Stock Control using Data Mining

M.S.Ankoshe
Pune University
Computer Dept.College Of Engineering
Kopargaon-423603

N.G.Pardeshi
Pune University
Computer Dept.College Of Engineering
Kopargaon-423603

J.N.Kale
Pune University
Computer Dept.College Of Engineering
Kopargaon-423603

ABSTRACT
Currently the firm owns three different shops at three different places.Present system is implemented using three different computers at all the shops which are not connected to each other by any means.The owner has to visit each and every shop and collect daily transaction and stock reports to get the data.These reports are then evaluated and used to order new stock. And hence "Stock Control using Data Mining" for shopping malls gives the idea about shopping mall's daily updations,details and recoveries, also we get decision over the malls.A centralized database management is very useful for any businessman who has more than one shops,outlets etc. Each and every shop is given a computer with the same software.All stock details entered by all the shops are maintained locally as well as centrally on server.Hence the proposed system will help in generating a decision support system for stock management, forecasting demands of the customer , calculating profitability of the shopping malls and comparing the stocks of two or more shops etc. All these operations can be implemented by using "Apriori" algorithm.Finally the system will be useful for any businessman who wants to have the control over stock of the items that are sold daily.

General Terms
Data Mining,Decision Making,Stock Management ,Database Management

Keywords
Apriori ,Prune Step,Join Operation,Frequent Itemsets

1. INTRODUCTION
"Data Mining is the application of statistics in the form of exploratory data analysis and predictive models to reveal patterns and trends in very large data sets".(Insightful Miner 3.0 user Guide)[1]Data mining has been defined in almost many ways by different authors "Data Mining is the process of exploration and analysis,by automatic or semiautomatic means,of large quantities of data in order to discover meaningful patterns and rules".(M.J.E.Berry and G.S.Linoff) "Data Mining is finding interesting structure in databases."(U.Fayyad,S.Chaudhari and P.Bradley)[1].With this Approach strong association rules also can be generated.

2. EXISTING SYSTEM
Currently the firm owns three different shops at three different places.Present system is implemented using three different computers at all the shops which are not connected to each other by any means.The owner has to visit each and every shop and collect daily transaction and stock reports to get the data.These reports are then evaluated and used to order new stock,etc.

3. PROPOSED SYSTEM
"Stock Control using Data Mining" for shopping mall gives the idea about shopping mall daily updations,details and recoveries.It will be helpful for taking decision over the malls Also centralized database management is very useful for any businessman who has more than one shops,outlets etc. Each and every shop is given a computer with the same

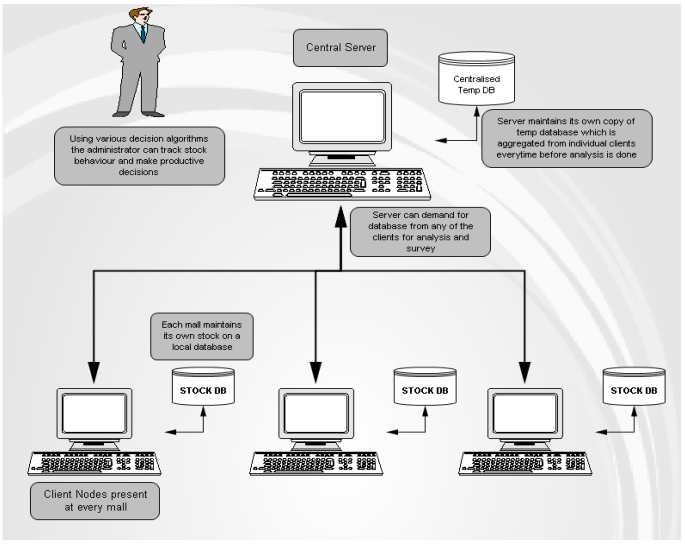

Fig 1: Proposed System

4. ALGORITHM
The Apriori Algorithm is an influential algorithm for mining frequent itemsets for boolean association rules[5].

1.Frequent Itemsets: The sets of item which has minimum support (denoted by Lifor ith-Itemset).

2.Apriori Property: Any subset of frequent itemset must be frequent.

3.Join Operation: To find Lk, a set of candidate k-itemsets is generated by joining Lk-1with itself.

4.1 Pseudo Code
Ck:Candidate itemset of size k

Lk:frequent itemset of size k

L1= {frequent items};

for(*k*= 1; *Lk*!=∅; *k*++) do begin

Ck+1= candidates generated from *Lk*;

For each transaction *t*in database do

increment the count of all candidates in *Ck+1*

that are contained in *t*
Lk+1= candidates in *Ck+1*with min_support
END

Return.

Table 1: Table of Transactions

TID.	List Of Items
T1	I1,I2,I5
T2	I2,I4
T3	I2,I3
T4	I1,I2,I4
T5	I1,I3
T6	I2,I3
T7	I1,I3
T8	I1,I2,I3,I5
T9	I1,I2,I3

1. Consider a database, D , consisting of 9 transactions.

2. Suppose min. support count required is 2 (i.e. min_sup = 2/9 = 22 %)

3. Let minimum confidence required is 70%.

4. We have to first find out the frequent itemset using Apriori algorithm.

5.Then, Association rules will be generated using min. support & min. confidence.

4.2 Steps

Step 1: Generating 1-itemset Frequent Pattern

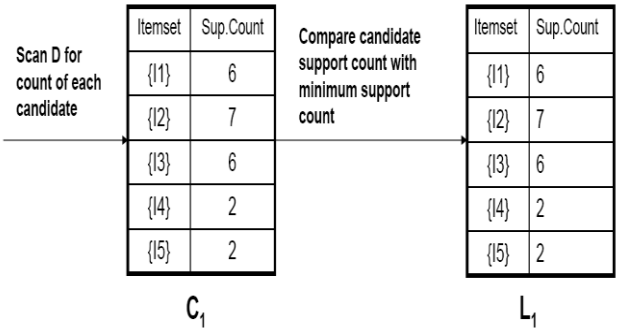

Fig 2: For comparing Candidate Support

The set of frequent 1-itemsets, L1, consists of the candidate 1-itemsets satisfying minimum support.

In the first iteration of the algorithm, each item is a member of the set of candidate.

Step 2: Generating 2-itemset Frequent Pattern

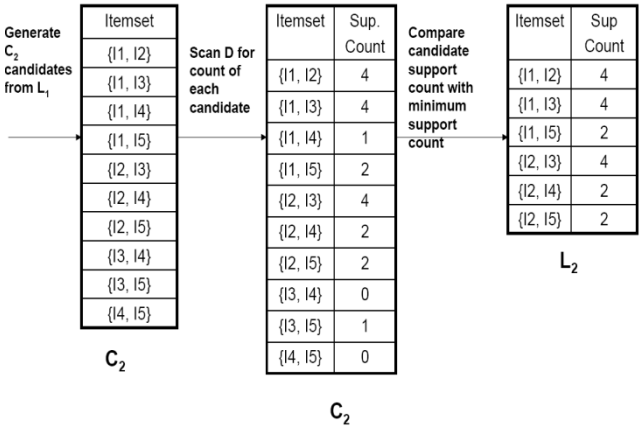

Fig 3: For Compare Candidate Support Count with Minimum Support Count

1 To discover the set of frequent 2-itemsets, L2, the algorithm uses L1 *Join* L1to generate a candidate set of 2-itemsets, C2.

2.Next, the transactions in D are scanned and the support count for each candidate

3itemset in C2is accumulated (as shown in the middle table).

3..The set of frequent 2-itemsets, L2, is then determined, consisting of those candidate 2-itemsets in C2having minimum support.

Step 3: Generating 3-itemset Frequent Pattern

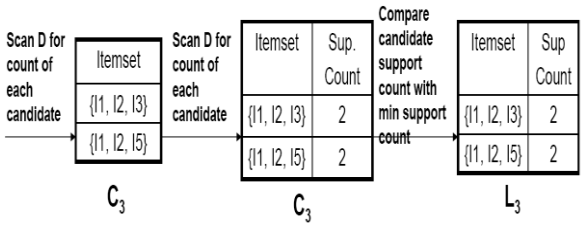

Fig 4: Fig For Generating 3-itemset Frequent Pattern

1 The generation of the set of candidate 3-itemsets, C3, involves use of the Apriori Property.

2. In order to find C3, we compute L2*Join*L2.

3.. C3= L2 *Join*L2 = {{I1, I2, I3}, {I1, I2, I5}, {I1, I3, I5}, {I2, I3, I4}, {I2, I3, I5}, {I2, I4, I5}}.

4. Now, Join stepis complete and Prune step will be used to reduce the size of C3. Prune step helps to avoid heavy computation due to large Ck.

•Based on the Apriori property that all subsets of a frequent itemset must also be frequent, we can determine that four latter candidates cannot possibly be frequent. How ?

Based on the Apriori property that all subsets of a frequent itemset must also be frequent.

1 .For example , lets take {I1, I2, I3}.The 2-item subsets of it are {I1, I2}, {I1, I3} & {I2, I3}. Since all 2-item subsets of

{I1, I2, I3} are members of L2, We will keep {I1, I2, I3} in C3.

2. Lets take another example of {I2, I3, I5}which shows how the pruning is performed. The 2-item subsets are {I2, I3}, {I2, I5} & {I3,I5}.

3. BUT, {I3, I5} is not a member of L2and hence it is not frequent violating Apriori Property. Thus We will have to remove {I2, I3, I5} from C3.

4. Therefore, C3= {{I1, I2, I3}, {I1, I2, I5}} after checking for all members of result of Join operationfor Pruning.

5 .Now, the transactions in D are scanned in order to determine L3, consisting of those candidates 3-itemsets in C3having minimum support.

Step 4: Generating 4-itemset Frequent Pattern

1. The algorithm uses L3 *Join*L3to generate a candidate set of 4-itemsets, C4. Although the join results in {{I1, I2, I3, I5}}, this itemset is pruned since its subset {{I2, I3, I5}}is not frequent.

2.Thus, C4= φ, and algorithm terminates, having found all of the frequent items. This completes our Apriori Algorithm.

Step 5: Generating 5-itemset Frequent Pattern

4.3 Procedure

1.For each frequent itemset ***"l",***generate all nonempty subsets of ***l.***

2.For every nonempty subset ***s***of ***l***, output the rule **"s ->(l-s)"**if support_count(l) / support_count(s) >= min_confwhere min_conf is minimum confidence threshold.

•Back To Example:

We had L = {{I1}, {I2}, {I3}, {I4},
{I5}, {I1,I2}, {I1,I3}, {I1,I5},
{I2,I3}, {I2,I4}, {I2,I5}, {I1,I2,I3},
{I1,I2,I5}}.
1–Lets take ***l*** = {I1,I2,I5}.

3.Its all nonempty subsets are {I1,I2}, {I1,I5}, {I2,I5}, {I1}, {I2}, {I5}.

4.Let minimum confidence thresholdis , say 70%.

The resulting association rules are shown below, each listed with its confidence.

1–R1: I1 ^ I2->I5

2.Confidence = sc{I1,I2,I5}/sc{I1,I2} = 2/4 = 50%

3.R1 is Rejected.

4–R2: I1 ^ I5 ->I2

5.Confidence = sc{I1,I2,I5}/sc{I1,I5} = 2/2 = 100%

6.R2 is Selected.

7–R3: I2 ^ I5 ->I1

8.Confidence = sc{I1,I2,I5}/sc{I2,I5} = 2/2 = 100%

9.R3 is Selected.

10–R4: I1 ->I2 ^ I5

11–R6: I5 ->I1 ^ I2

12.Confidence = sc{I1,I2,I5}/ {I5} = 2/2 = 100%

13.R6 is Selected.

5. TESTING OF SYSTEM

In this system initially server application is executed on server machine and client application is executed on client machine.After this once the successfull login of client side,

different trancactions are given on client side and submitted to server side and on server side decision will be taken on the stock of items by using apriori algorithm.

6. EXPERIMENTAL RESULTS

TID.	List Of Items
T1	I1,I2,I5
T2	I2,I4
T3	I2,I3
T4	I1,I2,I4
T5	I1,I3
T6	I2,I3
T7	I1,I3
T8	I1,I2,I3,I5
T9	I1,I2,I3

0

Consider a database, D , consisting of 9 transactions.

•Suppose min. support count required is 2 (i.e. min_sup = 2/9 = 22 %)

•Let minimum confidence required is 70%.

•We have to first find out the frequent itemset using Apriori algorithm.

•Then, Association rules will be generated using min. support & min. confidence.

Lets take ***l*** = {I1,I2,I5}.
Its all nonempty subsets are {I1,I2}, {I1,I5}, {I2,I5}, {I1}, {I2}, {I5}.
Let minimum confidence thresholdis , say 70%.
The resulting association rules are shown below, each listed with its confidence.

Based on minimum confidence values association rules are generated and they are selected if it is equal to or more than minumu threshold values.

1.R1: I1 ^ I2->I5

Confidence = sc{I1,I2,I5}/sc{I1,I2} = 2/4 = 50%

R1 is Rejected.

2.R2: I1 ^ I5 ->I2

Confidence = sc{I1,I2,I5}/sc{I1,I5} = 2/2 = 100%

R2 is Selected.

3.R3: I2 ^ I5 ->I1

Confidence = sc{I1,I2,I5}/sc{I2,I5} = 2/2 = 100%

R3 is Selected

Following observations are made from Apriori algorithm

1 Apriori algorithm scans transactions one by one hence it takes long time as compared to other data mining algorithms such as incremental and other algorithms

2. Incremental algorithm uses Hash Table for storing transactions so no need to rescan the transaction database and hence incremental algorithm requires less time as compared to Apriori algorithm

3 Apriori algorithm generates strong association rules based on minimum confidence value .Hence as the minimum threshold value increases more challenging association rules will be generated so that owner can have more benefits to operate their Shopping Malls.

7. CONCLUSION

We are able to control stock for shopping mall system by using concept of data mining. This helps us to build and control a MIS System for Inventory management. It also provides client-server architecture for centralized stock control for all branches.By using Apriori algorithm we can find association rules. In this we give the offers over the products by considering its minimum support count and minimum confidence. System provide best product offer generation. In Future by using data mining more strong association rules can be generated to take proper decisions over stock control

8. REFERENCES

[1] Savasere,E,Omiecinski,& S.Navathe,An efficient algorithm for Mining association Rules in Large Databases,Proc.of 21st Int.Conf.on Very large Databases,1995:175-186

[2] D.W.Choung,J.Han,V Ng, Maintainance of discovered association rules in large databases:An incremental updataing Technique,in Proc,1996 Int.Conf,Data Engineering.

[3] E.H.Han,G Kary Pis,V Kumar,Scalable parallel data mining for associate rules,In proc,1997,ACM-SIGMOD Int.Conf.Management of Data.

[4] J.Han,Y Fu,Discovery of multiple level association rules from large databases,In Proc.1995 Int.Conf.Very Large Databases

[5] Professor Anita Wasilewska explained , **"Apriori Algorithm"** in detail with example