frequent item sets, a major characteristics known as the Apriori property, is employed to cut back the search area [1]. This property belongs to a special cluster of properties known as opposing monotone within the logic that if a collection doesn't overstep a check, all of its supersets can fail constant check additionally. It is known as opposing monotone as a result of the property is monotonic within the framework of failing a check.

## 5. Implementation

The ideas explain during this paper are enforced in Microsoft.Net 2010. All processes are operated user-interactively. The simulation of shopping mall prediction is shown below. Time-series knowledge sets will be scan and smoothened by moving average. The minimum support and minimum confidence values are given. By taking the shopping mall knowledge as knowledge set the potential cluster generated supported the statistic knowledge.

S.No	Product id	Product Name	Type	Purchased items
1	108	T-Shirt	Clothing	20
2	262	Toshiba Hayabusa 8GB Pen drive	Memory devices	20
3	271	Whirlpool Ace 68i Semi Automatic 6.8 kg	Washing Machine	8
4	109	Designer Saree	Clothing	6
5	122	Silver Women Party wear	Foot wears	6

**Table 1:** simulation of shopping mall prediction

Association generated are

**Table 2: Shopping mall prediction using association**

S.no	Product id's	Product Name's	Purchased Items
1	108,262	T-Shirt,Toshiba Hayabusa 8GB Pendrive	13
2	262,122	Toshiba Hayabusa 8GB Pen drive, Silver Women Party wear	6
3	108,271	T-Shirt,Whirlpool Ace 68i Semi Automatic 6.8 kg	4
4	271,109	Whirlpool Ace 68i Semi Automatic 6.8 kg, Designer saree	3

## 6. Conclusion

Mined in several shopping malls, and gathered knowledge like the kinds of shoppers, varieties of merchandise their sales supported the time and distinctive, manipulating the depreciation of the merchandise and giving discounts.

Despite the time-series pattern mining will perform affordable forecast for consecutive part of the products value trends, and therefore the correctness of the results will increase on if the correlation will increase, in this research have to propose a tree based mostly data processing algorithmic rule that treats market's behavior and interest as input and filter the specified output expeditiously and a mining model of stream knowledge time-series pattern in a very dynamic store, here to use the mix of decision tree for prediction of next day's or next month's sales prediction or prediction the decision tree and classification are similar and have equivalent property thereby a decision tree is accustomed offer a scientific style methodology for a good deciding.

## 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., 2011, "Prediction of Rain-fall flow Time Series using Auto-Regressive Models", *Advances in Applied Science Research*, pp. 128-133.
- [4] Shaminder Singh, Pankaj Bhambri and Jasmeen Gill., 2011, "Time Series based Temperature Prediction using Back Propagation with Genetic Algorithm Technique", *IJCSI International Journal of Computer Science Issues*, 8,5, pp. 1694-0814.
- [5] Weiwei Deng, Hong Peng, Jinsong Hong., 2007, "Data Stream Pattern Reliance Mining and Application", *computer Engineering*.
- [6] Jade M, Jayaraman V K and Kulkarni B D., 2006, "Improved time series prediction with a new method for selection of model parameters", *Journal of physics a: Mathematical and general*, pp.483-491.