

APPLICATION OF APRIORI ALGORITHM FOR ANALYZING CUSTOMER BEHAVIOR TO IMPROVE DEPOSITS IN BANKS.

MR.DEELIP. B. DESAI.
Asst.Professor,KIT'S IMER,
Kolhapur, India.
dbdesaikop@gmail.com

DR ABHIJEET KAIWADE
Professor, Dr. D.Y. Patil Institute of MCA.
Pune,India.
kaiwade@gmail.com

ABSTRACT

Many banks are in need of funds as their is lack of deposits from customers. There is growing competition among various sectors for banking industry, Banks are in search of customers who can deposit amount for fixed deposits, recurring deposits and so on. Various data mining tools can be used for this purpose, Specifically Apriori algorithm can be used to find association between customers and their behavior to keep deposits. This paper studies the process of applying Apriori algorithm with customer behavior to increase deposits with an empirical analysis. The result depicted in the study builds a model and help managers for decision making.

KEYWORDS: Data mining;, Apriori algorithm; frequent itemset; customer behavior;Banks.

I. INTRODUCTION

Today ,Banks to survive and grow it becomes critical to manage customers, build and maintain a healthy relationship with customers. . Data Mining in Banks can play a significant role ,the areas in which Data mining Tools can be used in the banking industry are customer segmentation, Banking profitability, credit scoring and approval, Predicting payment from Customers, Marketing, detecting fraud transactions, Cash management and forecasting operations, optimizing stock portfolios, and ranking investments. Various Data Mining techniques for data modeling are Association, Classification, Clustering, Forecasting, Regression, Sequence discovery Visualization etc. Some examples of some widely used data mining algorithms are Association rule, Decision tree, Genetic algorithm, neural networks, k-means algorithm, and Linear/logistic regression. Apriori algorithm is an influential algorithm for mining frequent item sets for boolean association rules. It can be used to find association between customer behavior and deposits. The goal of this study is to find, association between customer behavior and deposits, it uses frequent transaction of a customer.

II.APRIORI ALGORITHM

The Apriori Algorithm is an influential algorithm for mining frequent itemsets for Boolean association rules. Apriori is designed to operate on databases containing transactions. Apriori uses a "bottom up" approach, where frequent subsets are extended one item at a time (a step known as candidate generation), and groups of candidates are tested against the data. The algorithm terminates when no further successful extensions are found. Apriori uses breadth-first search and a tree

structure to count candidate item sets efficiently. It generates candidate item sets of length k from item sets of length $k - 1$. Then it prunes the candidates which have an infrequent sub pattern. According to the downward closure lemma, the candidate set contains all frequent k -length item sets. After that, it scans the transaction database to determine frequent item sets among the candidates.

The Apriori Algorithm is an influential algorithm for mining frequent itemsets for Boolean association rules. Following are the key concepts:-

- Frequent Itemsets:** The sets of item which has minimum support (denoted by L_i for i^{th} Itemset).
 - Apriori Property:** Any subset of frequent itemset must be frequent.
 - Join Operation:** To find L_k , a set of candidate k -itemsets is generated by joining L_{k-1} with itself.
 - Pseudo-code:**
 - Join Step:** C_k is generated by joining L_{k-1} with itself
 - Prune Step:** Any $(k-1)$ -itemset that is not frequent cannot be a subset of a frequent k -itemset
- C_k : Candidate itemset of size k
 L_k : frequent itemset of size k
 $L_1 = \{\text{frequent items}\};$
(for($k= 1; L_k!=0;k++$) do begin
 C_{k+1} = candidates generated from L_k ;
for each transaction t in database do
increment the count of all candidates in C_{k+1} that are contained in t
 L_{k+1} = candidates in C_{k+1} with min_support
end
return $U_k L_k$

How Apriori Works

1. Find all frequent itemsets:
 - i) Get frequent items:
 - (1) Items whose occurrence in database is greater than or equal to the min. support threshold.
 - ii) Get frequent itemsets:
 - (1) Generate candidates from frequent items.
 - (2) Prune the results to find the frequent itemsets.
2. Generate strong association rules from frequent itemsets rules which satisfy the min. support and min. confidence threshold.

III.APPLYING APRIORI FOR MINING CUSTOMER ASSOCIATION WITH DEPOSITS:
Sample Data

Customers deposits range and code

TABLE I

Account ID	Date	Withdrawal	Deposit	Closing Balance
1001	4-Mar-2014			44000.00
	15-Feb-2014		5000	39000.00
	10-Feb-2014	2000.00		41000.00
	6-Feb-2014		3000	38000.00
	4-Feb-2014		18000	20000.00
	28-Jan-2014	3000.00		23000.00
	15-Jan-2014	10000.00		33000.00
				6.00
1003	5-Mar-2014			24000.00
	17-Feb-2014	3000.00		27000.00
	5-Feb-2014	5000.00		32000.00
	2-Feb-2014	5000.00		37000.00
	1-Feb-2014		18000	19000.00
	1-Feb-2014		2000	17000.00
				5.00
1005	8-Mar-2014	6500.00		10000.00
				1.00
1006	25-Feb-2014			15000.00
	3-Feb-2014	2500.00		17500.00
	5-Feb-2014		2000.00	15500.00
				2.00
1007	28-Feb-2014			75000.00
	23-Feb-2014		1000	74000.00
	12-Feb-2014	3000.00		77000.00
	10-Feb-2014	500.00		77500.00
	1-Feb-2014	600.00		78100.00
	28-Jan-2014	8000.00		86100.00
				5.00
1008	3-Feb-2014			10000.00
	1-Feb-2014		3000	7000.00
				1.00

Average Transaction Deposits Calculations:

$$(ATD) = \frac{\text{Total Deposits for particular period}}{\text{Total Transactions for particular period}}$$

Deposit Code is given for all the average transactions done by each customers. The range for the Average Transaction Deposit(ATD) is given according to Deposit Code.

TABLE-II

Deposit Code	Average Transaction Deposits Ranges in(Rs)
D1	0-5000
D2	5001-10000
D3	10001-15000
D4	15001-20000
D5	>20001

For Example applying for above data.

Consider for a Account ID 1001.

26000

Average Transaction Deposits ----- = 4333

6

Types of Deposits table

TABLE-III

Deposit Code	Type of Deposit
FD	Fixed Deposit
ND	No Deposit(Customer have no any FD'S)

Types of Transactions:

Here Transactions Code are given for various probable FD and ND combinations.

TABLE-IV

Transaction Code	Deposits Code
T1	D1, FD
T2	D2,FD
T3	D3,FD
T4	D4,FD
T5	D5,FD
T6	D1,ND
T7	D2,ND
T8	D3,ND
T9	D4,ND
T10	D5,ND

Now support is calculated as percentile which in terms Confidence and Support

For Example

Applying for above sample data from table-I

TABLE-V

Account ID	Transaction Code
1001	T5
1002	T2
1003	T8
1004	T1
1005	T2

Now count the transactions of every Customers.

For. Ex.

Applying for above sample data from table-I

TABLE-VI

Transaction Code	Count
T1	23
T2	42
T3	35
..... and so on...

Total Count = Addition of all Transactions Count

Now calculate confidence as percentile
no.of count

$$\text{Confidence} = \frac{\text{no.of count}}{\text{Total no.of transactions.}} * 100$$

Transactions are co-related with each other which are above 60%. of total count so remove items which are less than 60 % . and thus we can get transactions (which may 3,4 or in any nos)

So for example if T4 IS 60% means D4,FD . D4 range is 15001-20000 and FD stands for Fix Deposit. Thus it concludes that

Customer which have range of 15001-20000 average transaction deposits have done Fix Deposits and are interested or have more chances of making Fixed Deposit's.

For Example :

Applying Apriori algorithm for available data for random sample bank, below results were found

TABLE-VII

Sr.No	Transaction Code	Count
1	T1	24
2	T2	85
3	T3	156
4	T4	188
5	T5	265
6	T6	220
7	T7	166
8	T8	110
9	T9	25
10	T10	50
Total Count		1279

no.of count

$$\text{Confidence} = \frac{\text{no.of count}}{\text{Total no.of transactions.}} * 100$$

1) For T1 Transactions

24

$$\text{Confidence} = \frac{24}{1279} = 1.87 \%$$

2) For T2 Transactions

85

$$\text{Confidence} = \frac{85}{1279} = 6.64 \%$$

3)For T3 Transactions

156

$$\text{Confidence} = \frac{\text{-----}}{1279} = 12.19 \%$$

and calculating in similar way for all other transactions following table is created.

TABLE-VIII

Sr.NO	Transaction Name	Confidence
1	T1	1.87
2	T2	6.64
3	T3	12.19
4	T4	14.69
5	T5	20.71
6	T6	17.20
7	T7	12.97
8	T8	8.60
9	T9	1.95
10	T10	3.90

From above table select 4 transactions which have maximum confidence,

Transactions T4,T5, T6 and T7 shows maximum confidence.

We can Conclude below from above interpretation:

Transaction T4 (14.69%) → D4,FD shows that customers have ATD(Average transaction Deposit) in range 15001-20000 and have done Fixed Deposit.

Transaction T5(20.71%) → D5,FD shows that customers have ATD(Average transaction Deposit) is >20000 and have done Fixed Deposit.

Transaction T6(17.20%) → D1,ND shows that customers have ATD(Average transaction Deposit) in range 0-5000 and not done Fixed Deposit (i.e No Deposits)

Transaction T7 (12.97%)→ D2,ND shows that customers have ATD(Average transaction Deposit) in range 5001-10000 and not done Fixed Deposit (i.e No Deposits)

From above interpretation it may be concluded that Customers with ATD range 15000/- and greater than 15000/- have done Fixed deposit and may be interested in future Deposits/Financial Schemes of the Bank. So whenever their may be any Financial schemes, Deposits Schemes this customers should be immediately contacted through e-mails/Sms.

IV. CONCLUSION

In this paper it has been described how Apriori algorithm is used for discovering frequent patterns for analysis of customers deposits. Various data mining techniques were used earlier for pattern analysis. However, for finding locally frequent items, Apriori is most suitable especially for transactional databases. This has lead to various improvisations of the core approach. Thus in Banking, data mining plays a vital role in handling transaction data and customer profile. From that, using data mining techniques a user can make a effective decision. Finally it can be concluded that Bank will obtain a massive profit if they implement data mining in their process of data and decisions.

V. REFERENCES:

- 1) Shao Fengjing, Yu Zhongqing. Data mining theory and algorithm [M]. Beijing: China Water Conservancy and Hydropower Press, 2003.
- 2) Zhang JianPing, Liu Xiya. k-means algorithm research and application based on clustering analysis [J]. Computer Application Research, 2007, (5) : 166-168.
- 3) Liu YingZi, Wu Hao. Customer segmentation method study review [J]. Journal of Management Engineering, 2006, 16 (1) : 53 -.
- 4) Ma HuiMin, Lu YiQing, Yin HanBin. Customer segmentation method based on the customer share [J]. Journal of Wuhan University of Technology, 2003, 25 (3) : 184-187.
- 5) Dr. Madan Lal Bhasin, "Data Mining: A Competitive Tool in the Banking and Retail Industries", The Chartered Accountant October 2006
- 6) R. Groth, Data mining: building competitive advantages, Prentice-Hall, 2001.
- 7) William W. Cooper, Lawrence M. Seiford, Kaoru Tone, Data envelopment analysis: a comprehensive text with models, applications, references and dea-solver software, Kluwer Academic Pub, 1999.
- 8) Rahul Mishra, Abhachoubey "Comparative Analysis of Apriori Algorithm and Frequent Pattern Algorithm for Frequent Pattern Mining in Web Log Data." International Journal of Computer Science and Information Technologies, Vol. 3 (4) 2012, 4662 - 4665 [9] A Comparative study of Association rule Mining Algorithm - Review Cornelia Györödi*, Robert Györödi*, prof. dr. ing. Stefan Holban
- 9) R. Agrawal, R. and R. Srikan, "Fast algorithms for mining association rules", Proceedings of the 20th international conference on very large data bases, 478-499, 1994. [11]W. Lin, S.A. Alvarez and C. Ruiz, Efficient Adaptive-Support Association Rule Mining for Recommender Systems, Data Mining and Knowledge Discovery, 6, Kluwer Academic Publishers, 2002, pp.83-105.
- 10) M. H. Dunham, Data Mining: Introductory and Advanced Topics, Prentice Hall , 2002
- 11) R. Hilderman, and H. Hmailton. "Knowledge Discovery and Interestingness Measures: a Survey", Tech Report CS 99-04, Department of Computer Science, Univ. of Regina, (1999)
- 12) J. Hipp, U. Guntzer, and G. Nakhaeizadeh. "Data Mining of Associations Rules and the Process of Knowledge Discovery in Databases", (2002)
- 13) A. Michail. "Data Mining Library Reuse Patterns using Generalized Association Rules", Proceedings 22nd International Conference on Software Engineering, (2000)
- 14) J. Liu, and J. Kwok. "An Extended Genetic Rule Induction Algorithm", Proceedings Congress on Evolutionary Computation (CEC), (2000)
- 15) P. Smyth, and R. Goodman. "Rule Induction using Information Theory", Knowledge Discovery in Databases, (1991)
- 16) T. Fukuda, Y. Morimoto, S. Morishita, and T. Tokuyama. "Data Mining Using TwoDimensional Optimized Association Rules: Scheme, Algorithms, and Visualization", Proceedings ACM SIGMOD Conference, (1996)
- 17) R. Srikant, and R. Agrawal. "Mining Generalized Association Rules", Proceedings 21st VLDB Conference, (1995)
- 18) <https://bizfluent.com/how-6633330-gain-banking-deposits.html>
- 19) <http://www.kanebankservices.com/WIB2%2009-05.pdf>