Descriptive and Predictive Analytics on Adventure Works Cycle: A Corporate Decision Making

Yew Liong Lim, Raheem Mafas

Abstract: Business Analysis has become one of the crucial elements of any business in this data-driven business world. This is at the frontline where the data analytics support the strategic management to make effective decisions with immense computing power. This paper investigates the big data problems of Adventure Works Cycles (AWC) by using analytical techniques and integrate different methods of knowledge discovery and data mining via descriptive and predicative analytics. The descriptive analytics revealed the prevailing business condition which could aid to make effective decisions. Consequently, an empirical study was performed to explore different types of predictive models to predict the future occurrences. Furthermore, a comparative analysis using different predictive algorithms which provides evidence that High-Performance Forest algorithm is particularly operative on the prediction of future occurrences with the accuracy of 80%, ROC index 0.878 and the cumulative lift value of 1.82. This study provides an intuitive grasp of the concept to forecast, find patterns and rules to increase AWC's overall sales performance and improve overall lead scoring more accurately.

Keywords: business analysis, visual analytics, descriptive analytics, predictive analytics.

I. INTRODUCTION

Adventure Works Cycles (AWC), is an international manufacturer and seller of bicycles and accessories. The company's headquarters is located in Bothell, Washington, USA and have 3 main regional sales offices in America, European and Pacific. AWC would like to expand their sales to their best customers, extending their product availability through an external web site, while maintaining the lower production costs. AWC sales mainly come from two sales channels, such as resellers and internet/online [1]. There are three major product categories for AWC online sales namely bikes, clothing and accessories & components. The core income for AWC is obtained by selling three main brands of bikes such as mountain bikes, road bikes and touring bikes bicycles. An additional income is generated by selling cycling accessories such as bottles, caps, gloves, jerseys, and components namely bike racks, brakes chains, handlebars, etc.

Business analysis is significant to any business to make sure that the business's needs are decisive and help to embrace improvements as well as effective decision making. This is expected to happen by bridging the gap between IT and business strategies confirming the core processes are in place and deliver the required benefits to the businesses and its stakeholders. The business analysis enables data-driven decisions and communicating complex technical details more easily. Business strategies and tactics are the main defined topics which are advisable and drive upon business changes whenever required [2].

The business analysis brings benefits with the power of data which makes the businesses to become analytically driven and helps to gain a new approach for making data-driven decisions. It brings deeper insights into business and considered an imperative facet for nurturing business transformation and innovation into a customer-centric operation. In recent days, the business analysis moves beyond reporting and dash boarding where the contemporary skills and knowledge of the business offer critical analysis. Storyboarding as exploratory and explanatory analysis which aids to convey the business context, engagement with the audience, highlighting the accountability of deliverable by relating the data leading to data-driven decision making [3]. This analytics approach helps to predict the future and overcome any prevailing business challenges.

The trend towards online business is overwhelming due to the paradigm shift in the business, thus shows a rapid growth in recent years [1]. However, AWC was comparatively behind the line as they started selling online in the year 2010. Indeed, AWC claimed that internet sales has not been achieving the set target, thus the sales performance showed a declining trend, whereby the online business not turning prospects into strong leads [1]. This resulted in decreased sales and missed revenue targets over the years. The sales and marketing team of AWC worked to improve the marketing alignments and increase online sales by refining lead scoring while reaching the decision-makers faster.

The technology evolution has drastically changed the consumer purchasing behaviour, thus the purchasing is accomplished in few clicks [4]. Convenience and price comparison are the main drivers for the majority of shoppers, where they like to have a shopping before they make the final purchase decision in this digital era [5].

Revised Manuscript Received on March 15, 2020.

***** Correspondence Author

Yew Liong Lim, School of Computing, Asia Pacific University of Technology and Innovation, Kuala Lumpur, Malaysia. Email: lionglim11@hotmail.com

Raheem Mafas*, School of Computing, Asia Pacific University of Technology and Innovation, Kuala Lumpur, Malaysia. Email: raheem@staffemail.apu.edu.my

However, a lack of vital analysis is evident which can enable and support the online shopper's purchasing behaviour and reward repeat shoppers more effectively as AWC aspires to increase customer acquisition.

A. Aim & Objectives

This paper aims to seek ways to improve overall internet sales of the AWC via analysing the data and generate sales performance dashboards for better decision making. This was done by identifying the highest & lowest selling products while drilling down into different dimensions to understand the hidden patterns, thus drafted the following SMART objectives to achieve from this study.

1. To examine the prevailing situation of the internet sales of AWC via descriptive analytics methods.
2. To build the most effective predictive model with optimum accuracy and evaluation measures to predict future occurrences.
3. To draft valid recommendations to AWC by evaluating the customer buying behaviour to improve their future internet sales.

II. METHODS AND MATERIALS

This study is conducted via descriptive and predictive analytical methods by using the Cross-Industry Standards Process for Data Mining (CRISP-DM) methodology on the data acquired from AWC. CRISP-DM is an open data mining methodology which brings advantages to people to avoid common mistakes, thus affords a good kick start to address the business issues and find feasible solutions [6]. Further, CRISP-DM is a trending data mining methodology which supports to reveal useful information and actionable knowledge from data [7].

Finding a more suitable solution for a business problem by understanding the business objectives is the most important phase. A basic understanding of the application domain is also a necessity to ensure the process of the quality knowledge discovery. Therefore, developing and understanding sales performance analysis, customer and product analysis are crucial and included in the respective stages of the process [8]. The data familiarization and pre-processing is an important step to improve the efficiency of data mining to unlock the hidden patterns followed by implementing data modelling methods. This leads to draft valid recommendations via the right interpretation of the results which could achieve the business objectives and meet stakeholder's expectations.

A. Data Understanding

There are many different versions (from 2005 to 2017) of AWC databases available shared by Microsoft and its communities for educational and proof of concept purposes. This study used the data source namely AdventureWorksDW 2014 which enclosed with OLTP Database, DW Database, Tabular and Multidimensional Model Databases. The full database backups/datasets were obtained from Microsoft GitHub (URL:https://github.com/microsoft/sql-server-samples/releases).

The dataset contains 60398 AWC internet sales transaction records together with customer's demographic details from December 2010 to January 2014 shared across 37 attributes. Based on the descriptive statistics generated using the *StatExplore* node on the SAS Enterprise Miner platform, the Class Variables and Interval Variable showed no missing values, thus, no imputation operation was carried out on this dataset.

B. Data Analysis

Data analysis is an important process with the aim of discovering useful information to make timely decisions. In this study, descriptive and predictive analytics were carried out to analyze the data.

Descriptive Analytics

Descriptive Analytics also identified as an Exploratory Data Analysis (EDA) is a crucial approach in big data analytics to gain useful insights and to draft valid conclusions [9]. The objective 1 of this study was achieved via building interactive dashboards using Tableau which eventually helped to draft valid conclusions and recommendations for the decision-makers.

Predictive Analytics

The predictive analytics was carried out by a number of machine learning algorithms using the SAS Enterprise Miner platform. The predictive analytics process helped to identify the most significant decision variables for the business while building a more suitable predictive model. This process supported to achieve objective 2 while cross-checking with the results obtained from the descriptive analytics.

III. RESULTS

The results were presented accordingly since the analysis was conducted in descriptive and predictive analytics pathways. In order to achieve objective 1, the descriptive analytics results were structured and presented as follows.

A. Sales Performance analysis

Data visualization is a great eye-catcher that helps to understand trends, hidden patterns and to make correlations, thus widely accepted as a typical instrument for reasoning about quantitative information [9]. A well-designed dashboard poised with suitable infographics coupled with storytelling always delivers valuable insights to the business decision-makers.

Sales Performance by country

Referring to Fig. 1, a total of 15 thousand customers across North America, European and Pacific were recorded as the customer base where the more sales were obtained from the United States and Australia. The USA is the highest performing where more profitable sales were recorded from the northeast and southern part of the country. The least profitable sales were recorded in Canada, UK, Germany, and France. Overall 42% of profit margin was gained by internet sales channel, with 53 thousand accumulated online orders. The internet sales have been a very low overhead relative to direct sales, the margins are much higher [16].

Even though the USA contributes the highest sales, the profit margin is 2.5% lower than Australia, thus proposes to do priority sales based on the regions.

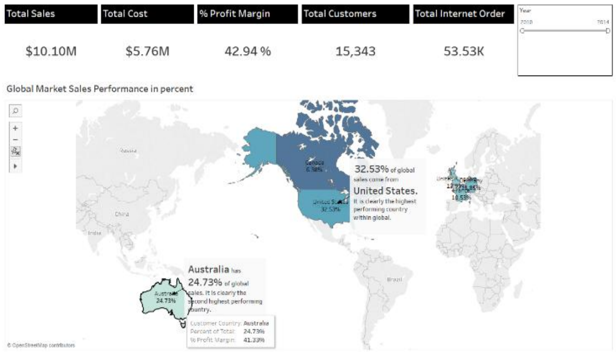

Fig. 1.Internet sales by country

In addition, AWC does offer free shipping, accessory packages, and occasional promotions. Price is the top reason that attracts internet customers purchasing AWC's bikes.

Sales Performance by promotion types

Plan an effective promotional campaigns would increase retail sales as it can provide more compelling sales in a timely manner [15]. As illustrated in Fig. 2, the "Volume Discount promotion" campaigns performed well and resonated better from the year 2011 to 2013 which increased AWC's short-term profits up to 68% by attracting both the existing and new customers. However, in January 2013 and 2014, internet sales dropped significantly whereas internet sales for accessories and clothing boosted up throughout the year 2013.

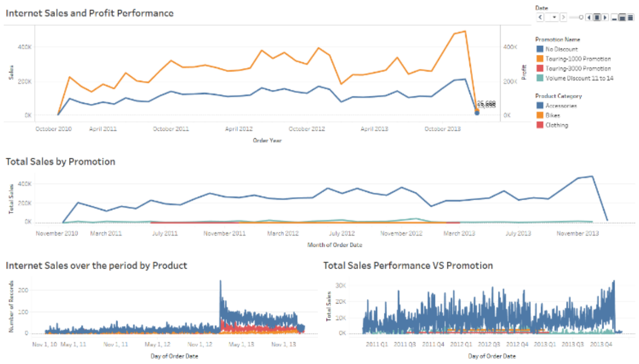

Fig. 2.Internet sales by promotion types

Sales Performance by product

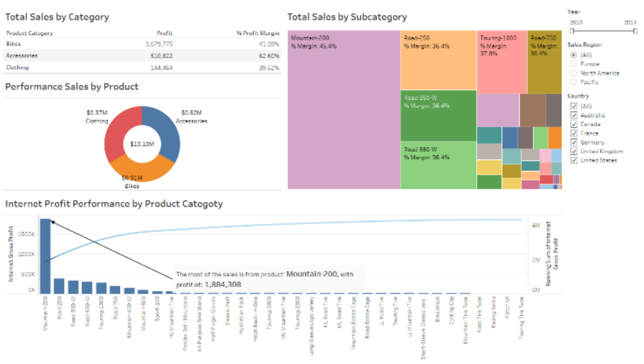

Fig. 3.Internet sales by product

As illustrated in Fig. 3, the Mountain-200 which generated a 45.4% profit margin followed by Road-250, Road-250-W, and Road-550-W with the profit margin of 36.4%. The USA and Australia have secured the highest profits as 1.1M and 0.9M respectively from Mountain-200. Touring Tire Tube made the least profit from the internet sales channel. HL Mountain Tire showed an association with the Mountain bikes identified from the customer purchasing pattern. The mountain tire brought an income of 62.6% of the gross profit margin in accessories category.

B. Overall Performance

The most profitable sales were recorded in December 2013, and the least profitable sales were recorded in January 2013 at the end of the promotional campaign buying season. The sales promotions allowed generating more revenue due to increased sales volume in the short run. To realize greater revenues, the sales promotions should attract more customers at very striking prices. Apparently the sales performance was low at the beginning of every year. Referring to Fig. 4, month over month ranking was done among different sales territory groups. In the month of December, Europe's sales rank order is number one with a sales amount of 0.44 M, while North America ranked number two with 0.52M. North America sales dropped from May (0.29M) to July (0.22M). This demonstrates that ranking for product categories change based on seasonality in different countries.

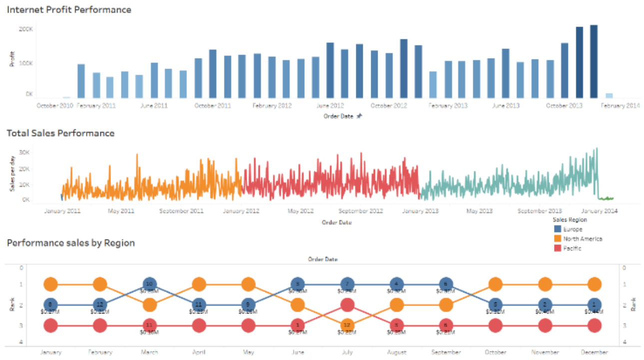

Fig. 4.Overall performance

C. Market basket analysis

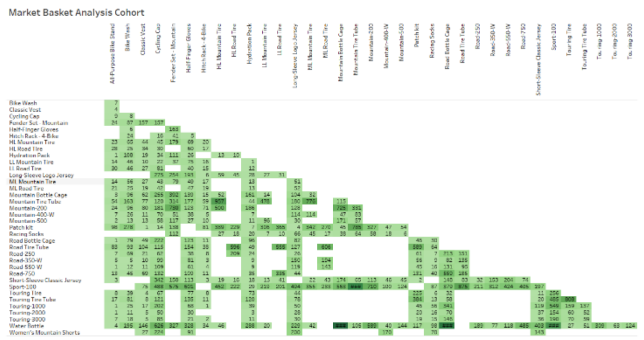

Fig. 5.Market basket analysis cohort

Market basket analysis is used to discover and understand customers' purchasing behavior [18]. Cross-selling is used to increase sales of existing customers by suggesting or recommending new products that may be of interest to the buyer. It is based on affinity analysis to define as the process of identifying a relationship that occurs together in a group. The output from market basket analysis helps in identifying which items should be placed together in an online website or can be a product recommendation when the Internet buyer chose a particular product [9].

Fig. 5 shows the co-occurrence product based on the sales records.

Customers who bought cycling cap generally bought a mountain tire tube. 1949 transactions sold during Volume discount promotion indicate people who bought cycling cap and mountain tire tube together.

D. Customer analysis

Referring to the Fig. 6, it summarizes the number of customers' order frequency who make one order, two orders, three orders and so forth. 13k out of 15k customers have made the only 1-time purchase which is equivalent of only 14% of customers has revisited in subsequent time to purchase. Also, it measures the percentages of the customer from each cohort (year of acquisition) who purchased at least 1, 2, 3, N times in a particular year. Nearly 90% of customers purchase at least 1 time in the year 2013, thus noted that the customers in more tenured cohorts purchase more frequently than the customer in recent cohorts. Also, it helps AWC to understand how well the regional marketing and sales are doing at generating new business. The figure shown that customer acquisition in Europe begins to flatten out in Feb 2013, some actions must be taken to increase lead flow [13].

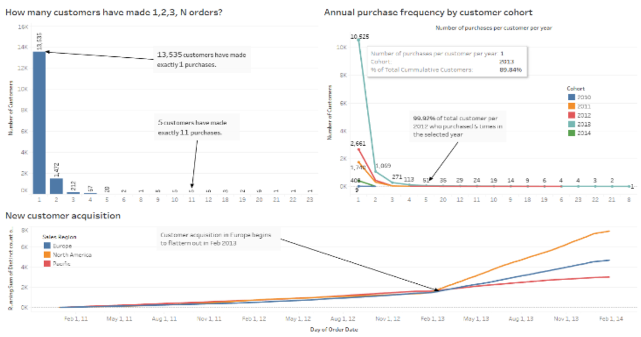

Fig. 6.Customer analysis

Purpose solutions

The RFM (Recency, Frequency, Monetary) analysis shown a proven marketing model based on customer buying behaviours and customer segmentation [11]. RFM explains the transaction history of the group of customers and identify which cluster of customers who likely to respond to promotions and also for future personalization [13]. To motivate the existing customer base to buy more often, 3 purpose solutions can help to increase overall sales.

Strategy 01:

Email marketing is always one of the effective marketing tools for increasing repeat purchase rate. Email in regular intervals to relevant customers to acknowledge them the new product, limited offers would be encouraging the customer to buy. Integration of a loyalty program on emails as an incentive to get customers to return would be an added value.

Strategy 02:

Improve customer engagement, reward the customer with coupon campaigns to support promotional marketing strategy are believed to a strong boost for ROI [12]. Offering a welcome coupon to a new customer would encourage them to spend more on their first purchase which would also offer a memorable first purchasing experience. Flash sales over the weekend, free shipping for specific time period and a new product coupon can drive the interest of customers to make more purchase. Consequently, compensating the customer with a compelling business offers will get some positive review and rating for easy social sharing [10].

Strategy 03:

Bundle-up promotions of two or more complementary products would encourage the customers to buy more goods. Bundle-up promotions positively support the new product items which indirectly increase the entire customer experience and the overall revenue.

E. Predictive Modelling

Building a predictive model to predict the future occurrences is very crucial for any business in this competitive data driven business world. Customer demographics and geographic details, sales related details were used to build a predictive model to predict whether customers tend to purchase a product or not. In this line, several predictive models were built using the SAS Enterprise Miner tool for AWC and evaluated using the accuracy rate along with other evaluation measures to choose the best model for future use. After analysing the data, 14 input variables were recognised to build a predictive model with a binary target variable "*IsBikeBuyer*". This model could be applied to a list of potential customers to predict the probability of responding to a targeted mailing activity that sells Adventure Works bikes.

According to the chi-square plot with "*IsBikeBuyer*" as a target variable, the plot highlights that variable like *SalesRegion* has the highest chi-square among all the variables, indicating that it has the strongest association with the target variable followed by *Education*, *CommuteDistance* and so on. On the other hand, *HouseOwnerFlag* has the lowest chi-square value among all the variables, indicating that it has the weakest association with the target variable. The dataset was split into training (50%) and validation (50%) to build different predictive models.

Predictive Model I - High Performance (HP) Tree

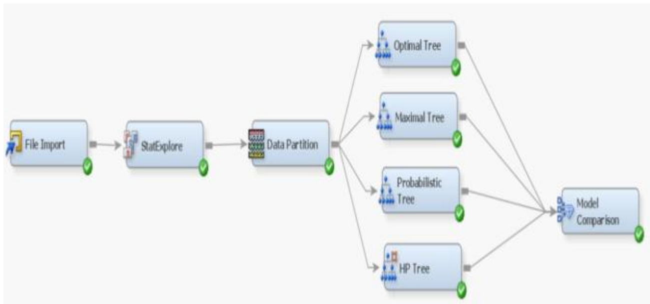

Fig. 7.Process Flow diagram - Decision tree

A decision tree is a supervised learning used in both classification and regression problems. It is a tree where each node denotes an attribute, branch denotes a decision and each leaf denotes an outcome [11]. A tree is created for whole data and process a single outcome at every leaf node. Fig. 7 shows the process flow diagram of different decision tree models such as optimal tree, probabilistic tree, maximal tree and HP tree built in SAS Enterprise Miner.

Based on the different decision tree models, HP tree gives the better results in terms of the tree map and cumulative lift values.

According to the fit statistic shown in the Fig. 8, the HP tree model is the better model based on the validation misclassification values (0.22844). The receiver operating characteristic (ROC) plots the trade-off between sensitivity and the specificity across all selected fraction of data, thus the index (0.851) value supports the selection of the suitable model.

Fit Statistics

Model Selection based on Valid: Misclassification Rate (_VMISC_)

Selected Model	Model Node	Model Description	Valid: Misclassification Rate	Train: Average Squared Error	Train: Misclassification Rate	Valid: Average Squared Error
Y	HPTree	HP Tree	0.22844	0.14531	0.21349	0.15660
	Tree4	Probabilistic Tree	0.29311	0.18924	0.28714	0.19355
	Tree2	Optimal Tree	0.29311	0.19220	0.28714	0.19621
	Tree3	Maximal Tree	0.29311	0.19220	0.28714	0.19621

Fig. 8.Fit statistics of decision tree models

Predictive Model II - High Performance (HP) Bayesian Network (BN)

Bayesian network is a probability graphical model that denotes a set of variables and their conditional dependencies using a Directed Acyclic Graph (DAG). In the BN model, each node represents a random variable, whereas the edge between the nodes represents probabilistic dependency among the corresponding random variables. The dependencies are assessed by using statistical and computational methods. There are 3 types of independence test statistics used in BN classifier model building such as G-square, chi-square as well as Chi & G-square. As shown in Fig. 9, all three of them were built and evaluated to find the most suitable BN classifier for this problem.

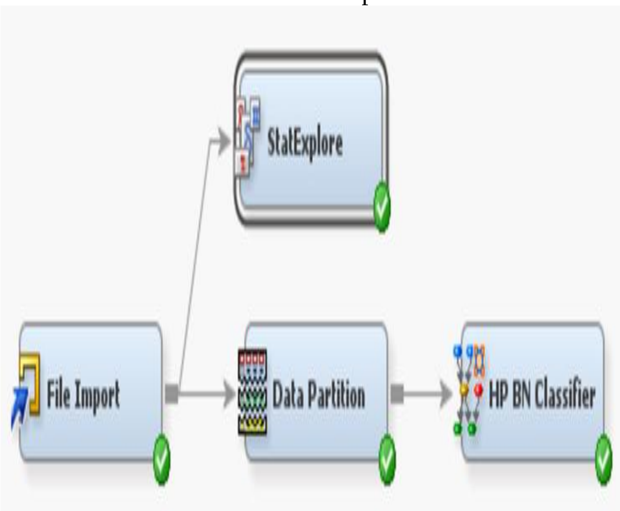

Fig. 9.Process Flow diagram - HP BN classifier

Referring to Fig. 10, HP BN classifier using Chi-square or Chi-G-square and referring to Fig. 11, G-square have the same conditional dependencies to 12 variables listed as *YearlyIncome, NumberChildrenAtHome, Education, NumberCarsOwned, SalesRegion, Occupation, CommuteDistance, HouseOwnerFlag, TotalChildren, MaritalStatus, Age, and Gender*.

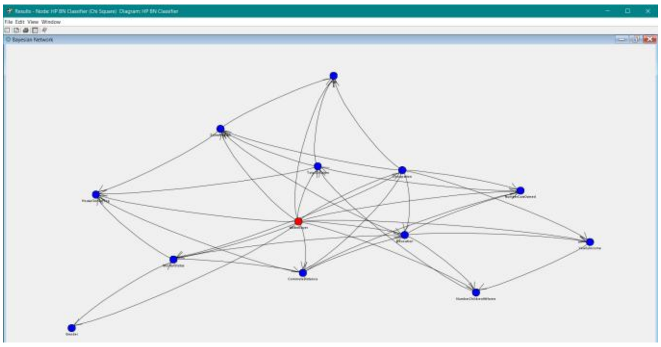

Fig. 10.HP BN classifier by Chi-square/Chi-G-square

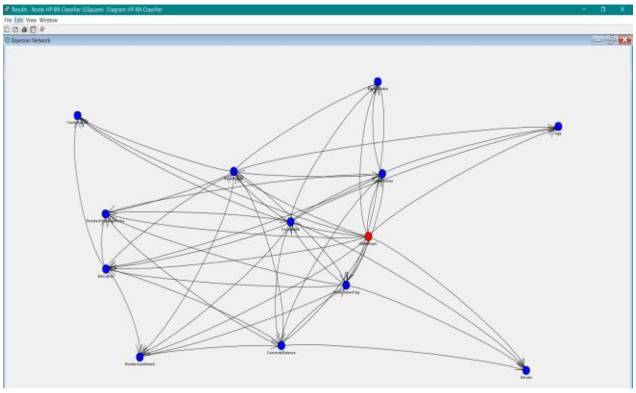

Fig. 11.HP BN classifier by G-square

Consequently, the cumulative lift for both training and validation data for all the 3 types of HP BN classifiers shows the same curve. Further, the fit statistic shown in Fig. 12, proves that the G-square HP BN classifier obtained lower validation misclassification rates (0.23811) among all the 3 types of HP BN classifiers. As a conclusion, G-square HP BN classifier is more suitable to this classification problem.

Fit Statistics

Model Selection based on Valid: Misclassification Rate (_VMISC_)

Selected Model	Model Node	Model Description	Valid: Misclassification Rate	Train: Average Squared Error	Train: Misclassification Rate	Valid: Average Squared Error
Y	HPBNC2	HP BN Classifier (GSquare)	0.23811	0.16273	0.22028	0.17668
	HPBNC	HP BN Classifier (Chi Square)	0.26255	0.17131	0.24637	0.18194
	HPBNC3	HP BN Classifier (Chi GSquare)	0.26255	0.17131	0.24637	0.18194

Fig. 12.Fit statistics of HP BN classifier

Predictive Model III - High Performance (HP) Forest

Random forest is a supervised machine learning algorithm built by merging multiple number of decision trees together to get a more accurate, precise and stable prediction. It is a wide and diverse model that gives extra randomness to the model by searching the best feature among a random subset of features. There are 2 types the sample options such as proportion and count are used to build the model and evaluated to get the most suitable random forest model. Besides that, the maximum number of trees is increased to 150 (default = 100) to improve the accuracy of the model.

The Fig. 13 shows the misclassification rate of all the random forests models.

The validation misclassification rate for the HP forests with 150 trees is 0.19894, while the 100 trees are 0.19980. Therefore, the HP forests with 150 trees regardless of the type of the random forest (proportion or count), are more accurate.

Fit Statistics
Model Selection based on Valid: Misclassification Rate (_VMISC_)

Selected Model	Model Node	Model Description	Valid: Misclassification Rate	Train: Average Squared Error	Train: Misclassification Rate	Valid: Average Squared Error
Y	HPDMForest3	Proportion 150	0.19894	0.14435	0.18114	0.15148
	HPDMForest4	Count 150	0.19894	0.14435	0.18114	0.15148
	HPDMForest	Count 100	0.19980	0.14423	0.18190	0.15140
	HPDMForest2	Proportion 100	0.19980	0.14423	0.18190	0.15140

Fig. 13.Fit statistics of HP Forest

F. Model Selection

As shown in Fig. 14, the different models such as HP Tree, HP BN Classifier, HP Forest and HP Neural were compared and the most suitable predictive model was selected for this problem. The model selection was done based on the lowest validation misclassification rates, thus the best model gives the highest accuracy.

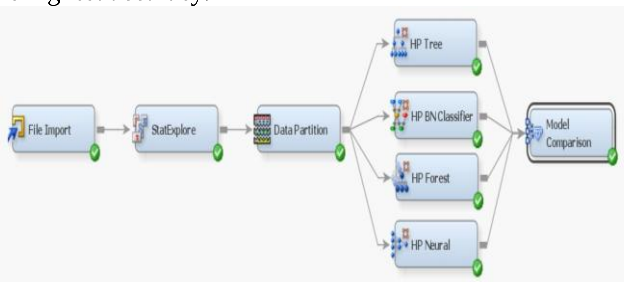

Fig. 14.Final model selection

The Fig. 15 shows the cumulative lift for training and validation data for HP Tree, HP BN Classifier HP Forest and HP Neural. Based on the results, HP Forest has the highest cumulative lift, followed by HP Tree, HP BN Classifier and the HP Neural which has the lowest cumulative lift in different depth.

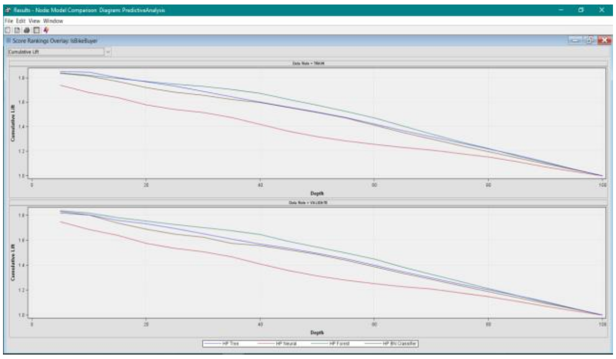

Fig. 15.Cumulative lift for predictive models

According to Fig. 16, the validation misclassification rate for the HP forest is 0.19894, the HP Tree is 0.22844, the HP BN Classifier is 0.23811 and the HP Neural is 0.33222. The lowest validation misclassification rate was scored by HP Forest, thus can be concluded as HP Forest is the most-fit model for this classification problem with the accuracy of 80%.

Fit Statistics
Model Selection based on Valid: Misclassification Rate (_VMISC_)

Selected Model	Model Node	Model Description	Valid: Misclassification Rate	Train: Average Squared Error	Train: Misclassification Rate	Valid: Average Squared Error
Y	HPDMForest	HP Forest	0.19894	0.14435	0.18114	0.15148
	HPTree	HP Tree	0.22844	0.14531	0.21349	0.15660
	HPBNC	HP BN Classifier	0.23811	0.16273	0.22028	0.17668
	HPNNA	HP Neural	0.33222	0.19836	0.32741	0.20040

Fig. 16.Fit statistics for final model selection

The comparison between the prediction models is interpreted using three indicators, the Receiver Operating Characteristics (ROC) index, then Cumulative life chart and the Misclassification rate of the validation data.

Table- I: Model Evaluation/Fit Statistics

Model	Valid Misclassification Rate	Accuracy (%)
HP Forest	0.19894	80.11
HP Tree	0.22844	77.16
HP BN Classifier	0.23811	76.19
HP Neural	0.33222	66.78

According to the Fig. 16 and Table- I, The HP Forest algorithm performs the best of the four (4) algorithms as it has the lowest validation misclassification rate as 0.198 (Fig. 16), thus obtains the highest accuracy (80%). It also obtains the area under the curve with ROC index 0.878 (Fig. 17) and is closer to sensitivity of 1 along the curve. It has the highest cumulative lift of 1.82 (Fig. 15), and consistently the highest value along the chart.

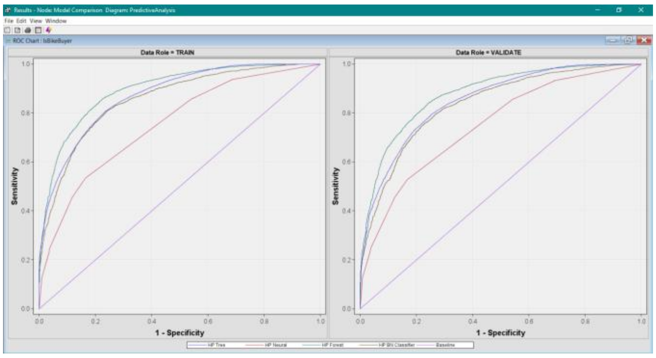

Fig. 17.ROC chart for final model selection

IV. RECOMMENDATIONS

Un-target marketing campaigns can be expensive. Thus, the retailers and e-commerce businesses can focus on a smaller and specific group of customers and significantly minimize marketing costs and improve return on investment. Target only those customers that would likely respond to these campaigns. Based on actual transaction history to understand customer's behaviour which can reveal insights about how the business's and customer's lifetime value can open a huge opportunity of growth. Get to know different customer segments and identify the best customer can improve customer loyalty and get detailed customer analytics and customer insights.

Different customer segments may recommend different marketing actions to make an effective outcome. This is because different customers react and respond to different messaging. Discounts and offers will only grab the attention of price-sensitive customers, but those regular customers will get excited only to a new product launch. Understand customers better and send them on more relevant campaigns will guarantee a higher success rate and indirectly get in touch with them and make them happier and stay loyal to your products. Since customer and purchase records are digitized, in order to quantify customer behaviour, the existing model can extend it to different models. For instance: RMF (Recency, Monetary, Frequency), RFD (Recency, Frequency, Duration), and RFE (Recency, Frequency, Engagement). RFM studies the groups of customers based on their transaction history. For instance: how recently, how often and how much did a customer perform a purchase [14]. RFD is analysing duration time spent for a customer on surfing-oriented products, while RFE is a composite value model based on customers' time spent on a page, pages per visit, bounce rate or social media engagement.

Do not go overboard by keep sending marketing campaigns to one segment. As a result, customers might get irritated and stop buying.

V. CONCLUSION

Descriptive analytics via a consistent interpretation offers wide range of actionable knowledge to businesses where it brings competitive edge. This actionable knowledge needs more support from the predictive analytics to formulate solid business strategies to win current the competitive business environment. Predictive analytics allows businesses to foresee the future events to effectively implement the formulated / designed actions plans to enjoy competitive advantages and gain more profits.

In this line, AWC as a data-driven organization summarized the past events for instance: customer lead score, customer attrition, sales performance, or success of marketing campaign and analysed to get better business insights to increase overall profitability, make optimization and improve their decision-making power more effectively [20]. The descriptive analytics revealed lots of useful insights on the sales performance, market basket analysis and the customer analysis which enlighten ways for AWC to understand the current situation.

Consequently, effective and suitable predictive models such as HP tree, HP forest, HP Neural and HP BN classifiers are built and evaluated to predict the future occurrences from which the AWC can plan accordingly. The predictive models are evaluated and compared using the measures such as accuracy, ROC index and cumulative lift values to get the best fit model. In the evaluation process, the HP forest has been identified as the best-fit model for this problem with the accuracy of 80%, ROC index 0.878 and the cumulative lift value of 1.82.

In conclusion, the performed predictive analytics results can support to identify the potential customers and predict the most likely response rate to a targeted mailing activity that promotes AWC sales to a greater level.

REFERENCES

1. M. Mitri, "Teaching Tip Active Learning via a Sample Database: The Case of Microsoft's Adventure Works," Journal of Information Systems Education, vol. 26(3), 2015.
2. "A Guide to the Business Analysis Body of Knowledge: Draft Material for Review and Feedback," International Institute of Business Analysis, 2006
3. Knaflic C., "Storytelling With Data: A Data Visualization Guide for Business Professionals", Wiley, 2015.
4. D. M. Velumani, "A Study on Consumer Buying Behaviour towards Nokia Mobile in Erode District," IOSR Journal of Business and Management (IOSR-JBM), Vol. 16, No. 12, pp. 52-54, 2014.
5. D. L. I. A. M. Asoc. Prof. Elīna Gaile-Sarkane, "Consumer behavior changing: methods of evaluation," Trends of Economics and Management, Vol. 3, No.4, 2013
6. Shafique U., Qaiser H., "A Comparative Study of Data Mining Process Models (KDD, CRISP-DM and SEMMA)," International Journal of Innovation and Scientific Research, Vol. 12, No. 1, pp. 217-222, 2014.
7. J. H. Rüdiger Wirth, "CRISP-DM: Towards a Standard Process Model for Data Mining", IFIP International Conference on Digital Forensics, 2017
8. Vente J, Waal A.D, Willers C., "Specializing CRISP-DM for Evidence Mining," SpringerLink.
9. M. Faber, "Visualizing Student Research though a Multi-Conditional Branching - Choose Your Own Adventure-Style Tutorial," Semantic Scholar, 2015. Accessed on Jun 19, 2019. [Online]. Available: https://www.semanticscholar.org//paper/Visualizing-Student-Research-through-a-Branching-%22-Faber/10684e7187e06ccadb11e185ab1d58c8cf90a35e
10. Kurniawan F., Hammad J., Nugroho S.M, Hariadi M., "Market Basket Analysis to Identify Customer Behaviors by Way of Transaction Data," Knowledge Engineering and Data Science (KEDS), Vol. 1, No.1, pp.20-25, 2018.
11. Chris L., "Good to Great: Quadrant Analysis", 11 Dec, 2015. Accessed on Jun 21, 2019. [Online]. Available: https://www.theinformationlab.co.uk/2015/12/11/tableau-good-great-quadrant-analysis/
12. B. Lyons, "Top 15 LOD Expressions", Apr 6, 2015. Accessed on: Jun 23, 2019. Accessed on Jun 21, 2019. [Online]. Available: https://www.tableau.com/about/blog/LOD-expressions
13. Khajvand M., Zolfaghar K., Ashoori S., Alizadeh S., "Estimating customer lifetime value based on RFM analysis of customer purchase behavior: case study," Procedia Computer Science, 2010.
14. Christy A.J., Umamakeswari A., Priyatharsini L., Neyaa A., "RFM ranking – An effective approach to customer segmentation," Journal of King Saud University – Computer and Information Sciences, 2018.
15. Adesoga D.A., "Promotional Strategy Impacts on Organizational Market Share and Profitability," ACTA Universitatis Danubius, Vol. 11, No.6, 2015.
16. Aftab. A, Kamisan G., "The Role of Promotion Strategies in Personal Selling", Far East Journal of Psychology and Business, Vol.12, No.3, 2013.
17. Kattamuri S. Sarma, Predictive Modeling with SAS Enterprise Miner: Pratical Solutions for Business Applications, 2th ed., Cary, North Carolina, USA, SAS Institute Inc, 2010.
18. Nidhi M.K, Dr. Snehil. D., "Market Basket Analysis using Apriori Algorithm in R Language," International Journal of Trend in Scientific Research and Development (IJTSRD), Vol. 2, No. 4, 2018
19. Monsen, Karen A., Intervention Effectiveness Research: Quality Improvement and Program Evaluation, Springer International Publishing AG, 2018
20. Banphot V., Antonis C.S., Magid I., "How to retain online customers", Communications of the ACM, Vol. 47, No.6, 2004

AUTHORS PROFILE

Yew Liong Lim is graduated in the stream of Bachelor of Information System Engineering from the University Tunku Abdul Rahman and currently pursuing a Master degree in Data Science at the Asia Pacific University. Yew Liong Lim currently works as a Data Specialist in automotive industry. Yew does research in customer's behavior analysis, Data Science and Analytics.

Raheem Mafas is an experienced full time lecturer specialized in the field of Data Science & Business Analytics and holding a Master of Business Administration and MSc in Data Science & Business Analytics. The main research domain is e-commerce and interestingly dealing with business intelligence, visualisations, text analytics and sentiment analysis though published number of journal articles.