MINING DATA TO REDUCE SHRINKAGE IN ORGANIZED RETAIL

[1]G.PAVITHRA, [1]R.S.SARANYA, [2]P.S.RAMESH

[1]Department of Information Technology, SASTRA University, Thanjavur, India
[1]Department of Information Technology, SASTRAUniversity, Thanjavur, India
[2] Senior Asst. Prof., Department of Information Technology, SASTRA University, Thanjavur, India

E-mail: pavi199226@gmail.com, rssaranya198@gmail.com, ramesh@cse.sastra.edu.
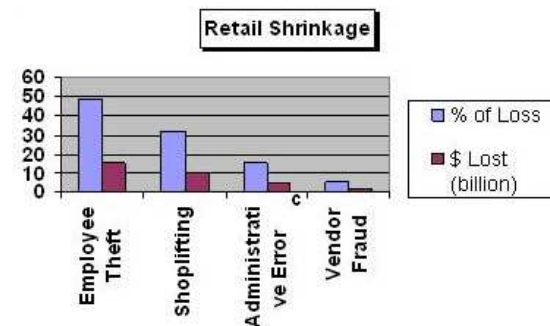
.

ABSTRACT

Shrinkage is a reduction in inventory due to shoplifting, employee theft, paper work errors and supplier fraud. If the shrinkage amount is large for a company obviously it will result in the net profit deterioration. In this paper we propose a model to reduce shrinkage in a retail industry using data mining. The barcode of a product is taken as the primary key. The access to this key should be with one trusted party and the remaining database can be accessed by anyone. Maintaining barcode for the product types such as stationery, groceries will solve this problem. This is done by using the data mining tool and presented in a pictorial representation. For instance, in the observed pattern if the quantity of the barcoded product exceeds the quantity of the product available, it can be identified as a shoplifted item. Using the tool the data pattern extracted from the dataset will contain the information of that particular item. Thus the theft can be identified and prevented. The reliability of our model lies in the trust of the administrator who can access the primary key.

Keywords: *Data mining tools, Primary key, Shrinkage, Clustering*

1.INTRODUCTION

Data mining is the process of discovering new patterns from large data sets. The goal of data mining is to extract knowledge from a data set in a human-understandable structure and involves database and data management. The future trends and behavior can be predicted by data mining tools, allowing businesses to make knowledge-driven and proactive decisions. Data mining offers a prospective analyses that moves beyond the analyses of past events that is provided by retrospective tools that are typical of decision support systems. The business queries that were too time consuming to solve were resolved using data mining tools. The hidden patterns that experts may miss can be found easily by the databases. Nowadays data mining techniques can be implemented more rapidly on existing software and hardware platforms to enhance the value of existing information resources, and can be coordinated with new systems and products as they are brought on-line.

Retail shrinkage is defined as the difference in the value of products as per the records and the actual amount of product available in the shop[1]. A survey made in the states of US among 118 retail chains says that the country has lost nearly $31 billion and shops lost 1.7percent of their total annual sales in a year due to shrinkage in retail. The reasons for retail shrinkage are primarily employee theft, shoplifting, administrative errors and vendor fraud.

Reasons for Shrinkage	% Loss
Employee Theft	48.5
Shoplifting	31.7
Administrative Error	15.3
Vendor Fraud	5.4
Total Shrinkage	$ 31 billion

Figure 1: Retail shrinkage in US in a year.

1.1 Employee Theft

Internal shrinkage is the other term used to represent an employee theft. Employee theft is caused due to the employees of the store. It means the act of stealing cash and products by the employees of the store. This is the reason for almost half of the shrinkage in retail industry. Shrinkage caused by employees occurs due to the wrong recording of transactions, forging bills, improper use of cash and computer and it is the reason for almost 61percent of the total employee caused shrinkage. 95percent of businesses experience employee theft and 75percent of theft is not identified. This problem is very difficult to identify because of the fact that the transaction is not visible. CCTV cameras, alarm systems, good security and an employee who can be trusted will solve this problem.

1.2 Shoplifting

Shoplifting is also termed as commercial burglary. It is one of the most common crimes. One by twelfth of every customer might me a shoplifter, a study says. Boosters is term used to represent professional shoplifters. Shoplifters need privacy. So the atmosphere of the store would have to be in a way that avoids this privacy and any blind spots. The expensive items and any other small items have to be kept in a locked case. The most successful way of preventing shoplifting is by using CCTV (Closed Circuit TeleVision) cameras. Alarm system and using a single door for a store can also help in avoiding this shoplifting.

1.3 Administrative Errors

Administrative are paperwork errors such as changing the prices due to carelessness cause around 15percent of the retail shrinkage.

1.4 Vendor Fraud

Vendor fraud occupies a very small percent of shrinkage. Vendor fraud mostly occurs when outside vendors stock the products within store.

2. EXISTING WORK

Database and data mining can be used in every departmental stores and supermarket for product entry and billing[2]. The daily work in a store will happen in the way given below:

1. The entry of the product into the store.

2. The details about all the products will be entered in the database.

3. The price of each item will be updated for the billing process.

4. The customer enters and take the basket and select the necessary product and put it in the basket and take it for billing.

5. The employee will check the products with the bar code detecting machine for a match with product-id and it will show the information and price for the products.

6. The bill will be calculated and total payment will be shown.

7. Customer will pay for the products bought.

8. All the products will be packed and given to the customer

This type of billing has been followed in all stores over the last many years. It has also improved in many ways according to requirements of sellers and customers. It does the same work of calculating the bill, giving it to the customer and maintaining proper database. They are accurate in calculation and printing, they also generate records. A new concept is also introduced in the stores is that they also maintain relationships with the customers who purchase more products from the store regularly. They are provided with customer cards and points are updated every time they buy items. Based on these points they are given gifts etc.The system also shows the overall profit and profit on a particular product and give reports about which items are required and which have cross their expiry date.

2.1 Bottlenecks Of The Existing System

The existing system has many advantages and disadvantages. So the bottlenecks of the existing system are as follows:

1. User Interface:

Full -fledged user interface is not available.

2. Graphical User Interface:

GUI is not good. Attractive user environment is missing.

3. Speed:

Processing speed of the software is not so much good to operate fast.

4. Flexible:

Existing system is not so much flexible that can be changed according to the operators and customers.

5. Generation of the Reports:

Not able to automatically generate the reports and documents.

6. Workload:

Sometimes the system hangs when a lot of database is used.

7. Error:

Sometimes the system makes error in calculating the bill and in the information of the products due to workload.

8. Staff requirement:

Existing model needs so many people to work on the database.

9. Resources:

System does not use the resources properly.

Dealing with large amount data it is difficult to frequently check for shrinkage in the database, so it can't be identified. Also theft or shrinkage can't be avoided or reduced.

3. PROPOSED MODEL

We try to identify the missing item by using data mining tools. Our model proposes that the barcode of an item should be made the primary key. Consider 4 databases each for Total_items, Available_items, Sold_items and Damaged_items[3][4]. Total_items database contains the entire details of all products in the store. Sold_items database contains the entire details of all the items that are sold. Damaged_items database contains the entire details of all the items that are damaged. Available_items database contains the entire details of all the items that are neither sold nor damaged[5].
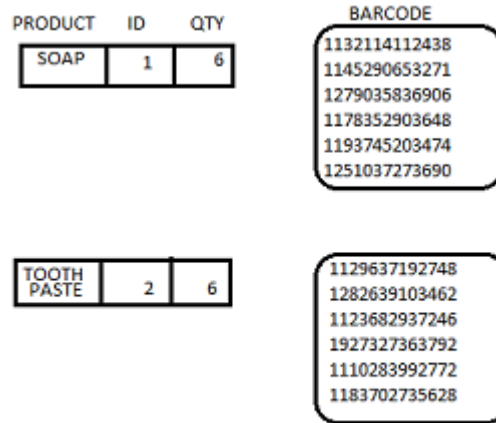

Figure 2: Example- Outline of the database.

Every day the new items are updated in the Total_items database[6]. Updation will be based upon the clustering algorithm. It groups the similar products into a cluster. The damaged items are updated in the Damaged_items database. As the day progresses a number of items will be sold. The details of these will be updated in the Sold_items database[7]. The items that are present in the Total_items database but not in the Sold_items and Damaged_items database are updated in the Available_items database.

Available_items=Total_items-(Sold_items+Damaged_items)

Finally at the end of the day we check on the present items and identify if any of them has been lost. This is done using any data mining tool. The quantity of all the items in the Available_items database is shown in a bar chart. The description of each item (Eg- Soaps, Shampoo) along with the

number of barcodes present is displayed in the bar chart[8]. This number is checked with the number of items of that description present in the store. If the number in the chart is found to be greater than that present in the store, we say that the item was shoplifted. The number of items shoplifted is equal to the difference between the number of items in the bar chart and that present in the store. Using this we can easily identify the lost item.

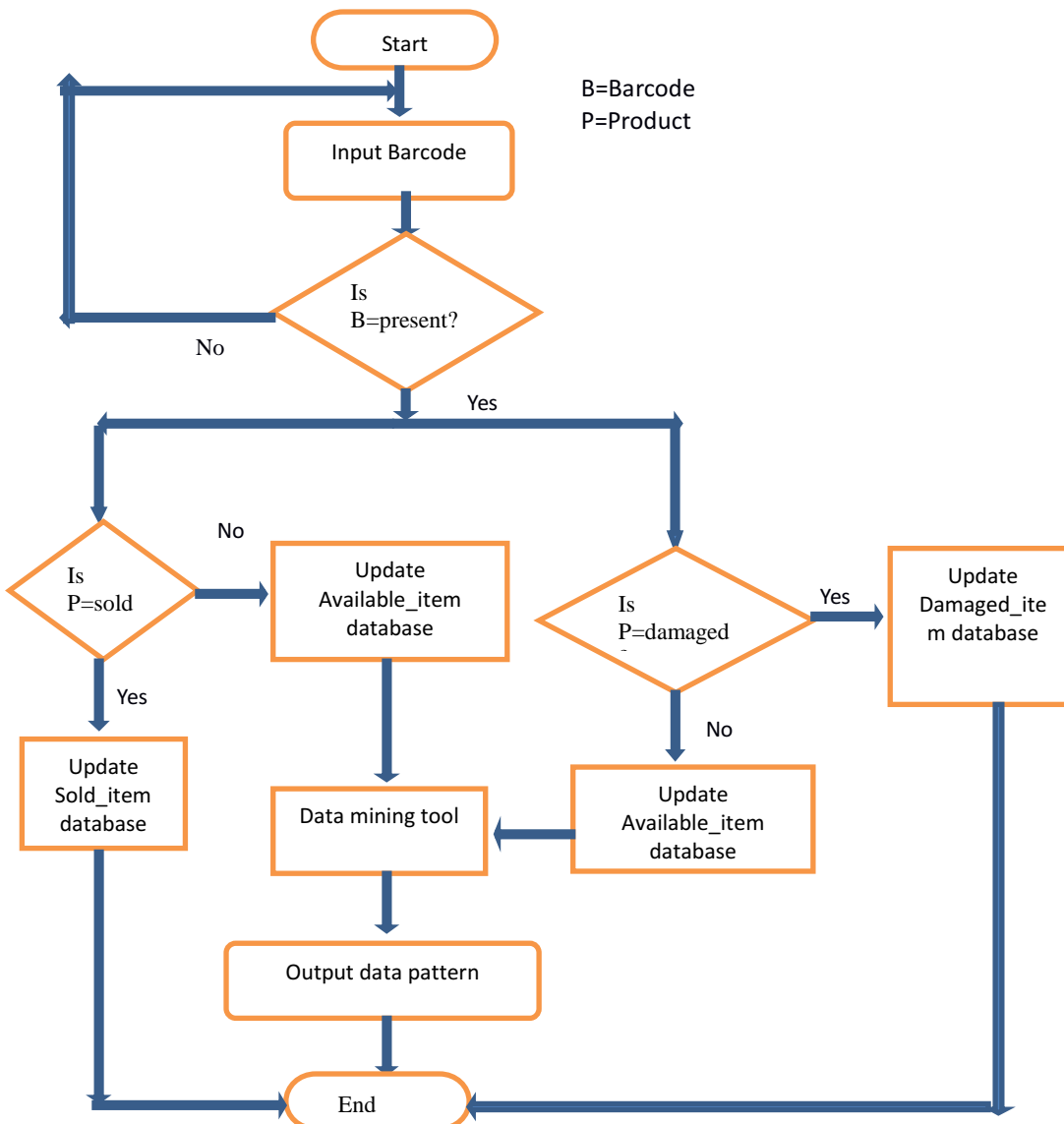*Figure 3: Barchart obtained on applying Data*

Figure 4: Flow chart- Execution procedure

There are chances where the employee may be the reason for the theft and he may change the values in the database. Because of this the barcode is taken as the primary key and only one trusted person has access to this unique key[9]. The remaining database can be accessed by all the staffs in the store. This depicts the efficiency of the model.

3.1 Data Flow

Figure 4 and Figure 5 depicts the sequence of activities taking place in our model.The execution starts with the reading of the barcode,if it exists the condition will pass to the next step. The barcode will be sorted out based on their status (ie) sold, damaged or available. If the test condition for either Sold or Damaged is satisfied their corresponding databases will be updated else the Available_item database will be updated. Then the Available_item database will be fed to the data mining tool which then displays the output in the form of a data pattern. The pattern thus obtained will be checked and if any discrepancies found ,that item will be identified as the shoplifted item. Then the process finally ends.
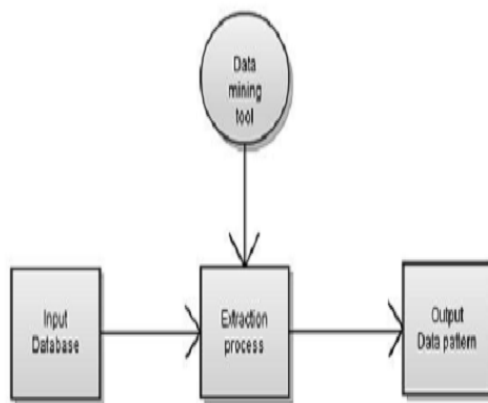

Figure 5: Block diagram

4. CONCLUSION

This model avoids shrinkage to a maximum extent. It's efficiency completely depends on the trusted party who has access to all the databases.

With the pictorial representation of the data obtained, it is easy to identify the discrepancies in values and it provides a better clarity for the user to view the database, thereby proving it's efficiency than other billing systems being used. Better results can be achieved with the proper maintenance of databases.

REFERENCES

[1] Data Mining for Network Intrusion Detection (2002). Paul Dokas, Levent Ertoz, Vipin Kumar, Aleksandar Lazarevic, Jaideep Srivastava, Pang-Ning Tan, Proc. NSF Workshop on Next Generation Data Mining, Baltimore, MD. pg 21-22

[2] Supermarket Billing System, Ankita Aggrawal, Neha Chatruvedi, Devendra Singh Sisodia(Students of MCA IV- Semester)

[3] State-of-the-art in Privacy Preserving Data Mining, V. S. Verykios, E. Bertino, I. N. Fovino, L. P. Provenza, Y. Saygin, Y. Theodoridis.. In SIGMOD Record, 33(1): 50-57. March 2004

[4] Statistical Inference and Data Mining, Clark Glymour, David Madigan, Daryl Pregibon, and Padhraic Smyth, Communications of the ACM, vol. 39, no. 11, pp. 35-41, November 1996.

[5] Integrating Data Mining with SQL Databases: OLE DB for Data Mining, Amir Netz, Surajit Chaudhuri, Usama Fayyad, and Jeff Bernhardt, in Proceedings of 17th International Conference on Data Engineering, IEEE Computer Society, 2001

[6] An Efficient, Scalable, Parallel Classifier for Data Mining (1996). Anurag Srivastava, Vineet Singh, Eui-Hong (Sam) Han and Vipin Kumar.

[7] Parallel Formulations of Decision-Tree Classification Algorithms (1999). Anurag Srivastava, Eui-Hong (Sam) Han, Vipin Kumar, and Vineet Singh, Data Mining and Knowledge Discovery: An International Journal, vol. 3, no. 3, pp 2-3, September 1999.

[8] Data Mining and Knowledge Discovery, Tian Zhang, Raghu Ramakrishnan, Miron Livny Volume 1,Issue2,1997,141-182.

[9] Toward Standardization in Privacy-Preserving Data Mining. S. R. M. Oliveira and O. R. Zaïane. In Proceedings of the 3rd. Workshop on Data Mining Standards (DM-SSP 2004), in conjunction with KDD 2004. Seattle, WA, USA. August, 2004.