



GLOBAL JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY: G
INTERDISCIPLINARY

Volume 20 Issue 2 Version 1.0 Year 2020

Type: Double Blind Peer Reviewed International Research Journal

Publisher: Global Journals

Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Application and Analysis of Retail Inventory using Data Mining Techniques

By MD Imtiaz Uddin Adnan, Redoyan Raz, Tanvir Ahmed & A. H. M Saiful Islam

Notre Dame University

Abstract- Data mining is one of the most essential tools for gathering information from different datasets in almost all recent industries. In this 21st-century, data mining gained attention because of its significance in decision making, and it has become a key component in various industries such as retail. Inventory management requires pre-planned goals and attention to detail, and prioritizing items that require less attention can be a waste of time and resources. Learning indications about customers' shopping patterns by showing associations among various provides significant value in managing retail inventory. In the present research paper, popular data mining techniques have been applied and analyzed for multi-item inventory management in retail sales stores to show how data mining techniques can optimize and organize the retail inventory.

GJCST-G Classification: H.2.8



Strictly as per the compliance and regulations of:



© 2020. MD Imtiaz Uddin Adnan, Redoyan Raz, Tanvir Ahmed & A. H. M Saiful Islam. This is a research/review paper, distributed under the terms of the Creative Commons Attribution-Noncommercial 3.0 Unported License (<http://creativecommons.org/licenses/by-nc/3.0/>), permitting all non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Application and Analysis of Retail Inventory using Data Mining Techniques

MD Imtiaz Uddin Adnan ^α, Redoyan Raz ^σ, Tanvir Ahmed ^ρ & A. H. M Saiful Islam ^ω

Abstract- Data mining is one of the most essential tools for gathering information from different datasets in almost all recent industries. In this 21st-century, data mining gained attention because of its significance in decision making, and it has become a key component in various industries such as retail. Inventory management requires pre-planned goals and attention to detail, and prioritizing items that require less attention can be a waste of time and resources. Learning indications about customers' shopping patterns by showing associations among various provides significant value in managing retail inventory. In the present research paper, popular data mining techniques have been applied and analyzed for multi-item inventory management in retail sales stores to show how data mining techniques can optimize and organize the retail inventory.

I. INTRODUCTION

With increased globalization and advancement in technology, the retail market has become more and more dynamic, and therefore, retailers need a new approach to identify different objectives to be more competitive and successful. Inventory management is one of those key sectors that determine the success of a retailer. In today's ever-changing climate with a high level of uncertainty, keeping up with the demands leads to positive result on the market. Mining or extracting customer insight from structured and unstructured data and other sources is of tremendous importance for inventory management in retail stores. The change in customers' taste plays a significant part of what product is to be stored. Predicting which product will give more profit, products that are sold in unison, information like that is useful to store products in the inventory. Knowing which that product is out of fashion can help us in optimizing an inventory effectively. Some of the popular data mining techniques are –

- a) Clustering
- b) Association rules
- c) Decision tree

Data mining is finding and predicting hidden information from databases. It is a powerful technology with great potential to help organizations focus on the most accurate data in their data warehouses [1, 2, 3].

Author ^α ^σ ^ρ ^ω: Department of Computer Science and Engineering, Notre Dame University Bangladesh. e-mails: izunan131@gmail.com, redwanraz007@gmail.com, tanvirindub@gmail.com, saiful@ndub.edu.bd

Every technique can play its part in decision making and storing products, some more than others. The analysis from each of these techniques produces patterns which helps us to find valuable information. For a successful business, identification of high-profit, low-risk customers, retaining those customers, and bring future customers are important tasks for business owners and marketers.

II. RELATED WORK

In the last few years, the internet gives us new business concepts and also much information. As competitive pressure rises, the application of data mining process in customer's behavior becomes a excellent tool. [6]

Customer relationship management (CRM) aims at stronger loyalty of customers with feasible market share. With competition for shelf space intensifying, there is a pressing need to provide shoppers with a highly differentiated value proposition through "right product mix in the right amount at the right time." [7]

Customer relationship management (CRM) and customer profiles: Federated department stores are combining customer and transaction data to identify the best customers and offer exclusive extras. [8]

III. METHODOLOGY AND ANALYSIS

a) Clustering

The inventory space in a retail store is a precious commodity. To represent products seasonality, retail stores need to organize the products with care. Festivals and holidays should also be kept in mind when reforming the shelf. If a product has a large amount of sales in a day, it shouldn't dominate over other product storage. The storage priority is given to a product that has a high sale rate for an extended period.

To represent products from a retail store we have used this set as an example-

Row No.	Channel	Region	Fresh	Milk	Grocery	Frozen	Detergents_...	Delicassen
1	2	3	12669	9656	7561	214	2674	1338
2	2	3	7057	9810	9568	1762	3293	1776
3	2	3	6353	8808	7684	2405	3516	7844
4	1	3	13265	1196	4221	6404	507	1788
5	2	3	22615	5410	7198	3915	1777	5185
6	2	3	9413	8259	5126	666	1795	1451
7	2	3	12126	3199	6975	480	3140	545
8	2	3	7579	4956	9426	1669	3321	2566
9	1	3	5963	3648	6192	425	1716	750
10	2	3	6006	11093	18881	1159	7425	2098
11	2	3	3366	5403	12974	4400	5977	1744
12	2	3	13146	1124	4523	1420	549	497
13	2	3	31714	12319	11757	287	3881	2931
14	2	3	21217	6208	14982	3095	6707	602

ExampleSet (440 examples, 0 special attributes, 8 regular attributes)

The data set given above is a series of data set representing the amount sells of each of the product weeklies.

To get useful information out of this data set, we use a simple clustering technique, which is k-means clustering.

K-means clustering

K means defines a prototype in terms of a centroid, which usually the mean of a group of points and it is used for objects in a continuous n-dimensional space. Centroid never corresponds to an actual data point.

To reduce the dominance of a product after one day of massive amount of sell over our inventory, we normalize the dataset. After normalizing the data set, we get-

Row No.	Channel	Region	Fresh	Milk	Grocery	Frozen	Detergents_...	Delicassen
1	1.447	0.590	0.053	0.523	-0.041	-0.589	-0.044	-0.066
2	1.447	0.590	-0.391	0.544	0.170	-0.270	0.086	0.089
3	1.447	0.590	-0.447	0.408	-0.028	-0.137	0.133	2.241
4	-0.690	0.590	0.100	-0.623	-0.393	0.686	-0.498	0.093
5	1.447	0.590	0.839	-0.052	-0.079	0.174	-0.232	1.298
6	1.447	0.590	-0.205	0.334	-0.297	-0.496	-0.228	-0.026
7	1.447	0.590	0.010	-0.352	-0.103	-0.534	0.054	-0.347
8	1.447	0.590	-0.350	-0.114	0.155	-0.289	0.092	0.369
9	-0.690	0.590	-0.477	-0.291	-0.185	-0.545	-0.244	-0.275
10	1.447	0.590	-0.474	0.718	1.150	-0.394	0.953	0.203
11	1.447	0.590	-0.683	-0.053	0.529	0.274	0.649	0.078
12	1.447	0.590	0.091	-0.633	-0.361	-0.340	-0.489	-0.364
13	1.447	0.590	1.559	0.884	0.400	-0.574	0.210	0.499
14	1.447	0.590	0.729	0.056	0.740	0.005	0.802	-0.327
15	1.447	0.590	1.000	0.497	0.436	-0.572	0.456	0.228

ExampleSet (440 examples, 0 special attributes, 8 regular attributes)

Basic K-means algorithm-

1. Select k points as the initial centroid
2. Repeat
3. For k number of clusters by assigning each point closest to its centroid
4. Recomputed the centroid of each cluster
5. Until centroids do not change.

To apply k-means clustering to our data set we have used software known as Rapid Miner studio.

After applying k-means algorithm, we get-

Cluster Model

```
Cluster 0: 240 items
Cluster 1: 1 items
Cluster 2: 2 items
Cluster 3: 103 items
Cluster 4: 1 items
Cluster 5: 57 items
Cluster 6: 3 items
Cluster 7: 2 items
Cluster 8: 31 items
Total number of items: 440
```

And the centroid table-

Attribute	cluster_0	cluster_1	cluster_2	cluster_3	cluster_4	cluster_5
Channel	-0.681	-0.690	-0.597	1.447	1.447	-0.690
Region	-0.071	0.590	0.113	-0.056	0.099	-0.056
Fresh	-0.239	1.965	2.094	0.313	-0.331	0.792
Milk	-0.384	5.170	-0.118	3.917	0.439	0.561
Grocery	-0.467	1.286	-0.214	4.271	0.647	-0.011
Frozen	-0.056	6.893	0.674	-0.004	-0.328	9.242
Detergents_Paper	-0.439	-0.554	-0.430	4.613	0.664	-0.464
Delicassen	-0.184	16.460	0.378	0.503	0.044	0.932

In our dataset, the optimal number of $k=6$ from the performance vector.

In this dataset, the average sell is -1 and higher the disparity from -1 the larger or smaller the amount of sell. We can see that for milk in week one, the amount of sell=-0.384 and for week two is 5.170. The highest disparity from all the weeks is week two that suggests the high amount of sell. If this is the 1st week of November, then there is a high chance to sell this time next year, so for future storage, we can use this information and store a high amount of milk or a high amount of frozen items for the 1st week of November.

As more milk gets sold, it should also give us the idea of which product will be out of stock first. It will also help to apply FIFO (oldest stock gets sold first). That means the product that to be out of stock early can be sold first. We can make an early prediction that milk to out of stock next year's November first and store milk as quickly as possible.

b) Association rule

Association rule mining analysis is used to find patterns that suggests how strongly associated features in the dataset. Implication rules represent these patterns [4]. Finding the most useful role in and collecting interesting patterns to improve the organization of storing products is one of our main goals, and association rule will help us in that regard. The popular algorithms that use association rules include AIS, SETM, Apriori, and variations of the latter.

Normalized 1	Normalized 2	Normalized 3	Normalized 4	Normalized 5	Normalized 6
0.500	0.390	0.280	0.560	0.500	0.610
0.600	0.300	0.200	0.700	0.100	0.600
0.730	0.450	0.550	0.640	0.450	0.360
0.350	0.650	0.180	0.410	0.240	0.410
0.130	0.670	0.530	0.200	0.270	0.400
0.270	0.180	0.640	0.550	0.270	0.730
0.800	0.300	0.700	0.800	0.700	0.200
0.250	0.580	0.500	0.250	0.420	0.330
0.400	0.470	0.270	0.530	0.800	0.600
0.420	0.420	0.830	0.460	0.290	0.710
0.130	0.630	0.560	0.750	0.130	0.310
0.440	0.110	0.670	0.440	0.330	0.780
0.530	0.470	0.320	0.530	0.580	0.950

And after applying the FP-growth algorithm

Size	Support	Item 1	Item 2	Item 3
2	0.617	Normalized 6	Normalized 5	
2	0.610	Normalized 6	Normalized 4	
2	0.615	Normalized 6	Normalized 2	
2	0.608	Normalized 6	Normalized 3	
2	0.620	Normalized 5	Normalized 4	
2	0.621	Normalized 5	Normalized 2	
2	0.613	Normalized 5	Normalized 3	
2	0.612	Normalized 4	Normalized 2	
2	0.605	Normalized 4	Normalized 3	
2	0.609	Normalized 2	Normalized 3	
3	0.586	Normalized 6	Normalized 5	Normalized 4
3	0.588	Normalized 6	Normalized 5	Normalized 2

Here we can see a performance measurement unit called support. It tells us the frequency of different or individual items occurs together.

$$\text{Support} (\{X\} \rightarrow \{Y\}) = \frac{\text{Transactions Containing both } X \text{ and } Y}{\text{Total number of transactions}}$$

As can be seen, from support normalized products 2 and 3 are sold together at a 60 percent rate. That tells us to store normalized products 2 and 3

together to increase efficiency. Perhaps a highly-priced normalized 2 product can be stored beside normalized product 3 to ensure a maximum profit.

Now applying association rule algorithm, we get-

No.	Premises	Conclusion	Confidence	Lift
36	Normalized 5	Normalized 4, Normalized 2, Normalized 3	0.820	1.418
37	Normalized 5	Normalized 2, Normalized 3, Normalized 1	0.820	1.421
38	Normalized 4	Normalized 6, Normalized 2, Normalized 3	0.821	1.408
39	Normalized 4	Normalized 5, Normalized 2, Normalized 3	0.821	1.396
40	Normalized 2	Normalized 6, Normalized 4, Normalized 1	0.821	1.424
41	Normalized 2	Normalized 4, Normalized 3, Normalized 1	0.821	1.418
42	Normalized 2	Normalized 5, Normalized 4, Normalized 1	0.823	1.424
43	Normalized 2	Normalized 6, Normalized 3, Normalized 1	0.825	1.427
44	Normalized 6	Normalized 5, Normalized 4, Normalized 2	0.826	1.410
45	Normalized 6	Normalized 5, Normalized 2, Normalized 3	0.826	1.404
46	Normalized 2	Normalized 6, Normalized 5, Normalized 1	0.827	1.430
47	Normalized 5	Normalized 6, Normalized 4, Normalized 2	0.827	1.424
48	Normalized 5	Normalized 6, Normalized 2, Normalized 3	0.827	1.418

Confidence is the conditional probability of an event if given a set event has occurred.

$$\text{Confidence} (\{X\} \rightarrow \{Y\}) = \frac{\text{Transactions Containing both } X \text{ and } Y}{\text{Transactions Containing } X}$$

From this, if someone already bought products 6, 5 and 1, the conditional probability of someone buying product number 2 is .827, which is the highest from this group of data sets. As can be seen, product

number 2 should be stored close to 6 or 5 or 1 to increase efficiency and selling. Lift suggests the randomness of the given rule.

$$\text{Lift} (\{X\} \rightarrow \{Y\}) = \frac{(\text{Transactions Containing both } X \text{ and } Y) / (\text{Transactions Containing } X)}{\text{Fraction of transactions containing } Y}$$

A positive value which is more than 1 suggests how reliable the rule is. From the dataset, we can see that association rule number 46 is the most useful rule.

branch suggests an outcome of a test, and each leaf node holds a class label [5].

c) *Decision tree*

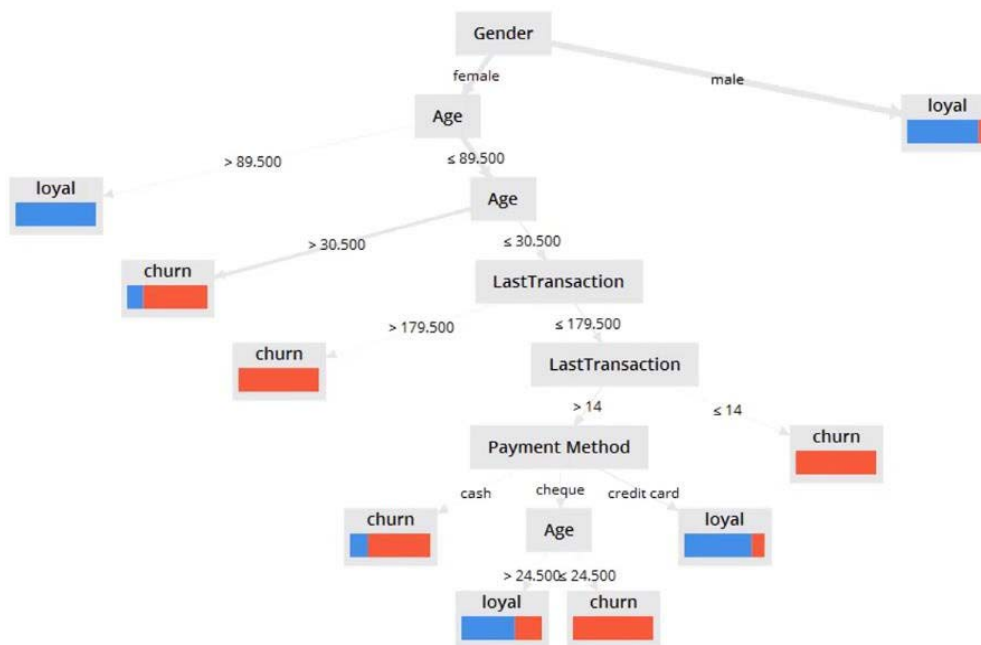
It is flow-chart like a tree structure, where each internal node denotes a test on an attribute, each

The data set we have used to apply decision tree algorithm is given below-

Row No.	Churn	Gender	Age	Payment Me...	LastTransa...
1	loyal	male	64	credit card	98
2	churn	male	35	cheque	118
3	loyal	female	25	credit card	107
4	loyal	male	39	credit card	90
5	churn	female	28	cheque	189
6	loyal	female	21	credit card	102
7	loyal	male	48	credit card	141
8	churn	female	70	credit card	153
9	loyal	male	36	credit card	46
10	loyal	male	22	credit card	51
11	loyal	male	27	cash	137
12	loyal	male	22	cash	147
13	churn	female	49	credit card	158
14	churn	female	24	cash	162
15	loyal	male	45	credit card	55

The data set shows us a few attributes, and to we have to detect which one is significant and needs priority. The decision tree algorithm we have used is known as chi-squared.

After applying the algorithm we get-



Here the algorithm came into the conclusion that gender is the root node. The decision tree tells us that, age group is the key element while storing for female customer. Age group of more than 89.5 are most likely to be loyal and under 89.5, we check churn and other attributes that tells us which one is in need of prioritization.

IV. CONCLUSION

As the retail industry gets ever so competitive, it is necessary for us to find every single opportunity to have the edge over everyone. Inventory management plays a major part of retail industry, and data mining techniques can be of use to store products efficiently with the future in mind. Customer insight is essential for any department even in storing products, and with these data mining techniques, valuable information can be extracted and used to our advantage. Our goal is to increase the attention in inventory management with the help of these techniques as it gets overlooked.

REFERENCES RÉFÉRENCES REFERENCIAS

1. Berry, M. J. A. and Linoff, G. Data mining techniques for marketing, sales and customer support, USA: John Wiley and Sons, 1997.
2. Fayyad, U. M; Piatetsky-Shapiro, G.; Smyth, P.; and Uthurusamy, R. 1996. Advances in Knowledge Discovery and Data Mining. Menlo Park, Calif.: AAAI Press.
3. Jiawei Han and Micheline Kamber (2006), Data Mining Concepts and Techniques, published by Morgan Kauffman, 2nd ed.
4. Association Rules in Data Mining: An Application on a Clothing and Accessory Specialty Store.
5. Efficient Classification of Data Using Decision Tree Bhaskar N. Patel, Satish G. Prajapati and Dr. Kamaljit I. Lakhtaria.
6. Data Mining for the category management in the retail market Jochen Garcke, Michael Griebel and Michael Thess.
7. Data Mining for Retail Inventory Management Pradip Kumar bala.
8. Growing Lifetime Value, Chain store Age(75), May 1999.

