IJITKM 0973-4414 Volume-9, Number-1 Jun-Dec 2015 pp. 68-73 **Impact Factor 2.8** available online at www.csjournalss.com

Analysis and Prediction Framework: Case Study in Fast Moving Consumer Goods

A. Aponso[1], K. Karunaratne[2], N. Madubashini[3], L. Gunathilaka[4] and I. Guruge[5]

SLIIT Computing, Sri Lanka Institute of Information Technology, Colombo 03, Sri Lanka.

[1]anoshaponso@gmail.com, [2]kaveeshakarunaratne@gmail.com, [3]madubashini1021@gmail.com, [4]josanthalg@gmail.com, [5]ivantha.g@gmail.com

Abstract*:* Big data analysis is a new trend in modern businesses. Analysing big data is not easy since data are tremendous. This system is knowledge based predicting system for any dataset in any size. Data mining provides the methodology and technology to change these bundles of data into useful human understandable information. As a solution, this system was introduced. Since knowledge based prediction system provide all information and knowledge related activities, it can be utilized to enhance the product manufacturing in any company. This leads manufacturers and managers to get more efficient decisions. This system was tested on real time data. A leading fast moving consumer goods company's database has been used as the case study. The historical data was used for the results based on the algorithms and perform it in an organized manner. Data set was analysed using Naïve Bayes, Linear Regression, Decision Tree, and Frequent Pattern Growth and ARIMA algorithms.

Keywords: Big data analysis, data mining, product manufacturing, fast moving consumer goods(FMCG), Naïve Bayes, Linear Regression, Decision Tree, Frequent Pattern Growth, ARIMA.

I. INTRODUCTION

Data Mining consists of finding interesting trends or patterns in large datasets to guide decisions about future activities. Data mining tools should be able to identify patterns in data using minimal user input. The patterns provide useful and unexpected insight to data analyst, which can be further, investigated subsequently using Decision Support tools. Manual data analysis has not archived its target, and methods for efficient computer-based analysis are indispensable. To fulfil this need, data mining was born. Decision support systems are defined as interactive computer-based systems intended to help decision makers utilize data and models in order to identify problems, solve problems and make decisions. They provide support for decision-making but not replace it [1].

This research is based on data mining. Fast moving consumer goods were taken as a case study. Nowadays fast moving consumer goods are essential items in the day today life. In fast moving consumer goods scenario, extracting useful information out from vast data volume causes greater difficulties. This company competes in the personal care products and accessories market, catering mainly to new-born's up to senior citizens. The project team think the output of the forecast results will be used to set promotional plans among regions and seasons, and measure amount of fast moving consumer goods to import in 2016 up wards. Potential business benefits include reducing costs (inventory costs and advertisement at lower sales seasons), and improving marketing strategies (targeting the right seasons). Managers, exporters, salespersons and manufactures face the continuous challenge of not being able to predict the products that would be market and demand well in the upcoming selling and also in which time period of the year they can earn more financial benefit by these products. Being able to do so will give them a definite financial advantage of requesting faithful manufactures to produce their products for the demand and receiving a good financial benefit that has happened. In order to conduct a Data Mining experiment, the project team recommends considering the following questions to provide a direction to the experiment with less guesswork. These are the research questions the project team discussed.

What is the most suitable algorithm to be used in mining data sets of FMCG?

As this is an experiment, a set of steps for a successful experiment needs to be developed. At this stage, project team needs to select the appropriate Data Mining techniques and apply them in a certain order and develop a methodology/algorithm to gain knowledge from the data selected for the experiment. As mentioned above, Data Mining is a broad term applied to several kinds of exploratory analysis. There are many different Data Mining techniques and algorithms such as Frequent Pattern Growth, Linear Regression, Naïve Bayes rule and Time Series algorithms.

What knowledge can be gained by data mining to FMCG?

This provides a direction and a goal for the Data Mining experiment without a proper goal; it is difficult to estimate effort, time and proper methodology for an experiment.

How can FMCG business be benefited by data mining?

Data mining exposes hidden patterns among vast volume of data. What makes data mining special is that it can be used to predict future behaviour by looking at past events. Data mining professionals analyse the relationships among relevant

patterns in data sometimes seemingly unrelated data to predict future trends. Data mining is just what it sounds like mining for data from any number of databases or other sources to extract key information. Ultimately, data mining should be the end result for every business with a database management system.

Main objective of the research is to identify hidden buying patterns that would benefit forecasting for the consumer products that needs to be produced in the upcoming seasons. These findings can lead the industry to further investigate the varieties of new arrivals that should be produced in the long run. Investigate and review existing literature related to the subject area and research on data mining tools and techniques appropriate for the study is also an objective. Perform selected data mining techniques to analyse many properties and their combinations data that has been gathered over the past years and evaluate research findings with the aid of senior contributors of the fast moving consumer goods sector to influence industry in utilizing the research outcomes are also the objectives. Not only that but also enhance the sales within time critical, valuable or small value scenarios and giving some promotions in the period of lack of sales and analysing the previous year's performance predict the key points of the sales improvement are the key objectives.

This research paper describes a framework that can be used for any dataset that discover knowledge for data/pattern analysis in fast moving consumer goods. This is a data mining based research, so this includes tools and techniques used for mine the data, comparison between other data mining projects, system assume to build, procedure and methodology for fill the relevant gap in this research.

II. BACKGROUND

There are many researches done in the field of data mining.

G. Vossen evaluates the problems, methods, and applications of big data, with an importance on future Business Intelligence architectures. The enormous business impacts and the effects that it will occur on advertising, commerce, and Business Intelligence on big data has already foresee from many companies already [2].

N. Rathee and S. Choudhary gave an introduction of Data Mining System and development of different data mining systems. When a company is working in different parts of country and producing large amount of data then it very important to analyse that data which account for productivity and cost benefits from that data. It is important to understand that which part of the organization is giving benefit and where the organization is losing. According to them data mining is an interdisciplinary subfield of computer science which involves computational process of large data sets' patterns discovery. They had used the juncture of artificial intelligence, machine learning, statistics, database systems and business intelligence [3].

L. LI et.al proposes a data mining based food risk assessment and early warning method, set up a scientific and reasonable system of risk evaluation, risk evaluation results are obtained, and then through the evaluation result is multilayer multidimensional association rules mining, food and testing project early warning information, provides powerful guarantee for food safety [4].

S. Sojan et.al says that application or an information system helps to take business decision more effectively. A company uses Customer Relationship Management as a strategy to handle customer interactions. For Successful CRM approach there are various data mining algorithms. The hike in the participation of today's business world has caused the data input to increase at a tremendous rate thus making data mining a very important part of industries. An interest of various companies has been attracted more and more due to the implementation of continuous Data mining processes [5].

In order to achieve goals of every organization and business, we need to fully exploit the data by extracting all the useful information from it. But it is not practical to manually analyse, explore, and understand the data because of it's the size and complexity. . Decision can take place on the basis classification of Dead-Stock, Slow-Moving and Fast-Moving of the sale. Segment by segment sales forecasting can produce very useful information. The information produced is very useful for business decision making. However it is very useful in understanding market trends. Sale data classification has different market trends such as when some clusters or segments of sale may be growing, while others are declining [5].

Metzger et.al analysed and compared predictive monitoring techniques with respect to five accuracy indicators based on observed evidence in the area of transport and logistics from an industrial case study. Even though numerous predictive monitoring procedures have been proposed in the before those techniques have been only assessed individually from each other, by making it hard to reliably compare their accuracy and applicability. [6].

H. Chen et.al provides a framework which identifies the development, applications, and emerging research in the areas of BI&A. Chen et.al have present an important area of Business intelligence and analytics (BI&A) for both practitioners and researchers. This shimmer the degree and effect of data-related problems in contemporary business organizations which are to be solved [7].

III. METHODOLOGY

Data mining denotes a rather difficult and specific field. A complicated and fixed approach is necessary for the use of data mining in order to help organizations use the data mining. For this project, Prototypes are expected and there will be very critical risk assessment when going through the project because the team does not have much experience within the domain. The spiral model emphasizes risk analysis. Remove all potential risks through careful analysis and, if necessary, by constructing a prototype.

Before data is being mined several steps have to follow.

1. Data integration:
   First collected and integrated all the data from different sources.
2. Data selection
   Then selected the data that can be used for data mining.
3. Data cleaning
   Most of the time, data gathered are not clean and may contain errors, missing values, noisy or inconsistent data. So they have to remove. This process was done manually.
4. Data transform
   This step is done for get more efficient results and to gain results in understandable manner. Mainly normalization was done for the data.
5. Data mining
   In this step the algorithms were used for mine the database. Data set was analysed using Naïve Bayes, Linear Regression, Decision Tree, and Frequent Pattern Growth and ARIMA algorithms.
6. Pattern Evaluation and Knowledge Presentation
   First some of the results that are not related were removed fine tune other results well to understand by the users of the system.

- View item sets which customer prefers to buy most.
  Frequent pattern Growth algorithm was used to implement this feature which finds frequent item set. It uses a divide-and-conquer strategy. The algorithm was coded manually. In this, there is report generated using most popular item set by the customers. This is useful to give promotion hamper packs to customer.

- Forecast monthly sales and customer buying items, in order to plan manufacture.
  This feature was created using ARIMA algorithm. The results are getting through a mining model in Microsoft Business Intelligence using DMX queries. In this also like in analysis, there are two parts. This gives a prediction of sales in future year.

- Analyse the previous months sales for identify important facts for marketing decisions.
  This feature was created using Microsoft Linear Regression algorithm which helps to calculate a linear relationship between a dependent and independent variable. Mainly in the system this is divided into two parts as Product Sales and Customer Product Sales. The algorithm was coded in C#. In Product Sales user can select Product Name and in Customer Product Sales, Customer Name and Product Name. Basically this gives an idea about past sales in chart format.

- Suggest packs to market uninterested items.
  This is the feature that generates a list of which is combination of lowest interests and highest interest item. This is for identify items to give promotions.

- Import any fast moving consumer goods dataset
  This system can be used to any dataset to do all the above features. So importing that database also a needed feature of the system.

- Predict Seasonal promotion items in order to be prepared for upcoming seasons
  This is dividing into two parts as in Sri Lanka mainly two seasons which the product sales go high which are New Year and Christmas. User can select the season and it displays a list of items which sold high in that season.

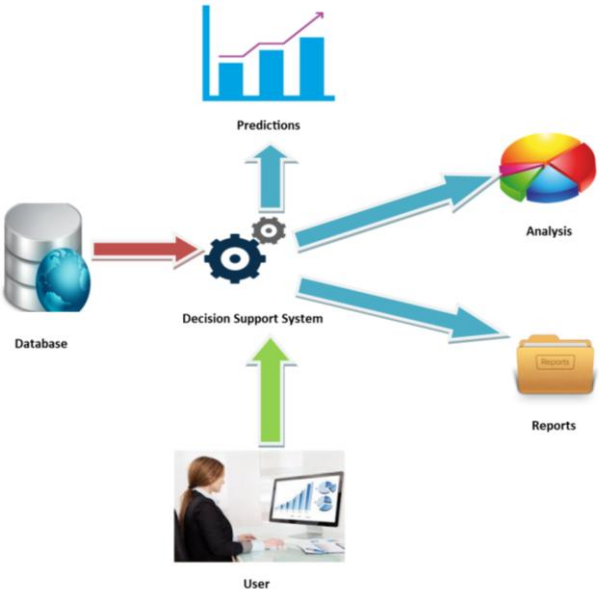


Figure 1-High Level diagram

IV. Results

First the project team have transformed data in required format by removing noise and any other inconsistencies.
The predictions are arrived at from two different forecasting models where all the results from these models will be refined. Since all the predictions should be long term predictions ARIMA algorithm was used because the ARIMA algorithm is optimized for long-term prediction. Mainly the predictions are divided into two parts.

- Product Monthly Sales Forecast
- Customer Buying Item Forecast

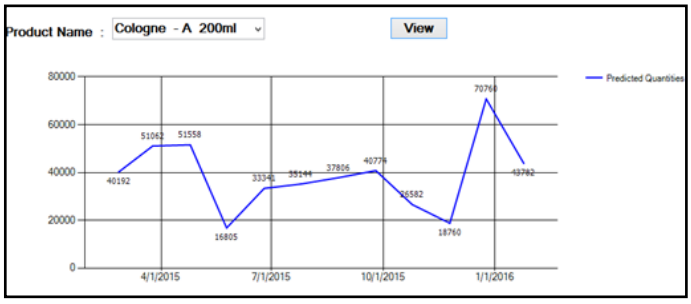


Figure 2-Sample Monthly Sales Forecast for a product

Figure 2 shows the predicting quantity of the product for the next year. It means the managers and manufacturers can get a clear idea about their upcoming sales and they can get ready to supply product.

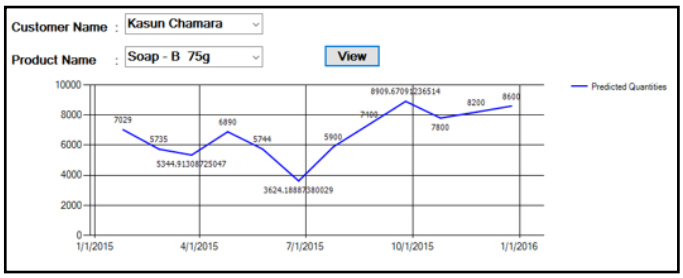


Figure 4-Sample Customer Buying Item Forecast

Figure 3 shows the predicting quantity of the product that can buy by the customer for the next year. It helps managers to give promotions to particular customers.

Listing 1 shows a sample code that used for predict customer buying item.

Error! Reference source not found. gives an analysis of past years monthly sales. Through this managers can get an idea about their up and downs in sales. Linear Regression was used to perform results.

```
SELECT FLATTENED

(PredictTimeSeries([Forecasting_Customer
s].
 [Quantity],12,EXTEND_MODEL_CASES)) as
 [Predicted Values]
From [Forecasting_Customers]
 PREDICTION JOIN
  OPENQUERY([Cheramy Care],
   'SELECT [Quantity],[date],
     [Customer_Index]
    FROM

[dbo].[View_SalesCustomerDetailsForecast
]
      ') AS temp
   ON
    [Forecasting_Customers].[Quantity] =
      temp.[Quantity] AND
    [Forecasting_Customers].[Date] =
      temp.[date] AND

[Forecasting_Customers].[Customer
_Index] = temp.[Customer_Index]
```

Listing 1-Sample DMX query of predict customer buying products

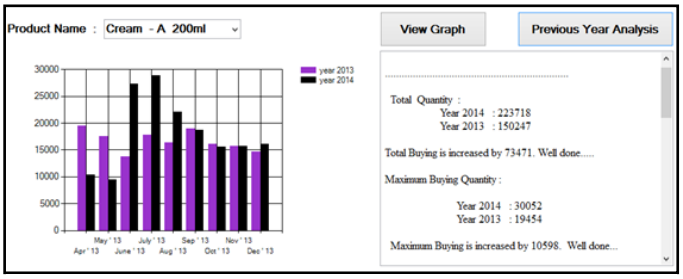
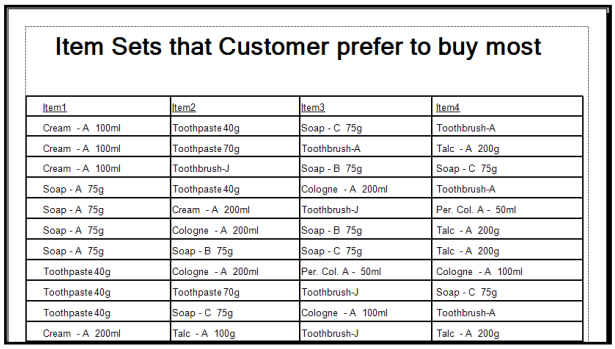


Figure 3-Past Years Product Sales Analysis

Figure 5-Past Years Customer Buying Item Sales Analysis

Figure 5 give the blue print of the report for a relevant customer for as unique product. Not only that but also it gives the information about is the total buying and maximum quantity is increased or not with a filtered details.



Item Sets that Customer prefer to buy most

Item1	Item2	Item3	Item4
Cream - A 100ml	Toothpaste 40g	Soap - C 75g	Toothbrush-A
Cream - A 100ml	Toothpaste 70g	Toothbrush-A	Talc - A 200g
Cream - A 100ml	Toothbrush-J	Soap - B 75g	Soap - C 75g
Soap - A 75g	Toothpaste 40g	Cologne - A 200ml	Toothbrush-A
Soap - A 75g	Cream - A 200ml	Toothbrush-J	Per. Col. A - 50ml
Soap - A 75g	Cologne - A 200ml	Soap - B 75g	Talc - A 200g
Soap - A 75g	Soap - B 75g	Soap - C 75g	Talc - A 200g
Toothpaste 40g	Cologne - A 200ml	Per. Col. A - 50ml	Cologne - A 100ml
Toothpaste 40g	Toothpaste 70g	Toothbrush-J	Soap - C 75g
Toothpaste 40g	Soap - C 75g	Cologne - A 100ml	Toothbrush-A
Cream - A 200ml	Talc - A 100g	Toothbrush-J	Talc - A 200g

Figure 6-Sample Item Sets

From this analysis, it identifies interesting buying patterns of customers for past two years. In the course of this managers and manufactures can get an idea about their sales and promotion packs they can introduce to the market. Frequent Pattern Growth was used to perform the results. This indicates the frequently moving items in a descending manner. From a relevant set display how many customers buys that product and the percentage.

This report shows the frequent items sets that have been bought together by the customers obtained by using FP-growth, which is an efficient data mining of frequent pattern in large databases. To design sales promotion packs according to buying behavior of customers and to improve customer satisfaction market basket analysis is an important algorithm used in retail organizations.

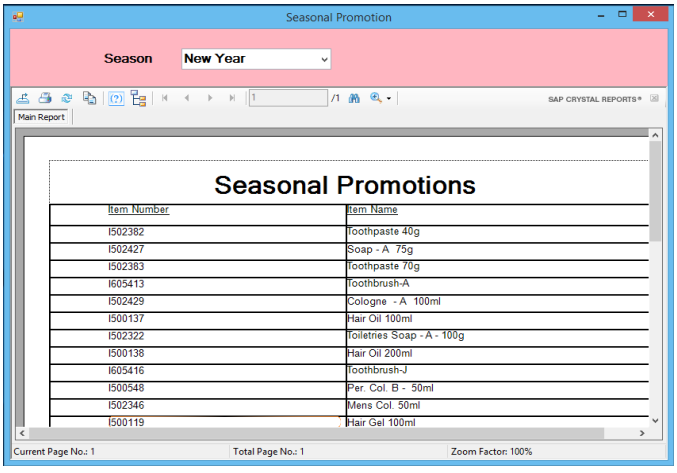


Figure 7-Seasonal Promotion Report

This interface shows the items that have been mostly sold in the selected seasons. User can select the Season ("New Year", "Christmas"). . Seasons are an opportunity to attract attention and bring in new business. Running promotions at times when there is a lot of demand is an obvious option, but inevitably there is a lot of competition too. The key with these is to plan ahead with this unique information.

V. Discussions

It has become clear that business process renovating and deploying automated solutions is not an easy task. Nowadays most business users have understood of the need to deploy data mining into direct marketing process.

Since efficient decision making and accurate predicting offers became a vital need to fast moving consumer goods industry. It can lead to higher risk. This is system focused on giving directly better making to the managers and to the manufacturers.

Processing a large number of record set for previous years from excel sheet at once was challenging to be performed on a local computer.

An organization may consist of data distributed and in a heterogeneous manner. Identifying and selecting necessary data to mine is crucial to ensure that the experiment meets its goals. An expert with experience and broad knowledge of the application domain (i.e. domain expert) can assist in the guiding and selecting the appropriate data items. This is a critical step in the experiment. The next step is to ensure that the data is selected and cleaned are in an appropriate format for mining. There are a plethora of tools available for extracting and cleaning data. Sometimes considerable effort is required at this stage as well.

Prepared the prediction for 2015 by using 2013 and 2014 figures and this predicted figure is compared with the actual figure up to 2015 August. The accuracy of the compared figure is nearly 58%.

VI. Conclusion

This system is knowledge based predicting system for any dataset in any size. This system was tested on real time data. A leading fast moving consumer goods company's database has been used as the case study.

The project team built quality product which any person can understand and operate the analysis. Most of the data mining software's provide short predictable data only. This product provides reasonable and explainable report to the customer who uses this product.

Finding of this research is a stepping stone in Knowledge Discovery in fast moving consumer goods by identifying seasonal patterns, which customer prefers to buy most, analysing the previous months sales for identify important facts for marketing decisions, suggest packs to market

uninterested items, forecast monthly sales and customer buying items, in order to plan manufacture, predict Seasonal promotion items in order to be prepared for upcoming seasons, import any dataset. Since this dataset was untouched for the sort of knowledge discovering this research was focusing, even the simplest findings from the basic analysis to data mining was validated against expert judgments.

The huge advantage of this system can be gained by the managers, manufacturers such as achieve more profits, high markets shares etc.

The project team also intend to do modifications in the data mining deployment methodology when there will be any indications requiring it.

In future work the project team are planning to develop a web site which will be compatible with any fast moving consumer goods company database and specially improve this system that can used for any other data set.

In concluding the findings of this research, identification of monthly similarities of the dataset is a major discovery of the research and satisfies the key research goal. Secondary research goals were also met by successfully extracting domain knowledge from various sources and performing data mining on the dataset.

References

[1] R. P. Chakraborty, "Integration of Data Mining Systems using Sequence Process," *International Research Journal of Engineering and Technology (IRJET)* , vol. 02, no. 01, pp. 204-211, Apr. 2015.

[2] G. Vossen, "Big data as the new enabler in business and other intelligence," *Vietnam J Comput Sci*, vol. I, pp. 3-14, 2013.

[3] N. Rathee and S. Choudhary, "Study of different Data Mining system & Platform," *International Research Journal of Engineering and Technology*, Vols. II, no. 1, Apr. 2015.

[4] X. Wu, X. Zhu, G. Wu and W. Ding, "Data mining with big data," *Knowledge and Data Engineering, IEEE Transactions*, Vols. 26, no. 1, pp. 97 - 107, 2015.

[5] A. Metzger, P. Leitner, D. Ivanovic, E. Schmieders, R. Franklin, M. Carro, S. Dustdar and K. Pohl, "Comparing and Combining Predictive Business Process Monitoring Techniques," *IEEE Transactions on Systems*, Vols. 45, no.2, pp. 276-290, 2015.

[6] O. El-Gayar and P. Timsina, "Opportunities for Business Intelligence and Big Data Analytics in Evidence Based Medicine," in *IEEE*, Waikoloa, HI, Jan. 2014.

[7] H. Chen, R.H.L. Chiang and V.C. Storey, "Business Inelligence and analytics: From big data to big impact," MIS Quarterly , Vol. 36 No. 4, pp. 1165-1188, Dec. 2012.