AN EXAMINATION ON INFORMATION MINING INNOVATION AND EXECUTING OF AN INFORMATION DIGGING APPLICATION FOR DEALS PRESCIENT INVESTIGATION

[1]Johar, MGM, S.M. Ferdous Azam, Mohd Shukri Ab Yajid

Abstract---Amway (M) has been taken into account in context of Data Mining technology and implementation of predictive analysis of sales. The references at this stage are mainly books, journals and articles on the Internet about Data Mining, as well as interviews and observations which are carried out to obtain the necessary information about Amway (M). The second part of this project is on the analysis of the data obtained. The analysis was done based on the database constructed. A model is built after this, as a guide to the real implementation in the later stage. The application developed was mainly to have the comparative studies showed here in visual forms. Various types of graphics were adopted in this program as to visualize the comparisons of the figures. Testing was well carried out after the development of the application. Testing was done as to ensure that this program worked well in the specified platform. Reports were produced based on the testing procedures.

Keywords---Technology, Sales Volume, Amway, Products, Oracle, Database, Graphics, System, Mining.

I. Introduction

Because of present day data innovation, which delivers always incredible PCs consistently, today is conceivable to gather, move, join, and store colossal measures of information at low expenses. In this way an ever-expanding number of organizations logical and legislative establishments can beer to develop colossal assortment of tables, archives and sounds in computerized structure. The idea is convincing that if individuals have enough information; they can take care of any issue (at any rate on a fundamental level). Notwithstanding, a closer assessment uncovers the information alone are not adequate. It very well may be said that in enormous database individuals for the most part can't see the wood of the trees (Strohmeier & Piazza, 2013; De Silva et al., 2018a; De Silva et al., 2018b; Nikhashemi et al., 2013). Albeit any single piece of data can be recovered and straightforward totals can be registered, general examples, structures, just as regularities for the most part go undetected and frequently these examples are particularly important. For example, if a general store finds that specific items are as often as possible purchased together, the quantity of things sold can here and there be expanded by properly organizing these items in the racks of the market.

Background of Amway (M)

[1]*Management and Science University*
mdgapar@msu.edu.my

An organization that has started with almost no workers in a little office with next to no office of stockroom in jalan Ipoh in 1976 and Amway (M) was one of the pioneers in the immediate selling industry around then. The dissemination is made through a board system, comprising of the individuals and deals focuses all through Malaysia (Anderson, Jolly, & Fairhurst, 2007). Today, Amway (M) has become the main direct selling organization in Malaysia with a center merchant power of 161,000 from all sides of the country and a business turnover of RM456 million, making Amway an easily recognized named in Malaysia. Amway (M)disperse a wide scope of amazing quality purchaser items, extending from day bay day merchandise, human services, home consideration, home innovation, auto care, individual consideration items to design. The distribution is made through an extensive network, consisting of the members and sales centers throughout Malaysia. Therefore, mining their data is one of the crucial steps in order to meet their challenges toward the future.

Procedural and Declarative Knowledge

It is important for a data mining project to have analyzed and characterized the information to be contained into either procedural or declarative, or the combination of both. It is known that procedural knowledge will normally contain the information that is sequential, or in other words, the information that must be maintained in sequence, while the declarative knowledge does not really require any of this. Whether or not to model the procedural or the declarative information differently is up to the analysts (Hajizadeh, Ardakani, & Shahrabi, 2010; Dewi et al., 2019; Pambreni et al., 2019; Tarofder et al., 2017). There are chances that there are differences among the types of questions for procedural and declarative data sets. Well, structuring them ahead of time so that each of them can be accessed when needed is a good option for the problem solving.

Distinguish Between Meta Knowledge and Actual Knowledge

It is a good practice to segment the classes of information along two dimensions, when thinking about the problem space of an application. The two dimensions are the real or actual status of information contained in the data source, and as well as the metadata about what information may that exist in the data sets. In this context, it is necessary to think the information in the way of known and unknown, and the analyst may or may not be correct in the judgment.
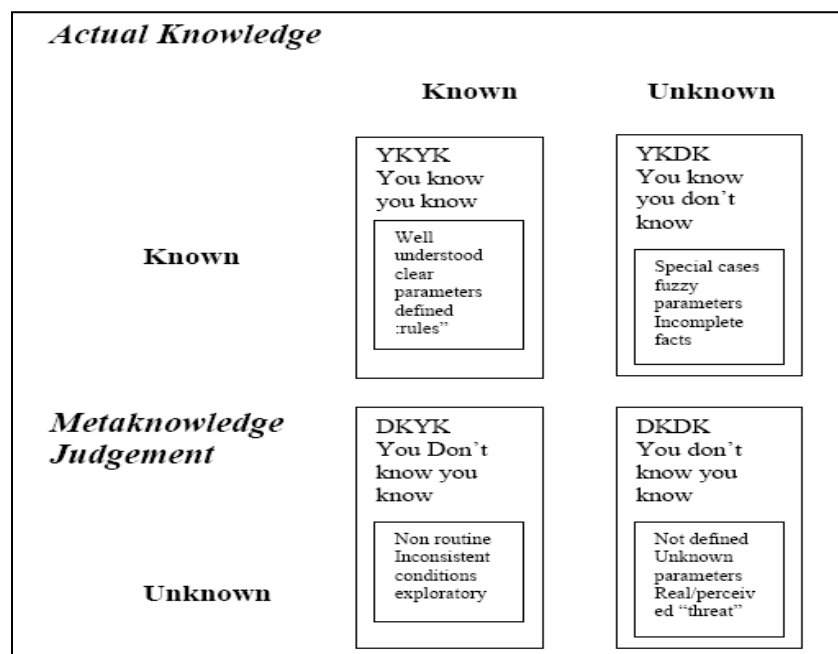
Figure 1: Dimensions of Met knowledge and Actual Knowledge within the problem

These dimensions can be used to construct two-by-two matrix as a method for description of the status of the information available during a data mining process.

Information that You Know You Know (YKYK)

Well, the kind of information that the organizations normally put into the YKYK cells are those standard activities, like generating reports, processing data, running customers' credit check, and many others. It is appropriate to say that this YKYK information is typically used for those "inside the box" kind of thinking in the accepted models.

Information that You Know You Do Not Know (YKDK)

This may in turn, reflecting in the business context. In this case, there are many events, or information that the managers know that they do not know. They intend to find out the products that would normally be purchased together in conjunction. They want to know whether the sales promotions are effective and are they targeting the best markets for particular products.

Information that You Do Not Know You Know (YDYK)

Well, it is always a tradition for the organizations to maintain the databases full of information. Many interesting results can actually be found in these large databases. Depending on the kinds of analyses, many sophisticated and useful results lead to the improvements of the operations quality. Forecasting of future behaviors, exposure of frauds, performance enhancements and other stuffs can be performed along with the DKYK discoveries (Zhang, Edwards, & Harding, 2007). The main goal of the data mining in terms of DKYK is actually to move this category up to the YKYK category, so as it is often clear and obvious and would be greater usage later on. This would lead to the easier detection of future recurrence and so they would be acted on accordingly.

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

The gaps in knowledge of this DKDK cell are not being defined. They are mostly based on unknown parameters, and are never been noted. It is thus, an important matter to discover and interpret the gaps in a way that would help to deal with the threats to organizations, either real or perceived. One way of helping the organizations here would be the detection of the missing information. This actually helps in various ways (Pujari, 2001; Doa et al., 2019; Maghfuriyah et al., 2019; Nguyen et al., 2019).

The types of questions that would be needed to answer in the data mining process would be considered in the very beginning. Talking to the clients, with the intention of finding out what they really need, is the correct way of determining the initial direction of the analysis. The kinds or types of findings that are most helpful to the organization must be sought after. This, at the broad conceptual level, the products of Amway (M) can broadly be divided into, for example, care products and non-care products. Within the Care Products Category, this can further be divided into People's Caring Products and Non-People's Caring products. This goes on and on, until the smallest level is me (Berry & Linoff, 2004; Pathiratne et al., 2018; Rachmawati et al., 2019; Seneviratne et al., 2019; Sudari et al., 2019; Tarofder et al., 2019).

II. Literature Review

In the late 1980s, traditional measurable investigation was improved with a lot of methods which are fluffy rationale, heuristic thinking, and neural systems. This was the brilliance long stretches of man-made conscious (artificial intelligence). Presently in the late 1990s, individuals have figured out how to adopt the best strategies from traditional measurable investigation, neural systems, choice trees, advertise crate examination, and other incredible methods, and play out a

significantly more convincing and successful way. Well, all these should be started in the very basic level of explanation(S. Sharma, Goyal, & Mittal, 2008; Nikhashemi et al., 2017; Tarofder et al., 2019; Ulfah et al., 2019). First of all, considering the fact that most organizations, no matter it is private business or even under governmental operation, most likely have a large amount of resources spent on the construction as well as maintenance of the large database that they have. Well, since the scale of this database is fairly large, those data are frequently cannot be analyzed by the standard statistical methods. This is mainly due to the fact that there would be missing records during the analysis, or they are simply in the form of qualitative measures, rather than in quantitative form. This would normally lead to the under-value and underutilization of the information in this database. These limitations are mainly due to the difficulties in the assessments and analysis of the data (Kantardzic, 2011; Pathiratne et al., 2018; Rachmawati et al., 2019; Seneviratne et al., 2019; Sudari et al., 2019; Tarofder et al., 2019).

With the general knowledge about data mining in mind, the author has come out with his own realization of the general ideas of data mining. What is obvious here is the fact that standard mathematical or even statistical analyses are not present in this analysis, for instance, given a set of data on the total sales volume of Amway (M) during the year 2004, some sort of patterns and trends of customers buying behaviors can be detected. These are based on the predictive analysis, and not with a specific, or predefined mathematical and statistical notions (Berkhin, 2006). These predictions might be reflecting the real sales patterns, or might be totally not reflecting the true situations at all. Predictive analysis in data mining sometimes tend to have mistaken, or inaccuracies, while analysis using standard predetermined notions are hardly inaccurate. As for the necessary tools and techniques for applying data mining is not of too much a concern. There are many options available. Although these tools and technologies are available, and some of them are better in terms of ease of use, and affordability, the users must have one important thing to be ensured of, this would certainly be the fact that the tools alone, no matter how effective it is, will never contribute to the entire solution (Han, 1998; Tarofder et al., 2016; Udriyah et al., 2019). The final decisions are still need to be made by the practitioner himself. The analyst must be responsible in finding out the better way of exploiting, manipulating, as well as exposing the critical patterns and relationships of the data. Various combinations of techniques are used in order to achieve this. It thus depends on the analysts themselves to produce an outstanding application, using combinations of several of the available tools and techniques.

As in the business community, data mining techniques are often adopted as to discover the new purchasing patterns among the customers. With this discovery, they can then plan a better investment strategy and also provide a better sales plan. This can help them to have better-planned marketing strategies and thus resulting in optimum turnovers as well as having optimum profits (Bukhbinder, Krumenaker, & Phillips, 2005). Data mining is of course applied in other industries besides those mentioned but they have one thing in common, to adopt the technique to study and analyze the hidden patterns and trends in the data, as to develop end products that are of interests to the majority, or to provide solutions that are vital for living a better and healthier life (Ullah, 2019). Among the six essential capacities over, the initial three capacities are the instances of coordinated information on mining. Right now, fundamental objective is to assemble a model portraying the specific variables of interests utilizing the accessible information for instance, ordering protection claims, assessing advance adjust, recognizing possibilities, and foreseeing weakening .it is to build up certain connection among all the factors, For instance, figuring out what items ought to be bunch together for a claim to fame list, discovering gatherings of peruses or audience members with cooperative preferences for books or music, and finding regular client portions for showcase investigation (Ahmed, 2004).

A case of ordering is to inspect an applicant client and appoint that client to a foreordained bunch or grouping. Another case of arranging is medical determination. In the two cases, a portrayal of the client or patient is encouraged into the arrangement calculation. The classifier decides to which group "centroid" the competitor client or patient is closest or

generally comparable. We might be characterizing clients as credit commendable or credit dishonorable, or we might be ordering patients as either require or not expect of treatment (Chien & Chen, 2008). There are, somewhat, a few connections exist between these two elements of grouping and estimation. Estimation is frequently used to play out an order task. For example, when deciding whether or not to increase the distributed items, the top level management of Amway (M), would perform some kind of decisive actions, such as running this in a model and assigning score for each of the decisions. They may have a range of numbers between 0-1. This is only an estimation of the likelihood that they may expand the things on offer (Delen, 2014).

Other grouping assignments may be reworked as estimation errands also, for an occurrence, there may be situations where the quantities of individuals are excessively enormous. There might be a consideration of whether or not to delete all the sleeping members, and list only the active members. In this case, there would be a situation where the management would figure out who are the active members and who are to be deleted. This may likewise be seen as grouping individuals as likely or probably not going to be dynamic, or it might be seen as the estimation of the time frame, or period , that a part may be dynamic or remain with Amway (Tan, Steinbach, & Kumar, 2016). The forecast is diverse in light of the fact that the records are grouped by some anticipated future practices or evaluated values. The main thing that should be possible to check for the precision of the grouping is to keep a watch out. Here are a portion of the instances of forecast right now, which item would be beating the diagram, foreseeing which clients/individuals would leave soon, anticipating which sort of items could be hitting on the plausible clients. The fundamental errand of partiality gathering is to figure out which of the things go together. Right now, gathering can be utilized to design the game plan of things to be shown together on rocks in their different outlet, just as sorting out the things which are regularly been bought together be seen together (Ngai, Hu, Wong, Chen, & Sun, 2011).

Grouping is regularly done as a prelude to some other type of information mining or displaying. For example, it may be initial step of market division, isolating the clients with comparative purchasing propensities before settling on the sorts of endeavors for any advancements and stuffs. The appropriate promotion can be come up with each type of the cluster(Gupta & Lehal, 2009)., Information representation, then again, is an incredible type of unmistakable information mining. Important representations are not simple to concoct, however it is extremely helpful and significant to have information perception. "Words generally can't do a picture justice" the privilege visual would extremely worth a thousand affiliation rules. Well, with this, the scope of decent variety of the chose component of interests, spoke to in the informational index, can be seen in top-down level (A. Sharma & Kaur, 2017). Apart from this, the abstraction is also allowing the processing of more data at a time within the data mining applications. Despite the usefulness and advantage that abstraction is providing, the main shortage of this abstraction is the fact that the details are missing out with the levels of abstraction. The more abstraction is used, the more details are missing out. It is thus, cannot be used to access in the top level of the analysis. This shortage is actually not an importance or serious matter, since there is a drill-down property to view the individual data elements if needed. Abstraction, with its great usage, is typically used in the sense of help detecting hot spots within the data, in order to guide the selection of data (Farooqi & Raza, 2012).

Apart from the object classes, the attributes, including the numeric values, can also be abstracted into groupings, which are called either bins or categories. In this case, if the amount or value of the sales is used in a specific data model, finding the patterns is pretty hard to be done with the specific values as $1230.50. Since the patterns are hard to find in these types of values, it is thus, useful to segment the numeric attribute values into a set of predefined groups that permits more generalized pattern detection. Since the gauge depends on the organization's past deals, it is important to know the dollar deals volume for as long as quite a long while (Mohammad, Ahmed, & Zaman, 2017). To finish an exhaustive deals figure, it likewise need to mull over the entirety of the components, both inward and outside, that can influence deals, Numerically,

it is conceivable to conjecture deals with some accuracy. Everything being equal, be that as it may, this exactness can be dulled due to outside market and financial elements that are past the individuals' control. Coming up next are a portion of the outside components that can influence deals, regularity of the business, relative condition of the economy, immediate and circuitous challenge, political occasions, style or forms, purchaser profit, populace changes, climate, efficiency changes (Bhattacharyya, Jha, Tharakunnel, & Westland, 2011). Deals anticipating requires adequately point by point examination of both the outer and interior elements identified with the business work. Inward factors that can influence deals are some degree progressively controllable, for example, work issues, credit arrangement changes, deals inspiration plans, stock deficiencies, working capital lack, value changes, change in appropriation strategy, creation capacity deficiency and new product offerings. The business estimate must be qualified by posing the accompanying inquiries, what are the things to be gauge (singular product offerings or specialty units), how far later on should the conjecture broaden, how as often as possible should the figure be made, how every now and again should the gauge be explored, what might establish an adequate resilience of figure mistake?

Generally, Predictive analysis is done based on the sets of data obtained, the correctness of the samples taken and this will thus develop a confidence level in statistics. For more accuracy, larger sample should be obtained or in other words, the larger the sample data, the more accurate the prediction is. These samples should be obtained and analyzed carefully and precisely in order to have a more accurate prediction of the future trends. The larger sample would produce better and accurate prediction, but in addition, the cost of the predictive analysis would go up as well. However, due to the advancement of today's technology, it is now possible to obtain a large data samples at minimal cost (Gorunescu, 2011). This encourages the predictive analysts to produce accurate analysis or predictions based on more accurate data sampling. This is a real advantage and should be used to the maximum. Directing a business estimate will furnish the business either an assessment of the past and current deals levels and yearly development, and permit individuals to contrast the organization with industry standards. It will likewise help individuals to setup the strategies so they can without much of a stretch screen the costs and working expenses to ensure benefits, and make them mindful of minor issues before they become serious issues. No one can gauge with 100 percent precision. However individuals can say with 100 percent exactness that a poor figure will contrarily affect the organization. Obviously, it will never totally take out the vulnerability in estimates, yet it very well may be decrease to a sensible level (Kaur & Singh, 2014). By gathering demonstrated realities and testing any significant suspicions in a figure before lead the gauge, it can extraordinarily lessen the mystery and increment the exactness of the estimate. Issues emerge when estimates are excessively high, yet in addition when they are excessively low. In the event that an estimate is excessively idealistic, money is regularly tied up in moderate moving stock, and net revenues are diminished because of squandered overhead (Olmeda & Sheldon, 2002).

Hypothesis

H1: Data mining application plays vital role in Data Management.

H2: Data mining helps to strive sales from different angles.

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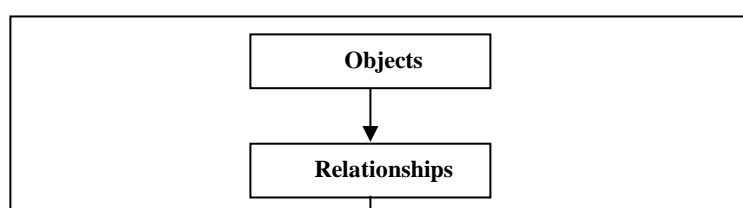
Figure 2: Hierarchical Framework for Knowledge Representation in Data Mining Applications

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

✧ Individuals can be registering as members

✧ 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.

Mathematically, it is possible to forecast sales with some precision. Realistically, however, this precision can be dulled because of external market and economic factors that are beyond the people's control. The following are some of the 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. Internal factors that can affect sales are somewhat more controllable, such as, labor problems, credit policy changes, sales motivation plans, inventory shortages, working capital shortage, price changes, and change in distribution method, production capability shortage and new product lines.

IV. Findings

Predictive analysis is done based on the sets of data obtained, the correctness of the samples taken and this will thus develop a confidence level in statistics. Conducting a sales forecast will provide the business with an evaluation of past and current sales levels and annual growth, and allow people to compare the company to industry norms. Perhaps the forecasting process was viewed separately from the rest of the organization; or external factors, such as trends or new legislation, that are not taken into consideration. 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
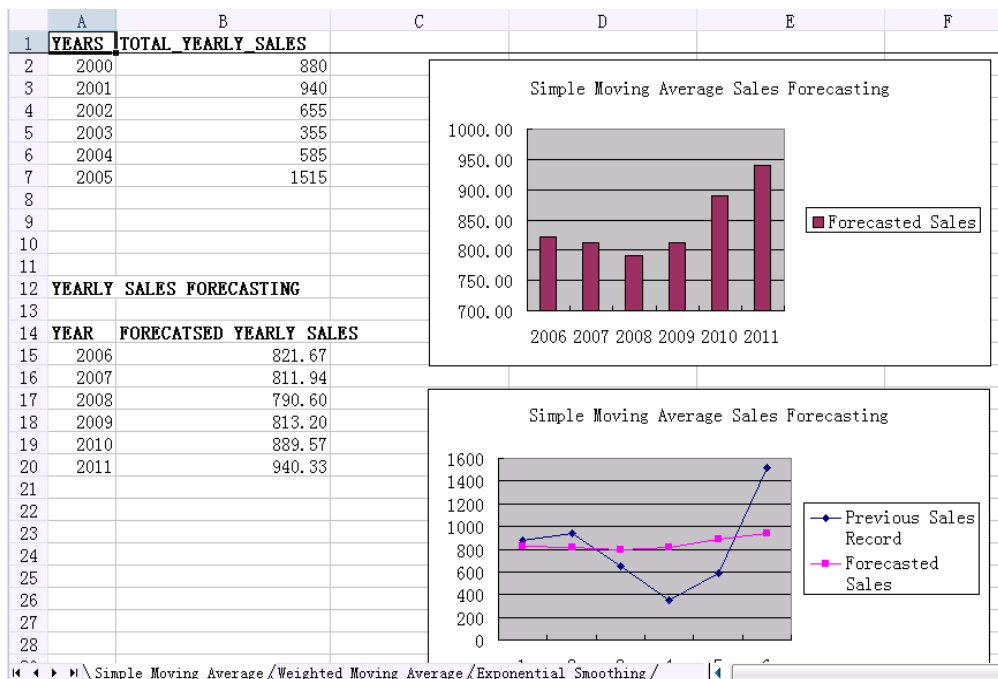
Figure 3: Sales Forecasting of the Project – Simple Moving Average
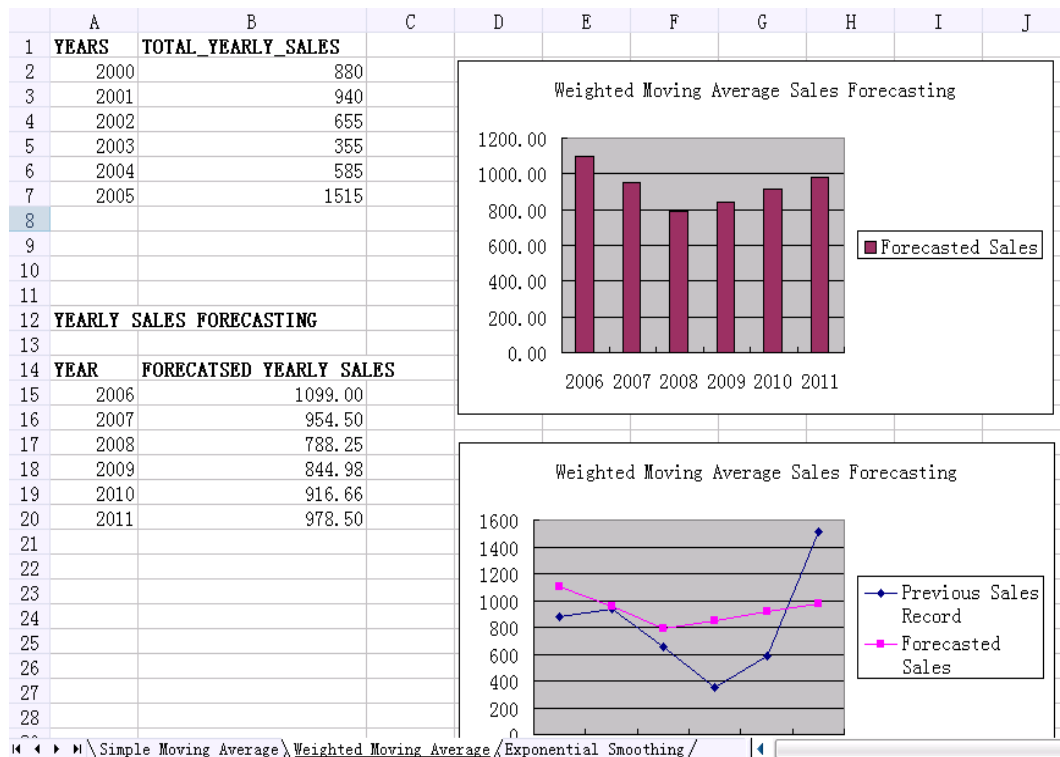
Weighted Moving Average

In a simple moving average the weight given to each of the 'n' observations is uniform. 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: Wt = 0.7, Wt-1 = 0.2 and Wt-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
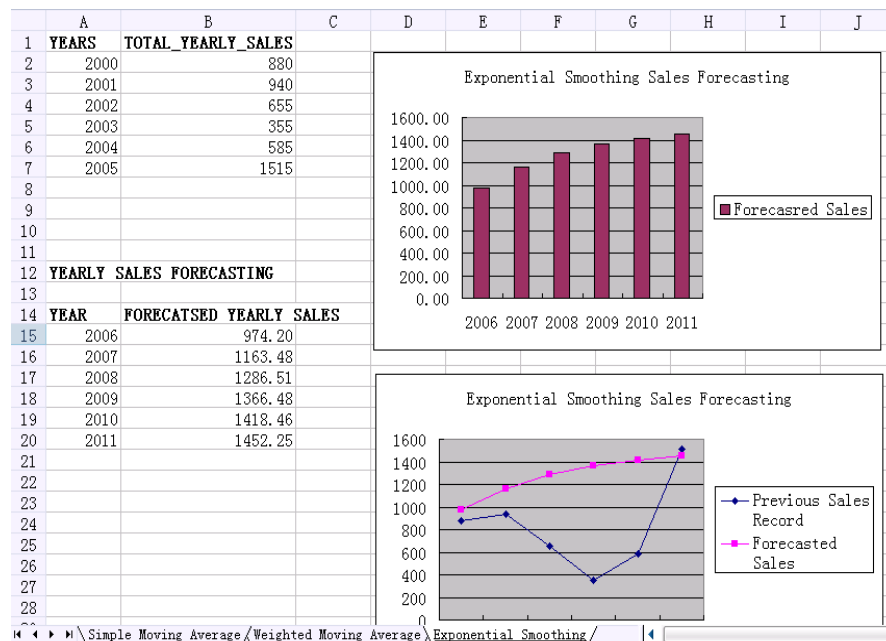

Figure 4: Sales Forecasting of the Project – Weighted Moving Average

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

Figure 5: Sales Forecasting of the Project – Exponential Smoothing

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.

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) x 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) x 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).

V. Conclusion

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 (Nawaz, Afzal, & Shehzadi, 2013).

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. Generally, this whole project needs to be polished to have a better effect (Nawaz, Azam, & Bhatti, 2019).

REFERENCES

[1]   Ahmed, S. R. (2004). Applications of data mining in retail business. Paper presented at the International Conference on Information Technology: Coding and Computing, 2004. Proceedings. ITCC 2004.

[2]   Anderson, J. L., Jolly, L. D., & Fairhurst, A. E. (2007). Customer relationship management in retailing: A content analysis of retail trade journals. Journal of Retailing and Consumer Services, 14(6), 394-399.

[3]   Berkhin, P. (2006). A survey of clustering data mining techniques Grouping multidimensional data (pp. 25-71): Springer.

[4]   Berry, M. J., & Linoff, G. S. (2004). Data mining techniques: for marketing, sales, and customer relationship management: John Wiley & Sons.

[5]   Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. C. (2011). Data mining for credit card fraud: A comparative study. Decision support systems, 50(3), 602-613.

[6]   Bukhbinder, G., Krumenaker, M., & Phillips, A. (2005). Insurance Industry Decision Support: Data Marts, OLAP and Predictive Analytics. Paper presented at the Casualty Actuarial Society Forum.

[7]   Chien, C.-F., & Chen, L.-F. (2008). Data mining to improve personnel selection and enhance human capital: A case study in high-technology industry. Expert systems with applications, 34(1), 280-290.

[8]   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

[9]   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

[10]   Delafrooz N., Paim L.H., Khatibi A. (2009). Developing an instrument for measurement of attitude toward online shopping, European Journal of Social Sciences

[11]   Delen, D. (2014). Real-world data mining: applied business analytics and decision making: FT Press.

[12]   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

[13] 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

[14] Farooqi, M., & Raza, K. (2012). A Comprehensive Study of CRM through data mining Techniques. arXiv preprint arXiv:1205.1126.

[15] Gorunescu, F. (2011). Data Mining: Concepts, models and techniques (Vol. 12): Springer Science & Business Media.

[16] Gupta, V., & Lehal, G. S. (2009). A survey of text mining techniques and applications. Journal of emerging technologies in web intelligence, 1(1), 60-76.

[17] Hajizadeh, E., Ardakani, H. D., & Shahrabi, J. (2010). Application of data mining techniques in stock markets: A survey. Journal of Economics and International Finance, 2(7), 109.

[18] Han, J. (1998). Towards on-line analytical mining in large databases. ACM Sigmod Record, 27(1), 97-107.

[19] Kantardzic, M. (2011). Data mining: concepts, models, methods, and algorithms: John Wiley & Sons.

[20] Kaur, B., & Singh, W. (2014). Review on heart disease prediction system using data mining techniques. International journal on recent and innovation trends in computing and communication, 2(10), 3003-3008.

[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] Mohammad, R., Ahmed, M. B., & Zaman, M. B. (2017). Predictive analytics: an application perspective. International Journal of Computer Engineering and Applications, 11(VIII).

[23] Nawaz, M. A., Afzal, N., & Shehzadi, K. (2013). Problems of formally employed women: A case study of Bahawalnagar, Pakistan. Asian Journal of Empirical Research, 3(10), 1291-1299.

[24] Nawaz, M. A., Azam, M. A., & Bhatti, M. A. (2019). Are Natural Resources, Mineral and Energy Depletions Damaging Economic Growth? Evidence from ASEAN Countries. Pakistan Journal of Economic Studies, 2(2), 15-28.

[25] Ngai, E. W., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. Decision support systems, 50(3), 559-569.

[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] Olmeda, I., & Sheldon, P. J. (2002). Data mining techniques and applications for tourism internet marketing. Journal of Travel & Tourism Marketing, 11(2-3), 1-20.

[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] Pujari, A. K. (2001). Data mining techniques: Universities press.

[33] 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

[34] 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

[35] Sharma, A., & Kaur, B. (2017). A Research review on comparative analysis of data mining tools, techniques and parameters. International Journal of Advanced Research in Computer Science, 8(7).

[36] Sharma, S., Goyal, D., & Mittal, R. K. (2008). Data mining research for customer relationship management systems: a framework and analysis. International Journal of Business Information Systems, 3(5), 549-565.

[37] Strohmeier, S., & Piazza, F. (2013). Domain driven data mining in human resource management: A review of current research. Expert systems with applications, 40(7), 2410-2420.

[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] Tan, P.-N., Steinbach, M., & Kumar, V. (2016). Introduction to data mining: Pearson Education India.

[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]   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

[45]   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

[46]   Ullah, M. A. (2019). A model for predicting outfit sales: Using data mining methods Emerging Technologies in Data Mining and Information Security (pp. 711-720): Springer.

[47]   Zhang, X., Edwards, J., & Harding, J. (2007). Personalised online sales using web usage data mining. Computers in Industry, 58(8-9), 772-782.