SALES REPORT GENERATION SYSTEM WITH RECOMMENDATION TOOL

[1] **MihirSolanki,** [2]**Tejaswini Nanaware,** [3]**Deepali Bharmal,** [4]**Mrs.Neeti Desai**

1 Student, computer Department, RGIT, Mumbai

2 Student, computer Department, RGIT, Mumbai

3 Student, computer Department, RGIT, Mumbai

4 Professor, computer Department, RGIT, Mumbai

ISSN: 2278 – 1323

1. ABSTRACT:

SALES REPORT GENERATION ANALYSIS is a project which generates reports dynamically (for their employees as per hierarchy: CEO, regional head, sales person) depending upon the attributes (state, region, customer, product, quantity, amount, in-voice number, order number, monetary, margin etc.). *Reporting system* is defined as software for dynamic report generation purpose and is based on the data-mining techniques. Generating a report for a large volume of data and giving meaningful reports as an output is very difficult as it generates multi-dimensional reports. Here, the concept of OLAP cube is implemented as the solution to above mention problem.
Thus, raw data which is processed using OLAP cubes which use OLTP that gives the user the necessary reports. On the basis of the above attributes we can find out the top-K sales person and the top-K product for any firm. This will help a company or a firm to plot their success graph. Association based algorithms will be used on the basis of Conventional frequency and for monetary implementation. The project also considers the rare item i.e. the company can take the necessary measures if there is a decrease in sales in the report which is been generated. Moreover, the company can then make a decision regarding the sales of a particular product. Database can be uploaded to the site and the required multi-dimensional reports are generated. Thus, this project can be used for improving the sales of any given company.In this wewill be highlightingfew Data Mining methods and outline how some of the more recently developed products used in the business industry.

Keywords-Data mining: slicing and dicing techniques, OLAP (Online Analytical Processing)

2. INTRODUCTION:

In today's competitive world firms need high quality backend to support their growth and a technique which will help them to sustain in business world. In such process of evolution in technology data mining plays an important role by discovering strategies for accurate analysis of transactional or any other type of data sets.Yet tools for such analysis of data are very rare. Data scientists had to search through bunch of printoutsand piles of datasets to find the gems in data that make up scientific achievements. In this paper, data mining system focused on extracting knowledge from the firm database for decision making.

Data

Data is a quantity which is collected and categorized as different quantities, characters or different type of symbols on which different operations are performed by the computer.

Companies collect the larger amount of the data and store it in the data warehouse and the data types can be-

- Database for the students details can be viewed as operational data.
- There can be different type of non-operational datatoo.
- Meta data – Meta data can be viewed as a data about data.

ISSN: 2278 – 1323

Information
Understanding the data, patterns and following and having collective information about the type of data and how it can be utilized to study for easy implementation.

Knowledge

Information can be transformed to valuable data about historical patterns and future trends, this is knowledge. For example, we can analyse the customer buying behaviour and study the pattern in which the customer buys the product. Thus, a retailer could determine which items are more bought based on the customer buying behaviour.

Data Mining

Data mining is a combination of database management and artificial intelligence. The AI is emerging day by day in the world of database. Data mining is the core step in the field of analysing data. Unlike old methods of getting queries on standard relationships, data mining findsmore dynamic and meaningful relationships that would have not existed or the relationships that would give a more deep view of the data. Using trends that the end-user would have never thought to query the computer about. Without adding any more data, data mining gives a huge increase in the value added by the database Data mining allow many business users as it can be widely used to analyse the data and save many dollars for them. Data Mining is a concept that is growing in the business as well as commercial sector as a source of useful information out of terabytes of datasets.

3. LITERATURE SURVEY:

Data mining strategies can be grouped as follows:
• Classification: It includes categorization of data, also classification of the data as per users needs.
• Estimation: Estimation model determines a value for an unknown output attribute.
• Prediction:The predictive model predicts a future outcome by analysing the current inputs.
• Association rule mining: The vital hidden rules called ‘association rules’ in a large transactional data base are mined out. For e.g. the rule: {bread, butter->cheese}.
• Clustering: Cluster is a collection of similar data items. For e.g. given 100 data items randomly, they can be grouped as clothing, footwear, accessories.

The main application areas of data mining are in Sales analysis, Bioinformatics, Web data analysis, text analysis, social science problems, biometric data analysis and various other domains, where there is scope for hidden information retrieval.

The ethics of data mining

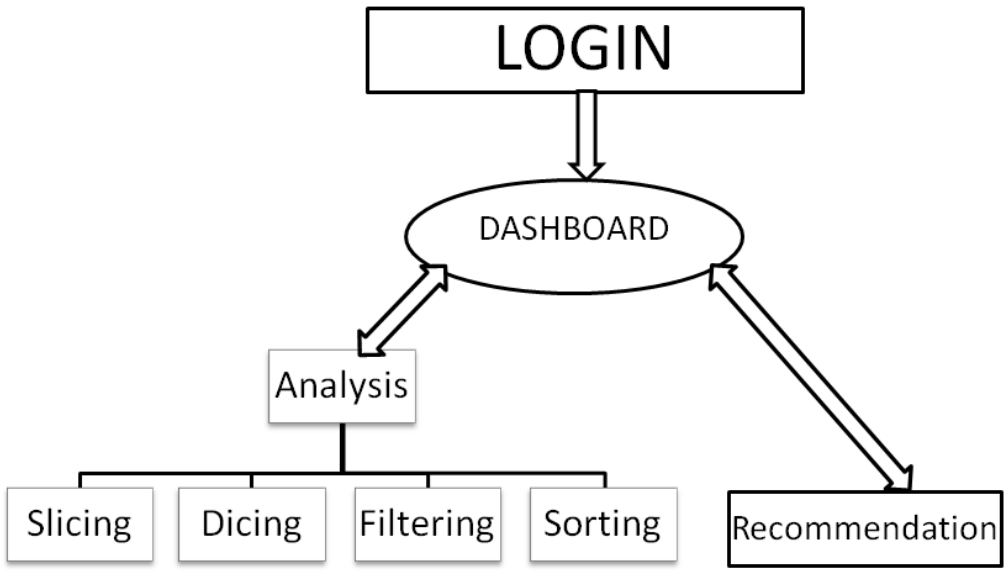
The use of data, particularly data about people, has serious ethical implications in data mining. Users of data mining techniques must act responsibly by making themselves aware of the ethical

ISSN: 2278 – 1323

issues that surround their particular application. It is people who take the results, along with other knowledge, and decide what action to be taken. Of course, anyone who uses advanced technologies should consider the wisdom of what they are doing. So the conclusion that can be drawn is like when data mining is conducted in a particular domain, one should seriously address the question whether the objective of mining is useful for the mankind and is there anything non-ethical hidden behind the rules extracted from the data mining process.

4. PROPOSED SYSTEM:

The project aims at developing a highly cost effective Multidimensional report for the database of a company. It researches models of computation in which actors (sales person, sales manager, regional head) operate on multidimensional arrays of data.



Data warehouses:

Is a system used for reporting and dataanalysis. DWs are central repositories of integrated data from one or more disparate sources. They store current and historical data and are used for creating analytical reports for knowledge workers throughout the enterprise. The data stored in the warehouse is uploaded from the operational systems (such as marketing, sales, etc., shown in the figure below). The data may pass through an operationaldatastore for additional operations before it is used in the DW for reporting.

- **Slicing:**

ISSN: 2278 – 1323

A slice in a multidimensional array is a column of data corresponding to a single value for one or more members of the dimension.Slicing is the act of divvying up the cube to extract this information for a given slice

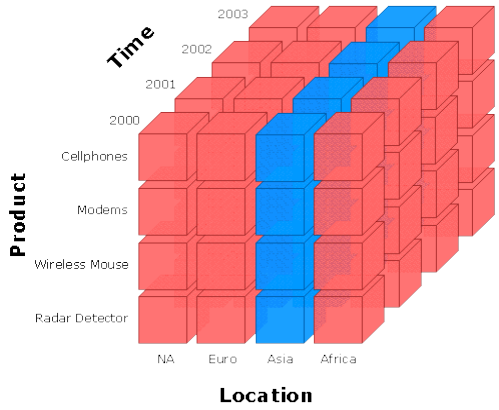


Fig3. Data Slicing

Dicing:

Dicing is more zoom feature that selects a subset over all the dimensions but for specific values of the dimension.

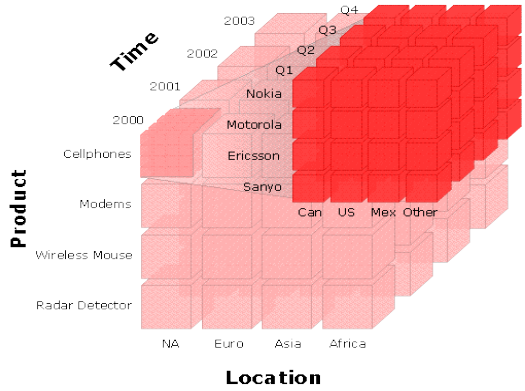
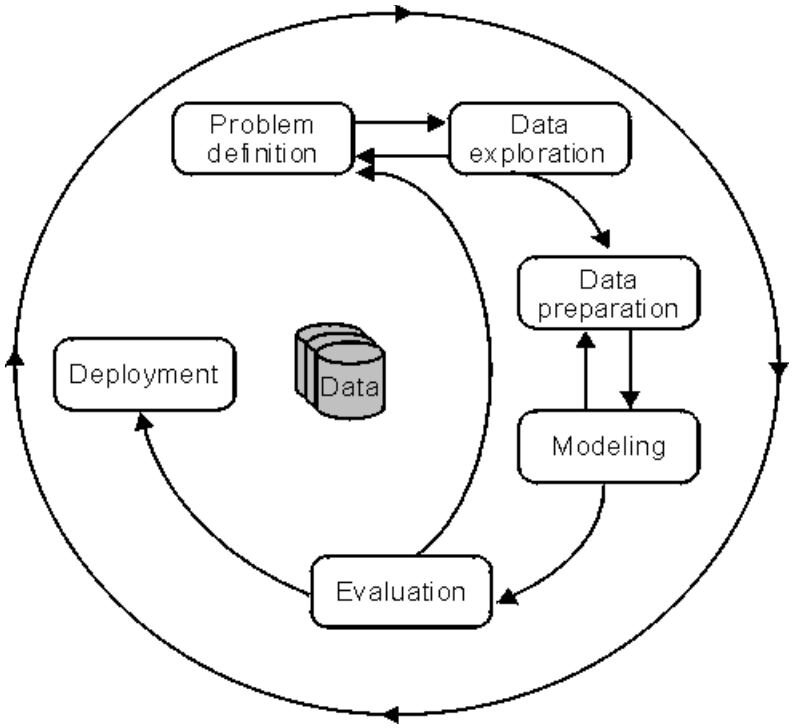


Fig4. Data Dicing

THE CYCLE OF DATA MINING

The life cycle of a data mining project consists of six phases. The shifting of phase to phase may differ. In this the phases can be moved in any direction.

ISSN: 2278 – 1323



The different phases focuses on understands the project and objective that is required for the business purpose.

- Data Understanding: It deals with understanding the data by capturing it from the source or the warehouse.

- Data Preparation: Unique data sets are collected and varieties are constructed using the raw data.

- Modelling: The actual modelling of how the mining is going to implemented is evaluated in this phase.

- Evaluation: Steps are been discussed and the ideas that are been modelled in the modelling phase are than implemented in the evaluation phase.

- Deployment: In this phase, the main objective of the deployment phase is to increase the knowledge of data so that it can provide more quick and easy access to the customers.

FEATURES OF THE PROPSED SYSTEM:

- Login
- Date selection
- Attribute selection
- Multi-dimensional report generation
- Filtering
- Sorting
- Graphical report generation
- Recommendation system

ISSN: 2278 – 1323

-

DATA ANALYSIS:

Sorting

A sorting algorithm is an algorithm that puts elements of a list in a certain order that ascending or descending order. This order gives row having maximum or minimum value with respect to particular attribute.

Attribute Selection

For generating report user can select attribute as basic parameter to compare all rows. For example, to see report according to employee name user has to select attribute as employee name so report will be arranged according name of employee as primary key.

Multidimensional Report Generation

User can view sales report with any attribute as primary key. This primary key attribute will be compared with rest to generate multi-dimensional sales report.
Multi-dimensional reports include two activities:

Filtering

Database will filter according attribute given, this is make view smaller. Because for larger firms total database will large hence report generated also will large.

Sorting

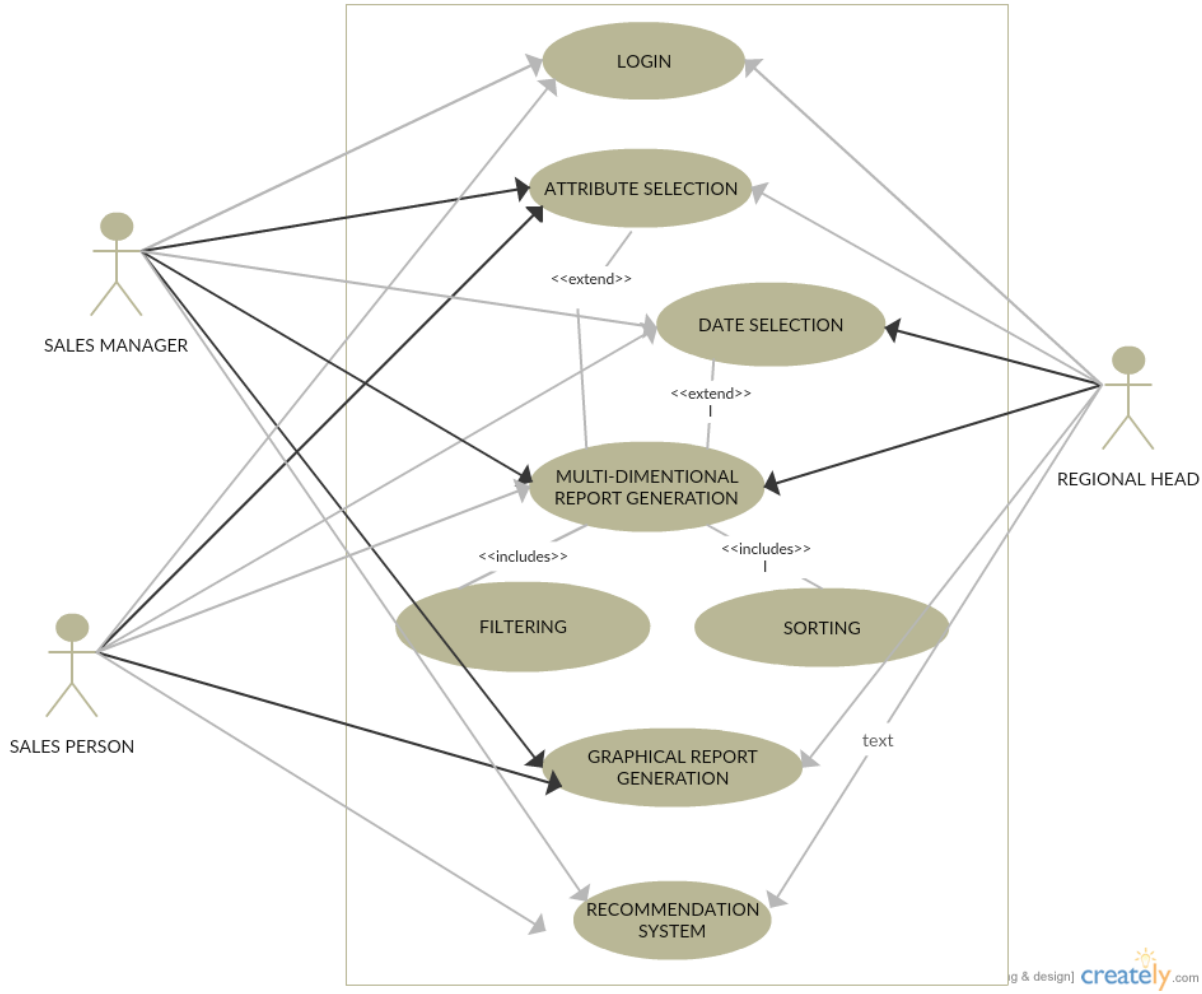
Database can be sorted in ascending or descending order. This makes comparison of different rows better. This also helps to find top-K product and top-K employee.

Top K Sales and Product

Top employee is decided on the basis of revenue generated by the employee. Employee who generates maximum revenue will be awarded with top employee. Similarly top product can be generated on the basis of maximum sales. This can be calculated using sorting algorithm.

Association rule mining is a procedure which is meant to find frequent patterns, correlations, associations, or causal structures from data sets found in various kinds of databases such as relational databases, transactional databases, and other forms of data repositories. Given a set of transactions, association rule mining aims to find the rules which enable us to predict the occurrence of a specific item based on the occurrences of the other items in the transaction. Collaborative recommender systems allow personalization for e-commerce by exploiting similarities and dissimilarities among customers' preferences. We investigate the use of association rule mining as an underlying technology for collaborative recommender systems. Association rules have been used with success in other domains. However, most currently existing association rule mining algorithms were designed with market basket analysis in mind.

ISSN: 2278 – 1323



5. DISCUSSION ON SYSTEM:

The system is implemented using Telerik tool. It is a data mining tool developed byTelerik Inc. Data mining techniques will be used for report generation. Different views of database will be created for generating reports depending on data slicing and dicing, this helps in minimizing the complexity of queries and will make our system more efficient.

Also, the system has a section for recommendation, which helps in maximizing the sales and the profits of the firm. Recommendation will be based on Cosine Similarity. Different items will be considered for recommendation and an n*n matrix will be generated. Column of the current item is compared with the columns of rest of the items and cosine similarity is calculated and stored in an array. The array is then sorted in descending order and the recommendation is displayed. The item with highest cosine similarity is recommended first (i.e. descending recommendation).

ISSN: 2278 – 1323

We are using asp .net, HTML, CSS, Telerik tool for implementing our project.

Algorithm for Recommendation:

INPUT

1. Initialize the items.
2. Store the data in an array.
3. Check for distinct click,
   if(distinct) ➔ increment the count
   else➔ keep the count same.
4. Maintain a symmetric matrix for recommendation
   i.ea [1] [0] = = b [0] [1].
5. Throw an exception whenever out of bond (i.e more than count)

   Then don't consider that items.

   Else ➔ return items.

6. Check for the item is valid.

   If yes ➔ return item[i]

   Else if NULL,

   Throw exception.

7. Initialize the rank array with the column of the index of current item.
   Rank [i] = data [i][index] ;
8. Take the sorted item except the current item.
9. If current item = valid.

   Return new list <Item count>( )

COSINE SIMILARITY

10. Store the current item index ➔ x.
11. Store the rank array ➔ a.
12. Get the associated value,
    Store count ←count.
13. If the same current index,
    Don't apply
    Else, b[] ← item rank.
14. Use cosine similarity(a,b).
15. Round off the result upto3 decimal place.

ISSN: 2278 – 1323

JACCARD INDEX

16. Store the current item index ➔ x.
17. Store the rank array ➔ a.
18. Get the associated value,
    Store count ←count.
19. If the same current index,
    Don't apply
    Else, b[] ← analyse item rank.
20. Use jaccard index similarity(a,b).
21. Round off the result upto3 decimal place.

OUTPUT

22. Display the order list in the descending order.
23. This provides the result upto 3 item sets.

6. CONCLUSION:

SALES REPORT GENERATION ANALYSIS provides a web-based solution to report generation. It helps in developing multi-dimensional reports dynamically and recommends the manufacturing of items. It generates flexible and versatile reports as per the users' needs. There is no restriction on date (from and to) selection. Also, yearly graphical reports will be generated.

A user can browse a spread sheet and upload it. The user can analyse the results from the report. It provides a user friendly interface for ease of operation.

7. FUTURE SCOPE:

System includes integration into computer-based, multi-function terminals for total automated support, cost/benefit analysis for managerial work, increased user participation in the design and revision of automated office systems, and increased attention to organizational and human factors in system design and implementation. Even with these methodologies, however, the implementation of sales report generation analysis has been difficult. Cost-benefit data is confusing, and overall productivity gains are difficult to be determined. If sales report generation analysis only provides faster generation of reports and easier access to files it will increase the problems inherent in current information flows and be rejected by the users. Security is a must. And apart from login ID and passwords, we are trying to find other methods for improving the security. A better customer-wise recommendation is essential. These improvements are a must and shall be improved by the iteration.

8. REFERENCES:

ISSN: 2278 – 1323

1. Agrawal, R., Imielinski, R., and Swami, A., "Mining association rules between sets of items in large databases", Proc. of ACM SIGMOD Conference on Management of Data, 207-216, Washington D.C., May 1993.
2. Chaudhuri, S. and Dayal, U., "An Overview of Data Warehousing and OLAP Technology", ACM SIGMOD Record 26(1), March 1997.
3. Dong, G., Han, J., Lam, J. Pei, J., and Wang, K., "Mining Multi-Dimensional Constrained Gradients in Data Cubes", Proc. of the 27th VLDB Conference, Roma, Italy, 2001.
4. Salton, G. and McGill, M. Introduction to Modern Information Retrieval, McGraw-Hill Book Company, New York 1983.
5. IBM (1995) Data Mining - An IBM Overview, IBM AlmadenResearchCentre.URL: http://www.almaden.ibm.com/stss/papers/overview.html
6. Simoudis, E. (1995) Reality check for data mining, IBM Almaden ResearchCentre.URL: http://www.almaden.ibm.com/stss/papers/reality/ UCLA Data Mining Laboratory (1996) UCLA Data Mining Laboratory Publications,UCLA.
7. URL: http://nugget.cs.ucla.edu:8001/publications/index.html
8. Data Mining Group at University of Helsinki (1996) Data MiningGroup at University of Helsinki, Department of Computer Science,
9. University of Helsinki. URL: http://www.cs.Helsinki.FI/research/pmdm/datamining/

ISSN: 2278 – 1323