

A STUDY AND IMPLEMENTATION ON DATA MINING TECHNOLOGY AND APPLICATIONS FOR ANALYSIS ON SALES PRODUCTIVITY

¹Johar, MGM, Mohd Shukri Ab Yajid, Ali Khatibi

Abstract---This project focuses on the study of Data Mining technology and implementation of product sales predictive analysis for Amway (M). Certain parts have been adopted to enumerate the grounds of study, where at this stage mainly books, journals and articles on the Internet about Data Mining, as well as interviews and observations have been performed to gather the data of Amway (M). After the collection of data, model was developed to interpret the significance of results performed on the basis of such data by integrating automation, visual representation and predictive analysis into business processes, companies can enhance the efficiency and accuracy of decision making by appropriate methodology. Lastly, regarding this project the author has learned a lot upon the completion of it. It has been, thus, a very satisfactory and fruitful journey of completing this project. Generally, this whole project needs to be polished to have a better effect.

Keywords---Technology, Sales, Amway, Products, Business Process

I. Introduction

While coming in from a historical background of technology, it is understood that current information technology has been upgraded far better than in past and as well it's the sector which is achieving most advancements in all over the world (De Silva et al., 2018a; De Silva et al., 2018b; Nikhashemi et al., 2013). Where the technology has made much easier things to do, it further has played vital role in the management of computer sciences where the data collection and arrangement procedure is much easier than in past (Bose, 2009). However, to explore these patterns and thus to use more of the information resides in the available data turns out to difficulty. To revamp the data into useful knowledge the tools are less quantity as compared to plentitude of data. People are inquisitive towards information and knowledge seeking. As a result a new area of research named Data Mining is emerged and which use the challenge to evolve techniques that help humans to discover convenient patterns in their data (Rygielski, Wang, & Yen, 2002).

Amway (M) establish a small office and warehouse facility in jalan ipoh in 1976 with only five employees with just five employees and At that time Amway (M) was one of the founder of direct selling industry. Amway (M) distribute a wide dimension of superb trait consumer commodity, ranging from daily goods, health care, home care, home technology, auto care, personal care products, to fashion (Tsai, 2012). The distribution is made through an extensive network, consisting of the members and sales centers throughout Malaysia. Today, Amway (M) has become the well-known direct selling company with 161,000 core distributor force and sales turnover of RM456 million, forging Amway a household name all over the

¹Management and Science University
mdgapar@msu.edu.my

Malaysia. Therefore, mining their data is one of the crucial steps in order to meet their challenges toward the future (Cabena et al., 1999; Dewi et al., 2019; Pambreni et al., 2019; Tarofder et al., 2017).

Questions concerning the knowledge representation had long been of the interests to many of the cognitive psychologists. They had been conducting lots of experiments on this issue. Many theoretical distinctions of different kinds of knowledge had come up. Procedural versus Declarative knowledge is the distinction that is particular relevance to data mining. Procedural knowledge is particular involving the “knowing how” to do things, while in the meantime, Declarative knowledge deals with “knowing that”, of the factual information of the real world. Procedural knowledge might be referred to as a cookbook, or recipe for cooking a meal. It describes the process, or how to get certain activity done, or in other words, it shows the process of getting a thing done in step by step form. Declarative knowledge is usually used as a representation of the factual information, such as the name of all the products distributed by Amway (M) or lists of all of the active customers and etc. (Linoff & Berry, 2011; Doa et al., 2019; Maghfuriyah et al., 2019; Nguyen et al., 2019).

Distinguish Between Meta Knowledge and Actual Knowledge

It is important to have a clearer view of the types of information that are to be extracted from the data sets. Looking or describing either the known quantities or just looking up for interesting relationships existing in the classes of the knowledge that have already been known, are what matters here.

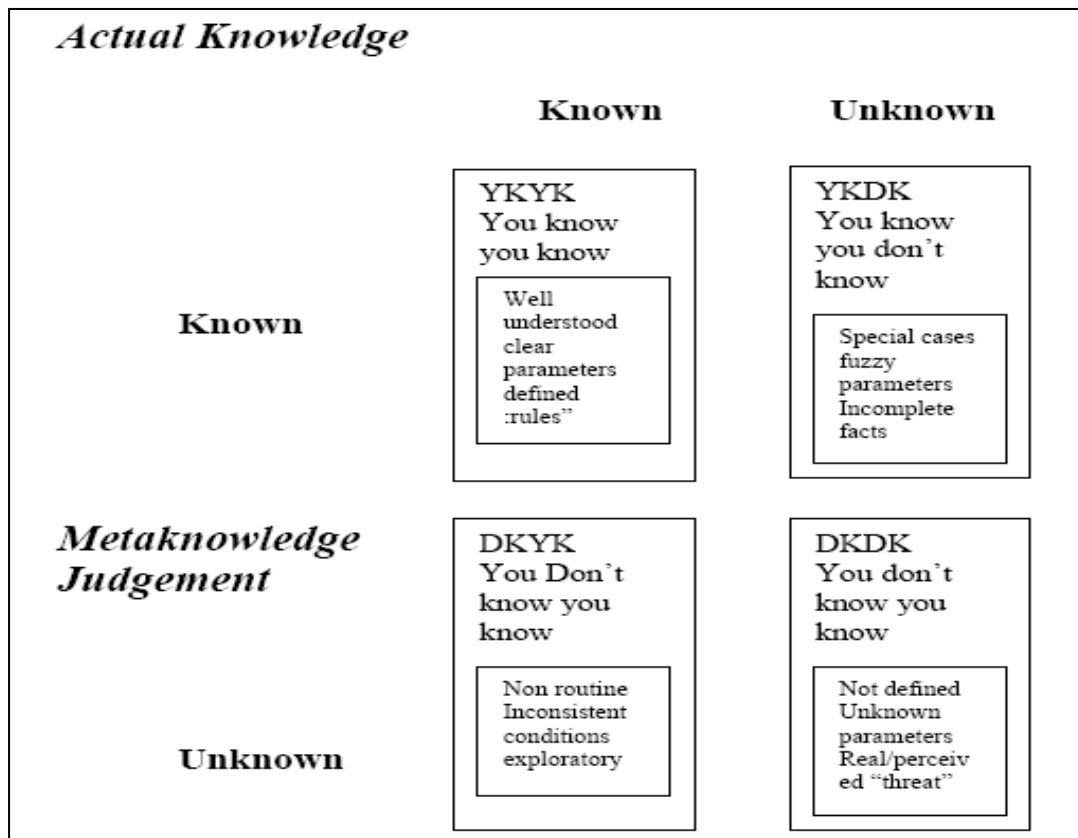


Figure1: Dimensions of Met knowledge and Actual Knowledge within the problem

In this figure, it can be seen that the Meta knowledge and the actual knowledge exist along orthogonal dimensions.

Information that You Know You Know (YKYK)

The information that you know you know (YKYK) is the simplest case in this matrix. There is actually information that exists in the data set itself, and it is clear to the users that they are aware of it. For example, it is clear and obvious to us that water boils at 100c, and froze at 0c. People know that the earth is round, know that student cannot be late for school; know that airplanes fly from one country to another and etc.

Information that You Know You Do Not Know (YKDK)

There are many things that can be characterized in this category. In daily lives, people know that they do not know a lot of things. For instance, they know that they do not know how many people die every single day; they know they do not know how many new born every day; they know they do not know the angle of the Earth's rotation on its axis, on its orbit around the sun; and etc.

Information that You Do Not Know You Know (YDYK)

The types of information that you do not know you know (YDYK) is targeted in exploratory analyses in which there might not be having definite answers about what it is expected to find. There are always the possibilities that new discoveries of interesting information are already contained within a data set though it might not currently being used. In business context, one may not be suspecting that there are some predictable patterns associated with a particular set of investment behaviors, though this may be achievable through the analyses of the existing data.

Information that You Do Not Know You Do Not Know (DKDK)

This is the last of the four cells. There are some important gaps in the data sources of this case. Not having the Meta knowledge of them might lead the users being unaware of the gaps. This is the most vulnerable situation to the system. It in fact, affects all governments as well as corporations throughout the world (Nemati & Barko, 2003; Pathiratne et al., 2018; Rachmawati et al., 2019; Seneviratne et al., 2019; Sudari et al., 2019; Tarofder et al., 2019).

Well, most of the time, the clients are not really sure what they really want or they often do not have specific minds or ideas about what the beginning of the project is all about. They are just seen as an exploratory exercise, and are just hoping to get some interesting results. If it is in this kind of situation, there is still work to do on the initial investigation. In People's Caring Products, the category can further be narrowed into Personal Care, and Health Care products. Under the Non-Personal Caring product category, there are Home Care, Auto Care as well as Home Technology Products. This completes the conception of the category of Care Products (Massaro, Maritati, & Galiano, 2018; Nikhashemi et al., 2017; Tarofder et al., 2019; Ulfah et al., 2019).

II. Literature Review

In early 1960s data mining was carried through statistical analysis. SAS, SPSS, and IBM were the colonizer of statistical analysis. Today in data mining filed All these companies are progressive and they have convincing presentation of element based upon their years of experience. Basically, correlation, regression, chi-square, and cross tabulation are basic statistical method which subsisted by statistical analysis. these classical approaches are still offered by SAS and SPSS in particular, but they carried these statistical measures to more insightful surroundings to justify or predict the data (Koh & Tan, 2011; Tarofder et al., 2016; Udriyah et al., 2019). The term of Data Mining may be used to describe wide range activities. In short, Data Mining is the process of exploring the structure of data. Aforementioned technology is mainly concerned with screening large quantities of data and evaluating the hidden information. The most beneficial aspects of this process is of course the fact that new knowledge can turn into litigable results. With this information, appropriate actions can be taken, such as increasing the customer's likelihood to buy, or reducing the number of deceptive claims. Data Mining is certainly one of the fastest growing fields in the computer industry. As mentioned above, Data Mining has been termed a various definitions. Apart from the short definition above, the author is digging more details and accurate definitions of Data Mining, and what it is often used as and in what kind of situation (Ahamed & Hariharan, 2011).

In some cases, the database may even grow larger than what the system administrators can handle. They do not always know exactly what information to be represented and how it might relate to the others. It is therefore, an advice for organizations to mine their database. This is to help visualize and understand the patterns underneath the large sets of data. And well, with today's advancement of technologies, more and more techniques and methods of data mining have grown rapidly. To discover the new hidden patterns and trends within the database can be performed with more accuracy nowadays (Singh, 2012). These are all the basic or general knowledge about data mining. This is a correlative process. Discovery is the progress, usually performed either automatically or even manually. Well, this progress is normally done in steps. First of all, an overall picture of the available data is obtained. Subsets of the data are then modeled and analyzed. New models would be formed if there are discoveries of any new interesting patterns, by re-sampling the data set. These new models are meant for emphasizing particular aspects of the data. This will be going on and on. There are no restrictions, or no specific rules or approaches in this iteration. It may occur within a single data source or across multiple sources. The most important key here is, it is important to adopt a flexible approach, as it may give more rooms to unexpected discoveries beyond bounds of the expected problem domain (Khajvand, Zolfaghar, Ashoori, & Alizadeh, 2011).

Data mining is a great problem solving technique as well as a natural activity that is best to be performed on a data warehouse. This leads to the entire data warehouse, data mart, and decision support communities being its largest target markets and attracts professionals from various industries, such as retail, health care, medical, insurance, transportation, and others. Apart from business industries, there are still many other areas that need data mining techniques as problem solvers, for example medical and microbiology industries. Prediction on the factors that are affecting the population rate and others like predicting the optimum respiration rate and also to prediction on the reaction of certain types of microorganisms towards certain kinds of chemicals, are some of the popular application of data mining in microbiology industry. As for the application in medical fields, data mining technique is often used as a predictive analysis on the reaction of a certain kind of chemicals towards the specific virus or bacteria and this will help them to produce better solutions for combating the fatal diseases (Dhond, Gupta, & Vadhavkar, 2000).

There are several functions for data mining to extract meaningful information from database. Overall, Classification, Estimation, Prediction, Association Rules or Affinity Grouping, Clustering, Description and Visualization that can be performed by six basic functions of data mining. The other three functions are the examples of the opposition of the first

three. They are actually the undirected data mining functions. The objective of directed data mining is different from the objective of undirected data mining. Predefined class can be assigned by examining the aspects of newly presented object through categorization. The main purpose of classification is to distinguish the unclassified and predefined variables through model building (Anderson & Kotsiopoulos, 2002). Database provides records which are helpful to classify objects. Thus, Class code is filled by updating records via classification. Estimation is practically dealing with the continuously valued outcome (as discrete outcome is what matters in classification) as compared to classification. Estimation is commonly used in a situation using some input data about unknown continuous variable such as height, weight and income etc.

There are still numbers of estimation tasks that might be practical, estimating the sales of a certain product, estimating the value or price of a product, estimating the numbers of members to be remaining as active members. As a conclusion here, classification and estimation are often used together. Prediction is, to some extent, related to both classification and estimation (Sumathi & Sivanandam, 2006). The difference is one of the stresses here. There would be no chance, or in other word, it would be expected not to be able to go back to make sure that there are no mistakes in the classification stage. Thus, no matter whether the classification is correct, incomplete knowledge is the only reason for the uncertainty. Here are some of the examples of prediction in this case, predicting which product would be topping the chart, predicting which customers / members would be leaving soon, predicting which types of products could be hitting on the probable customers. Prediction can also be adapted with the help of classification and estimation. By the use of examples the predicted value of variable can be compared by historical data (Hui & Jha, 2000).

Affinity grouping is done to evaluate the correlation between the objects. In this case, affinity grouping can be helpful to organize the items to be displayed together on shelves in their various outlet, as well as organizing the items which are often been purchased together be seen together. In the process of clustering diversified groups are segmented into smaller groups or subgroups of similar trait. It seems that both clustering and classification are quite the same, but they are not still. Predefined variables are not categorized in clustering as in classification (Chen, Han, & Yu, 1996). Only similarity counts in clustering. Self-similarity is the main pillar of clustering on which data is arranged. Data mining is all about prediction and classification of complicated database to achieve accurate knowledge. Behaviour of the variable is described and explain in efficient manner (Gong & Xia, 2009).

It is important to be remembered that the information that are needed when representing the actual values that are used to comprise the objects and their attributes can be well abstracted from the low level details into the composite representations. This is a useful technique, in which it provides a wide range of information needed in the earlier analytical process, on the classes of relationships and patterns in the data set. Composite representations can be formed, once the objects had been classified into each group of objects having similar user-selected characteristics, into a single object. The number of the underlying objects had been the main attribute here (Deshpande & Thakare, 2010). The original specific values would still be available for reference in the data source, but the analysis would still be performed on the abstracted values. For the groupings, the range and size of the values depends on the actual values represented in the abstraction (Han, Pei, & Kamber, 2011). The numbers of segments are not supposed to be either too large or too small. This might also be applicable to qualitative data, where there are natural levels of abstractions. A good example would be the separation of geographical regions (Shmueli, Patel, & Bruce, 2011). This is very much applicable in direct sales organization, where the stockiest would be distributed all around in Malaysia. They can be abstracted in the way that stores are located in districts, the districts are in cities, cities are in states, states are in regions, and etc. Predictive analysis is a study to find out or to forecast what will happen in the future. People have to face the reality that the analysis, no matter what is the nature of the analysis, it takes time. Careful thoughts and insights are needed to be done in order to interpret those indications that one have already taken

time to collect. In order to have a clear and accurate prediction of the future elements, careful and accurate analysis should be taken in action (Olson & Delen, 2008).

A prediction of sales forecast is based on the performance of previous sale and analysis of expected market conditions. Peoples are forced to take a look on future objective by true value of making in forecast. Because of limited product availability a forecast is pessimistic and the result is poor delivery performance, customers or unsatisfied and deficit revenue. There is need to establish the trends between the actual performance and past forecast to feedback the forecaster for perfect future forecasting (Seifert, 2004). There are a number of factors are responsible for slight deviations between projections and actual performance. Perhaps the forecasting process was considered different from the rest of the institution ; or external factors, such as trends or new legislation (Nemati & Barko, 2001). Significant discrepancies, however, such as a deviation of 15 percent or more in a month, or a cumulative deviation of 10 percent in a year, indicate a need for much more detailed analysis to determine what has occurred. A forecast is similar to a weather forecast is an educated guess about what the future will bring. People can predict all kinds of elements for example the future stock price, the sales and cost, the weather and so on (Köksal, Batmaz, & Testik, 2011). Wishful thinking compel people to make their guesses and forecast. Dollar sales value from the past several years predict about the forecast which is based on company's previous sales. Thorough sales forecast can be completed by the evaluation of all internal and external elements (Eze, Onwuegbuchulam, Ugwuishiwu, & Diala, 2017). Mathematically, it is possible to forecast sales with some precision. Realistically, however, this accuracy can be worn out due to external market and economic factors, discrepancies. There are some external factors that can affect sales, seasonality of the business, relative state of the economy, direct and indirect competition, political events, styles or fashions, consumer earnings, population changes, weather, productivity changes. External and internal factors need some sufficient information for sales forecast requirements (Ngai, Xiu, & Chau, 2009). There are also some internal factors that can affect sales are somewhat more controllable, such as, labour problems, credit policy changes, sales motivation plans, inventory shortages, working capital shortage, price changes, change in distribution method, production capability shortage and new product lines. After answering these following questions sales forecast are called qualified, what are the items to be forecasted (individual product lines or business units), how far in the future should the forecast extend, how frequently should the forecast be made, how frequently should the forecast be reviewed, what would constitute an acceptable tolerance of forecast error? It is significant that if a company sells more than one type of product or service, they should prepare a separate sales forecast for each service. More precise its outcome will be, if the more focused the sales forecast.

Hypothesis:

H1: There is significance of Controlling Sales via Data Mining application.

H2: Data Mining significantly enables account management.

III. Material & Methods

Hierarchical Framework

It is essential for a data mining practice to map the problems onto an explicit conceptual framework. Flexibility of thought should be given great emphasis in order to have this mapping done effectively, since it is likely to go out of bounds that anyone would be familiar.

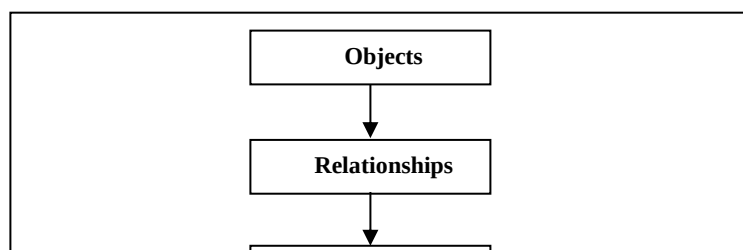


Figure 2: Hierarchical Framework for Knowledge Representation in Data Mining Applications

Some of the sample variables that can be observed in this project are listed below:

- ✧ Sales volume for different categories of products can be viewed and compared.
- ✧ Customers' accounts can be updated weekly.

Database

There is abundance of database application software everywhere now, due to the expansion of the database application systems that mushroomed today. Among those famous and well recognized are Microsoft Access, SQL Server, and Oracle.

Deriving Meta Data from dates, values, names, and addresses

Metadata here in data mining application is slightly different to those used in Artificial Intelligent. The metadata in this context is basically referring to the data within data. The purpose and goals of using this metadata is actually to be able to exploit it in a way that will make sense for the application and yield interpretable results. There are, however, some of them that occur commonly which are use of dates, values, names, and addresses.

Descriptive and Transactional Model

Declarative information belongs to descriptive model, while episodic information is part of transactional model. Meanwhile, transactional data model is contained of the episodic information about the time and place of the events, or transactions. This is to decide on the necessary models, based on the study of the data obtained.

Reactive and Proactive Analysis

Beginning the analysis in either a proactive or reactive mode can be decided on after the earlier activities. After the thinking of what the application would do, and the characterization of the problems into the conceptual frameworks, decision of the mode to be applied can be carried out. These two techniques are good at each area. We would like to review some of the major differences between both, and do a combination of both.

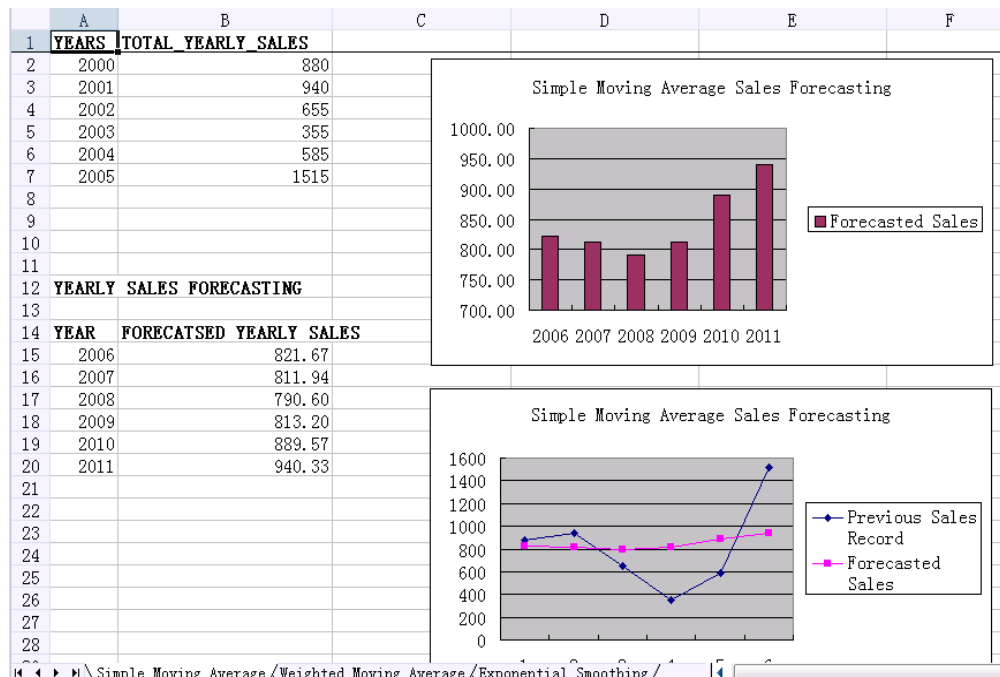
IV. Findings

Predictive analysis based on the sets of data obtained, the correctness of the samples taken and this will thus develop a confidence level in statistics. Annual growth and evaluation of sales level can flourish via forecast conduction and helpful for people to compare the company with industry. Sales Forecasting is practically done in part as the analysis in this project. This is done mainly for calculating and estimating the performance and the profitability of the organization. This is to have a better data mining on the system.

Simple Moving Average

The following example will demonstrate the idea of simple moving average forecasting.

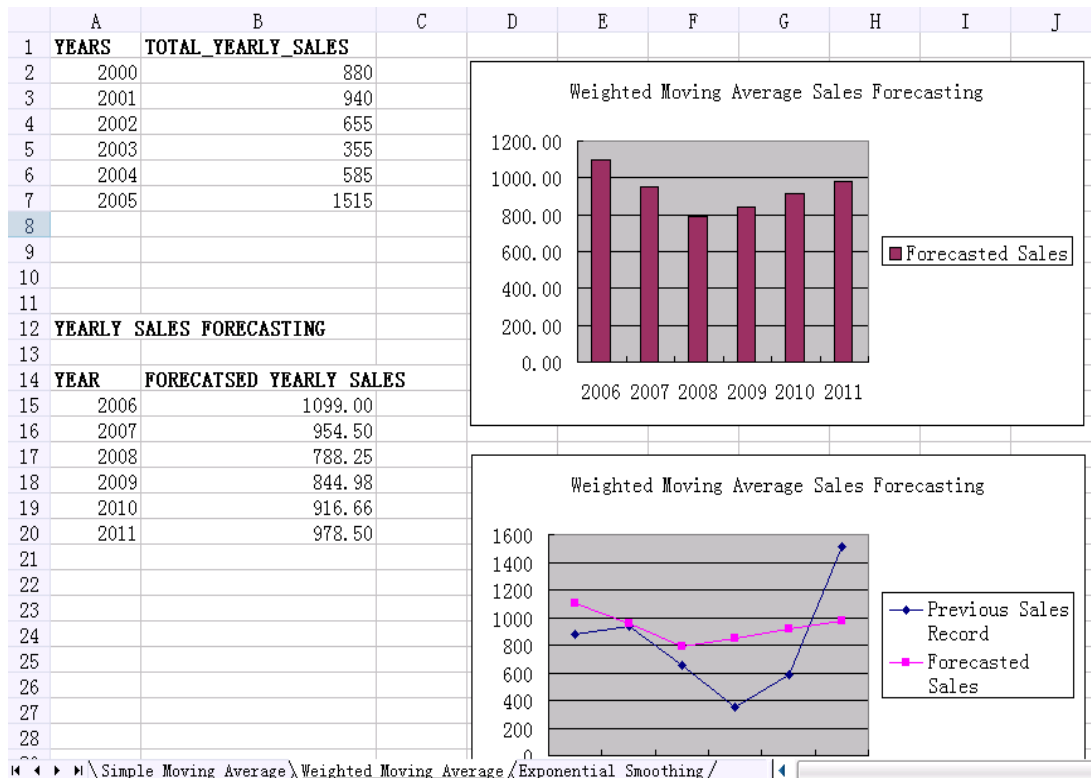
Forecast for year 2005 (for $n = 3$) = Actual Sales in 2002 + Actual Sales in 2003 + Actual Sales in 2004/3



Simple Moving Average

In a simple moving average the weight given to each of the 'n' observations. If 'n' is to equal 3, then each observation will add a third of the weight to the average. If so, they can use a weighting scheme such as: $W_t = 0.7$, $W_{t-1} = 0.2$ and $W_{t-2} = 0.1$. This scheme weights the current observation 't' at 0.7, the previous observation at 't-1' as 0.2 and the observation two time period back, 't-2', as 0.1. The weights employed are open to personal judgment and choice. The following example will demonstrate the idea of simple moving average forecasting.

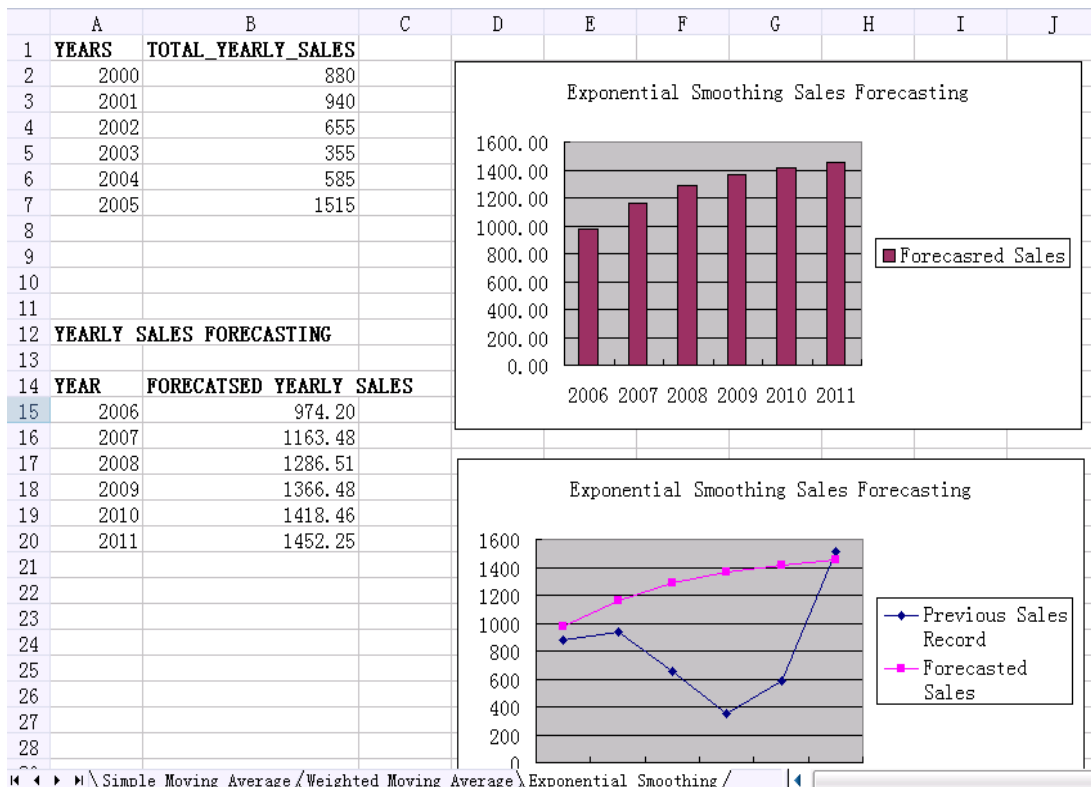
Forecast for year 2005 = Actual Sales in 2002 x 0.1+ Actual Sales in 2003 x 0.2 + Actual Sales in 2004 x 0.7



Exponential Smoothing

Alpha is open to choice, but a value of 0.2 to 0.3 is common.

Forecast for year 2005 = Alpha x Actual Sales in 2004 + (1- Alpha) x Forecasted Sales in 2004



Ratio Analysis

They are mainly the liquidity ratio, profitability ratio, use of assets, capital structures and the return of investment. As a matter of fact, only the first three ratios will be discussed here. This is mainly due to the irrelevance of the other ratios.

Liquidity Ratio

In this case, two of the commonly used ratios which addressed the situations about liquidity ratios, highlight such a situation.

Current ratio is one of them. It is obtained from the equation below.

Current Ratio= Current Assets / Current Liabilities

This current ratio actually measures the relationship between an organization's current assets and together with its current liabilities. As for an example, let's say Amway (M) has the current assets of RM90,000 and the current liabilities of RM30,000, then the current ratio would be 3:1 which is 90,000/30,000. With this, it is clear that for every RM1 of current liabilities, there would be RM3 of current assets.

Acid Test, on the other hand, is the second type of the liquidity ratio. The formula goes on this way:

Acid Test = (Current assets - Stocks) / Current liabilities.

In this example, the author would like to have the continuation from the first example. Let's say the stocks in Amway (M) then the acid test ratio would be:

$$(90,000-45,000)/ 30,000 = 1.5:1$$

Significantly high current and acid test ratios indicate that the cash flow is not having a problem. As such, it confirms the importance of liquidity ratio.

Profitability Ratio

There are two different percentages of profits under this category. They are actually the Gross profit percentage, or gross margin, and the net profit percentage. Gross Profit Percentage, or Gross Margin, is the first of the two. It is represented as the equation below:

Gross Profit Percentage or Gross Margin = (Gross Profit/ Sales) x 100%

In this case, the gross margin expresses the gross profit as a percentage of total sales. As an example, if Amway (M) makes a gross profit of RM150,000 on the total sales of RM250,000, then the gross margin would be $(150,000/250,000) \times 100\% = 60\%$. In other words, a gross margin of 60% would be obtained with the examples above. To simplify it, for every RM100 of sales, Amway is making a gross profit of RM60.

The second percentage of profit is the Net Profit Percentage, which can be expressed in this equation of:

Net Profit Percentage = (Net Profit / Sales) x 100%.

Well, the net profit percentage expresses the organization's net profit as a percentage of total sales. If, let's say Amway (M) is having RM25,000 of net profit for the sales of RM250,000, then the net profit percentage would be $(25,000/250,000) \times 100\% = 10\%$.

As in the case of this project, the author had done some analysis and a simple calculation on the profitability ratio for Amway (M).

The Use of Assets

Sales to fixed assets ratio is one of the ratios that can be used. The equal goes:

Sales to Fixed Assets Ratio = Sales/Fixed Assets

The sales to the fixed assets ratio usually indicates how much each of the Ringgit Malaysia (RM) invested generates in sales. For example, if the recorded total annual sales of Amway (M) is RM250,000, and the fixed assets is RM100,000, then the ratio of its sales to fixed assets would be 2.5:1. In other words, for every RM1 invested in the fixed assets, then there would be RM2.50 of sales is generated. In general, the relationship between the two (fixed assets and sales) is directly proportional. The higher the ratio of sales to fixed assets, the more efficiently fixed assets is being utilized.

Stock turnover ratio is another ratio using assets. The equation goes this way:

Stock Turnover Ratio = (Average stock held / Annual Cost of Sales) x 365

As for the case of this project, direct sales organizations like Amway (M), the stocks would not be lasting that long, as with the case of transport industry. The stock turnover ratio is therefore rather low. This is normally due to the fact that the common products that are distributed by the direct sales organizations are of high demand products. And the fact that it is direct sales organization, there will not be a product that are too many in quantity, however, there will not be too low level of stocks as well, Amway (M) would normally ensure that they are holding optimum levels of stocks.

Ratio Analysis of the Project

There are some of the relevant and important ratios that had been adopted for the purpose of this project. The ratios like the Efficiency and Profitability Ratios had been performed on the available data. The results of the analyses are stored in the program and can be retrieved as reference.

Ratio Analysis	
Income after Financial Items	0
Adjustment of Owners salary	0
Other adjustments	0
Income after adjustments	0
Equity, Subord. Reserves / Assets	0
Adjustment of Assets, Liability values	0
Adjusted Equity, Reserves / Assets	0
Quick Ratio	0
Current Ratio	0
Net interest-bear Debt/Equity	0
Interest coverage Ratio	0
Return on Assets	0
Return on Equity	0
Gross Profit	0
GP in % of Sales	0
GP incl personnel costs	0
GP incl personnel in % of Sales	0
Net Worth of Investments	0
NWI in % of Sales	0
Days in Receivables	0
Inventory Turnover	0 times/year
Equipment depreciation	0 year
Building depreciation	0 year
Intangible assets depreciation	0 year

Calculator Compute Update Cancel

MAMU MBA FINAL PROJECT BY A. VINCEN Beta Version 1.0

V. Conclusion

Data mining is an evolving technology whose implementation in the future will become mandatory in order for business organizations to remain competitive. There are some concerns with using data mining. The difficulty and complexity of data mining tool is one example. The proposed system is basically divided into three modules which are management, analysis and reporting. In the management module, the proposed system provided four kinds of management function which are customer, product, sales and financial management. With these functions, the organization can easily manage its customer relationship, maintain inventories, sales and financial records as for all of the data will be stored in database for future analysis. While in the analysis module, the proposed system provided two kinds of analytic functions which are financial analysis and sales forecasting. For the financial analysis, the proposed system will performed cash flow analysis, investment analysis and ratio analysis base on the selected income statement and balance sheet information that are previously stored in database. On the other hand, result of sales forecasting will be generated also based on the selected previously stored sales records in database. In the last module of the proposed system – reporting it provided insightful reports and also transformed numbers into easy-to-understand visual representation. With this, the decision makers are allowed to draw a better decision if compared with the traditional way of interpreting the sales records. In conclusion, by integrating automation, visual representation and predictive analytics into business processes, companies can dramatically improve the efficiency and accuracy of decision making. Lastly, regarding this project the author has learned a lot upon the completion of it. It has been, thus, a very satisfactory and fruitful journey of completing this project. This project cannot really be smooth sailing without the help of the people that the author had mentioned in the acknowledgement of this report. Nevertheless, there is a strong urge for the author to thank them once again here. The author would like to take this opportunity to express his sincere gratitude to all those people who had made this project a smooth sailing one. Finally, the author is very glad to have this project done in time. Since this project is a prototype developed within a time constraint, the functionalities are limited. It would have been better if more powerful programming languages such as Java or C++ have been selected in this case. As for the database, it is recommended to adopt MS SQL than Oracle as it is as powerful as Oracle with less complexity in the configuration. The database size in this project can be expanded as for a better forecasting. The larger the size of the database, the more accurate the forecasting become. Generally, this whole project needs to be polished to have a better effect.

REFERENCES

- [1] Ahamed, B. B., & Hariharan, S. (2011). A survey on distributed data mining process via grid. *International Journal of Database Theory and Application*, 4(3), 77-90.
- [2] Anderson, J., & Kotsiopulos, A. (2002). Enhanced decision making using data mining: applications for retailers. *Journal of Textile and Apparel, Technology and Management*, 2(3), 1-13.
- [3] Bose, R. (2009). Advanced analytics: opportunities and challenges. *Industrial Management & Data Systems*.
- [4] Cabena, P., Choi, H. H., Kim, I. S., Otsuka, S., Reinschmidt, J., & Saarevirta, G. (1999). *Intelligent miner for data applications guide*. IBM RedBook SG24-5252-00, 173.
- [5] Chen, M.-S., Han, J., & Yu, P. S. (1996). Data mining: an overview from a database perspective. *IEEE Transactions on Knowledge and data Engineering*, 8(6), 866-883.
- [6] De Silva A.D.A., Khatibi A., Azam S.M.F. (2018a). Can parental involvement mitigate swing away from science? Sri Lankan perspectives, *Cogent Education*
- [7] De Silva A.D.A., Khatibi A., Azam, S. M. F. (2018b). Do the Demographic Differences Manifest in Motivation to Learn Science and Impact on Science Performance? Evidence from Sri Lanka, *International Journal of Science and Mathematics Education*
- [8] Delafrooz N., Paim L.H., Khatibi A. (2009). Developing an instrument for measurement of attitude toward online shopping, *European Journal of Social Sciences*
- [9] Deshpande, S., & Thakare, V. (2010). Data mining system and applications: A review. *International Journal of Distributed and Parallel systems (IJDPS)*, 1(1), 32-44.

- [10] Dewi N.F., Azam, S. M. F., Yusoff S.K.M. (2019). Factors influencing the information quality of local government financial statement and financial accountability, *Management Science Letters*
- [11] Dhond, A., Gupta, A., & Vadhavkar, S. (2000). Data mining techniques for optimizing inventories for electronic commerce. Paper presented at the Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining.
- [12] Doa N.H., Tham J., Khatibi A.A., Azam S.M.F. (2019). An empirical analysis of Cambodian behavior intention towards mobile payment. *Management Science Letters*
- [13] Eze, U., Onwuegbuchulam, C., Ugwuishiwu, C., & Diala, S. (2017). Application of data mining in telecommunication industry. *International Journal of Physical Sciences*, 12(6), 74-88.
- [14] Gong, H., & Xia, Q. (2009). Study on application of customer segmentation based on data mining technology. Paper presented at the 2009 ETP International Conference on Future Computer and Communication.
- [15] Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*: Elsevier.
- [16] Hui, S. C., & Jha, G. (2000). Data mining for customer service support. *Information & Management*, 38(1), 1-13.
- [17] Khajvand, M., Zolfaghar, K., Ashoori, S., & Alizadeh, S. (2011). Estimating customer lifetime value based on RFM analysis of customer purchase behavior: Case study. *Procedia Computer Science*, 3, 57-63.
- [18] Koh, H. C., & Tan, G. (2011). Data mining applications in healthcare. *Journal of healthcare information management*, 19(2), 65.
- [19] Köksal, G., Batmaz, İ., & Testik, M. C. (2011). A review of data mining applications for quality improvement in manufacturing industry. *Expert systems with applications*, 38(10), 13448-13467.
- [20] Linoff, G. S., & Berry, M. J. (2011). *Data mining techniques: for marketing, sales, and customer relationship management*: John Wiley & Sons.
- [21] Maghfuriyah A., Azam, S. M. F., Shukri S. (2019). Market structure and Islamic banking performance in Indonesia: An error correction model, *Management Science Letters*
- [22] Massaro, A., Maritati, V., & Galiano, A. (2018). Data Mining model performance of sales predictive algorithms based on RapidMiner workflows. *International Journal of Computer Science & Information Technology (IJCSIT)*, 10(3), 39-56.
- [23] Nemati, H. R., & Barko, C. D. (2001). Issues in organizational data mining: A survey of current practices. *Journal of data warehousing*, 6(1), 25-36.
- [24] Nemati, H. R., & Barko, C. D. (2003). Key factors for achieving organizational data-mining success. *Industrial Management & Data Systems*, 103(4), 282-292.
- [25] Ngai, E. W., Xiu, L., & Chau, D. C. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert systems with applications*, 36(2), 2592-2602.
- [26] Nguyen H.N., Tham J., Khatibi A., Azam S.M.F. (2019). Enhancing the capacity of tax authorities and its impact on transfer pricing activities of FDI enterprises in Ha Noi, Ho Chi Minh, Dong Nai, and Binh Duong province of Vietnam , *Management Science Letters*
- [27] Nikhashemi S.R., Paim L., Haque A., Khatibi A., Tarofder A. K. (2013). Internet technology, Crm and customer loyalty: Customer retention and satisfaction perspective , *Middle East Journal of Scientific Research*
- [28] Nikhashemi S.R., Valaei N., Tarofder A. K. (2017). Does Brand Personality and Perceived Product Quality Play a Major Role in Mobile Phone Consumers' Switching Behaviour? *Global Business Review*
- [29] Olson, D. L., & Delen, D. (2008). *Advanced data mining techniques*: Springer Science & Business Media.
- [30] Pambreni Y., Khatibi A., Azam, S. M. F., Tham J. (2019). The influence of total quality management toward organization performance, *Management Science Letters*
- [31] Pathiratne S.U., Khatibi A., Md Johar M.G. (2018). CSFs for Six Sigma in service and manufacturing companies: an insight on literature, *International Journal of Lean Six Sigma*
- [32] Rachmawati D., Shukri S., Azam, S. M. F., Khatibi A. (2019). Factors influencing customers' purchase decision of residential property in Selangor, Malaysia , *Management Science Letters*
- [33] Rygielski, C., Wang, J.-C., & Yen, D. C. (2002). Data mining techniques for customer relationship management. *Technology in society*, 24(4), 483-502.
- [34] Seifert, J. W. (2004). Data mining: An overview. *National security issues*, 201-217.
- [35] Seneviratne K., Hamid J.A., Khatibi A., Azam F., Sudasinghe S. (2019). Multi-faceted professional development designs for science teachers' self-efficacy for inquiry-based teaching: A critical review, *Universal Journal of Educational Research*
- [36] Shmueli, G., Patel, N. R., & Bruce, P. C. (2011). *Data mining for business intelligence: Concepts, techniques, and applications in Microsoft Office Excel with XLMiner*: John Wiley and Sons.
- [37] Singh, H. (2012). Implementation Benefit to Business Intelligence using Data Mining Techniques. *International Journal of Computing & Business Research, Proceedings of „I-Society*.
- [38] Sudari S.A., Tarofder A.K., Khatibi A., Tham J. (2019). Measuring the critical effect of marketing mix on customer loyalty through customer satisfaction in food and beverage products, *Management Science Letters*
- [39] Sumathi, S., & Sivanandam, S. (2006). *Introduction to data mining and its applications* (Vol. 29): Springer.
- [40] Tarofder A.K., Azam S.M.F., Jalal A. N. (2017). Operational or strategic benefits: Empirical investigation of internet adoption in supply chain management, *Management Research Review*

- [41] Tarofder A.K., Haque A., Hashim N., Azam, S. M. F., Sherief S. R. (2019). Impact of ecological factors on nationwide supply chain performance, *Ekoloji*
- [42] Tarofder A.K., Jawabri A., Haque A., Azam S.M.F., Sherief S.R. (2019). Competitive advantages through it-enabled supply chain management (SCM) context, *Polish Journal of Management Studies*
- [43] Tarofder A.K., Nikhashemi S.R., Azam S. M. F., Selvantharan P., Haque A. (2016). The mediating influence of service failure explanation on customer repurchase intention through customers' satisfaction, *International Journal of Quality and Service Sciences*
- [44] Tsai, H.-H. (2012). Global data mining: An empirical study of current trends, future forecasts and technology diffusions. *Expert systems with applications*, 39(9), 8172-8181.
- [45] Udriyah, Tham J., Azam, S. M. F. (2019). The effects of market orientation and innovation on competitive advantage and business performance of textile SMEs, *Management Science Letters*
- [46] Ulfah R., Amril Jaharadak A., Khatibi A.A. (2019). Motivational factors influencing MSU accounting students to become a certified public accountant (CPA), *Management Science Letters*