AN APPLICATION OF DATA MINING TECHNIQUE ON SOCIAL MEDIA ADVERTISING: A FIELD STUDY

Yu-Chin Lee, Department of Distribution Management, Shu-Te University, No.59, Hengshan Rd., Yanchao Dist., Kaohsiung City 82445, Taiwan, +886-7-615-8000#4516, yclee@stu.edu.tw

Jinsheng Roan, Department of Information Management, National Chung Cheng University, No. 168, Section 1, University Road, Minxiong Township, Chiayi County 621301, Taiwan, +886-5-2720411#34615, bmajsr@ccu.edu.tw

Chun-Nan Lin, Department of Information Science and Management Systems, National Taitung University, 369, Sec. 2, University Rd., Taitung, Taiwan, +886-89-517668, cnlin@gm.nttu.edu.tw

ABSTRACT

Retailers launch their marketing campaigns on social media platforms, hoping to yield quick results and expand the geographical region of customers. However, the linkage between social media advertising and sales response remains debatable. This study uses data mining techniques to derive a decision model for a fashion shoe company to set out its social media advertising campaign. The findings yield four types of decision models with over 82% accuracy and provide the company with practical suggestions. The follow-up investigation suggests effective exposure to social media advertising leads to increased conversions and sales. Suggestions for future research are discussed.

Keywords: Data mining, Decision tree, Social media advertising effectiveness.

INTRODUCTION

The retail industry is undergoing a drastic change since Jack Ma, the founder and executive chairman of the Alibaba Group, proposed a new concept of "New Retail" in 2016. The concept revolves around the true integration between online and offline retailing channels (Wang and Ng, 2020). Moreover, this new model requires the input of the modern logistic system and innovative technologies such as big data and cloud computing. In new retail, brick-and-mortar stores, e-commerce, and big data analysis are essential components. Information and data on physical merchandise, membership, transactions, and marketing methods are interwoven to achieve online and offline integration. Through data analysis, companies will be able to identify customers who are most valuable and worthy of investing resources and provide customized services. In addition, the recovery of lost customers is also made possible. Scrutinizing data at hand and pinpointing the precise problems that caused the dissatisfaction, listening to customer opinions and persuading them to re-experience rectified products and services, and finally, seizing the opportunity to win back lost customers.

The success of the new retail business model lies in the significant amount of information collected from customers, i.e., big data. Big data concerns large-volume, complex, growing data sets with multiple, autonomous sources (Wu *et al.,* 2014). Discussions about big data are nonstop and fierce in many fields in this era of information explosion. In particular, big data analysis has become the center of discussion in retailing management, e.g., how to make better decisions based on the data at hand, how to use the data to predict future trends and increase sales performance or reduce the processing time required for supply chain activities (Fisher and Raman, 2018; Madsen *et al.,* 2020).

This study investigates a sole agent of a Korean fashion shoe brand in Taiwan. This company's business operation has expanded from an online platform to include three physical stores and four pop-up stores across Taiwan. Other than its attempt at marching toward the new retail paradigm, the company has also taken interest in various new marketing strategies and promotional activities.

This study has developed a sales decision model by applying the data-mining technique to examine this fashion shoe company's physical store sales figures and customer database. The results enable a better understanding of the customer characteristics and preferences and offer suggestions for the configuration of the future target audience of their marketing campaign. The results can also be applied to other industries and offer a good testimony of the success of new retail.

RESEARCH BACKGROUND

Data Mining

Data mining is the process of discovering hidden information from a massive database, such as trends, patterns, and relationships (Kaur and Wasan, 2006). It is also known as knowledge discovery in databases (KDD), data archaeology, data pattern processing, information discovery, and information harvesting (Fayyad *et al.,* 1996). Modern enterprises generate massive amounts of data daily, such as information about the market, customers, suppliers, competitors, and future trends. Through the application of data mining techniques, companies can extract meaningful information and knowledge from the massive database for the use of decision-making and building competitive advantage (Bose and Mahapatra, 2001; Chan and Lewis, 2002; Daskalaki *et al.*, 2003).

Data mining generally follows these steps: first, a comprehensive understanding of the data and the task at hand; second, an acquisition of the relevant knowledge and techniques; third, an integration and checkup of the data; fourth, removal of incorrect or inconsistent data, i.e., data cleaning; fifth, development of models and hypothesis; sixth, digging into the data; seventh, testing and verify the data; and lastly, interpretation and the application of data (Kurgan and Musilek, 2006).

Researchers may decide which of the following functions to be performed: classification, estimation, prediction, association, and clustering (Chou *et al.,* 2004; Mittal *et al.,* 2019; Ngai *et al.,* 2009; Rodrigues *et al.,* 2018; Yang and Li, 2018). Classification is done by defining and sorting data into predetermined groups. Classification methods include decision trees, memory-based reasoning, and link analysis. Estimation is performed through the approximation of unknown values. Common tools used for estimation include correlation analysis, regression analysis, and neural networks. Prediction is derived by making inferences about future trends or possible outcomes based on the data at hand. Techniques used include regression analysis, time series analysis, and neural networks. Association involves the discovery of items that are frequently associated together, i.e., the discovery of frequent patterns (Chiclana *et al.,* 2018; Djenouri *et al.,* 2018; Ting *et al.,* 2018). For instance, Walmart discovered a positive association between sales of diapers and sales of beer on Fridays (Peng *et al.,* 2020). Market basket analysis is one example of an early attempt to explore frequent patterns in association rules, it helps corporate decision-makers to make choices on cross-selling, catalog design, and consumer behavior analysis (Chen *et al.,* 2012). Clustering sorts data into several categorical groups (or clusters). It is different from the classification function in that the groups are determined from the data. In other words, the clusters are defined naturally based on the characteristics of the data. Therefore, the process is also called unsupervised learning. The techniques include the *k*-means method, the two-step approach, and discriminant analysis.

Social Media Advertising

Social media is defined as a group of internet-based applications that enables the creation and dissemination of user-generated content (Constantinides, 2014; Kaplan & Haenlein, 2010). The fast-paced development in information and communication technology has generated different kinds of social media platforms, including blogs, forums, bulletin boards, and social networks (Constantinides, 2014). These platforms have their unique ways or rules of delivering user-generated content, e.g., YouTube is known for users to broadcast their video-based content, and each offers a different interface to attract their intended users (Voorveld, van Noort, Muntinga, & Bronner, 2018).

The latest statistics show that the top three social media networks used by most users in the world are Facebook, YouTube, and WhatsApp (Statista, 2021). Other than the differences in their user composition, past research has also distinguished these platforms by different characteristics. For instance, Zhu and Chen (2015) developed a social media matrix that categorizes social media platforms into four quadrants: relationship, self-media, creative outlet, and collaboration, based on the characteristics of the nature of connection and the level of customization of messages. Marketers may apply different social media strategies according to designated platform types (Zhu & Chen, 2015). Furthermore, the number of active social media users has skyrocketed by 4.2 million, a 13.2% increase from 2020 (We Are Social, 2021). Daily time spent on social media was an average of 2.4 hours for users aged 16 to 64 (We Are Social, 2021). Due to the heavy usage of social media, marketers must divert their attention and resources to marketing practices in this domain (Constantinides, 2014).

Digital Advertising Effectiveness

Due to the difficulty in obtaining actual purchasing data, researchers often measure advertising effectiveness with some proximal variables, e.g., attitude toward the brand or ad (e.g., Deshpandé & Stayman, 1994; Lee, Lee, & Yang, 2017), brand awareness (Macdonald & Sharp, 2003; Rubinson, 2009), and purchase intention (Stewart, Kammer-Kerwick, Koh, & Cunningham, 2018; Yim, Kim, & Lee, 2020). These variables are traditionally believed to be predictive measures of advertising effectiveness (Lavidge & Steiner, 1961).

However, the scope and customer reach of social media platforms are much more complex than those of traditional media. The measure of digital advertising effectiveness takes several forms, e.g., level of media engagement (Voorveld *et al.*, 2018), ad intrusiveness (Belanche, Cenjor, & Pérez-Rueda, 2019), or user engagement metrics provided by social media platforms (Yousef, Dietrich, & Rundle-Thiele, 2021). While these measures provide insight into the understanding of consumers' attitudes toward specific online ads or social media platforms, they do not reflect the real actions taken by consumers after being exposed to these ads. Therefore, our study intends to add value to this line of research by adopting data mining techniques to discover the effect of social media advertising on consumers' eventual purchasing behavior.

RESEARCH METHODS

This study follows the KDD process (Fayyad *et al.*, 1996) to classify and cluster our data. The KDD procedure comprises the following process: first, loading the collected data into a database; second, data cleaning e.g., deleting duplicate and missing data; third, data exploration by using statistical inferences and correlation analysis; fourth, data cross-validation; fifth, interpretation of results and extract useful information; and lastly, apply the information to develop relevant business strategies.

Data Collection

The data used in this study comprises transactional and membership data from three physical stores. Customer data was given voluntarily by seeking consent from the customers. A total of 2,939 transactional data was collected from customers. However, after deleting incomplete and duplicate data, we had 545 valid data (18.54% response rate). The membership and transactional information included in the database were birthdays of customers, residential addresses, places of purchase, purchased items, amount of purchase, sources of brand knowledge and recommendation, etc.

Data Preprocessing

As mentioned, our database includes customer transactional records and membership information. The massive amount of data may contain erroneous, missing, or inconsistent data. Hence, we cleaned up the raw data before analysis. We integrated the data from three stores and encoded the data into five categories: region, birthdate, age, source of information, and items of purchase. Moreover, we transformed the customer region into 18 areas, birthdates into 12 star signs, age into eight categories in five-year intervals, and source of information into 19 categories. This way of organizing data makes data mining easier and more efficient. After data preprocessing, we then performed data analysis, developed the training model, made predictions, and interpreted results.

Data Mining

This study utilizes the clustering and classification functions of the data mining technique to search for meaningful patterns in our transactional data. WEKA software was applied as our analytical tool in this study. It is implemented in the JAVA programming language and is used for running machine learning and data mining tasks (Tseng *et al.*, 2015).

Cluster analysis, or clustering, is used to group highly similar data together according to their characteristics. The grouping is done without any predetermined conditions. This study adopted the *k*-means clustering algorithm to find the respective data points in a large dataset, it is one of the most popular algorithms used by researchers (Jarboui *et al.*, 2007). These data points are also called cluster centers, prototypes, codewords, etc. After finding these cluster centers, we then continued with data compression and processing.

Classification analysis, on the other hand, uses predetermined categories to predict behaviors (Chi-Hsien and Nagasawa, 2019; Exenberger and Bucko, 2020; Yin *et al.*, 2018). The procedure begins with the evaluation of transaction records and scrutinization of the linkage between customer profiles and consumed goods. Finally, by adopting the decision tree or decision rules, we identify the most suitable classifiers (Cho *et al.*, 2002; Kim *et al.*, 2002; Yao and Xiong, 2011). We applied the C4.5 algorithm to develop decision trees for the purpose of data classification in this study.

Validation Method

We used cross-validation analysis and clustering algorithms to identify customer groups and attributes. The result contributes to the decision of audience settings in social media advertising on Facebook. We also used a decision tree algorithm to develop a sales decision model for the sales floor clerk or marketing staff to consider when selling.

RESULTS

Descriptive statistics

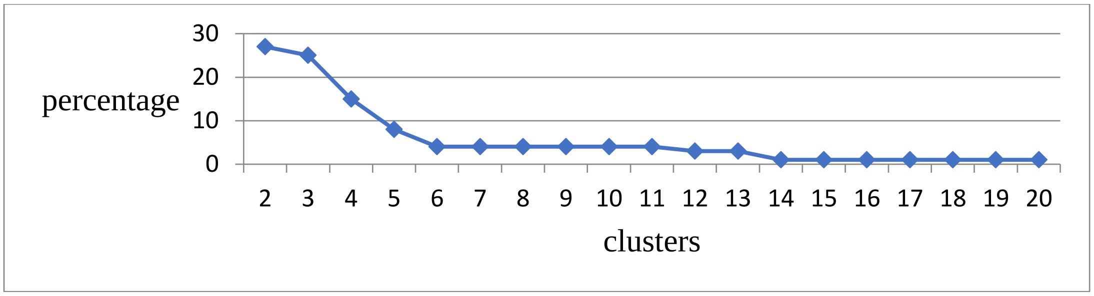
Our data was collected from three business trading areas (A, B, and C) of Company X, including 2,939 transaction records from June to November 2019. After deleting incomplete data, the final number of valid records comes to 545.

Cross Analysis

To understand the customer preferences in Areas A, B, and C, we analyzed the city (place of residence), color, and source of information based on 545 standardized data. Most of the customers in Area A came from New Taipei City (northern Taiwan); most of the customers in Area B came from Tainan City (southern Taiwan); most of the customers in Area C came from Kaohsiung City (southern Taiwan). The most popular colors in Area A, B, and C were green, black, and green, respectively. Customers acquired brand knowledge information mostly from Facebook in Area A and heard from a friend or a colleague in Area B and C.

Cluster Analysis

Our data was coded in terms of City (place of residence), Star Sign, Age, Item(s) of Purchase, Color, and Source of Information. As we can see from Figure 1, when the number of clusters reached five clusters, the percentage dropped below 10% (unrepresentative). As a result, this dataset is determined to have four clusters.



Note: X-axis = number of clusters; Y-axis = percentage of data points left for the group with minimal amount of data.

Figure 1. Comparison of clusters with the lowest number of data in each cluster

As Table 1 shows, there are 188 data in the first cluster (34%), 114 data in the second cluster (21%), 162 data in the third cluster (30%), and 81 data in the fourth cluster (15%). Characteristics of each cluster are described as follows: (a) consumers in the first cluster are from Taipei City, Libra, under 20 years old, purchased the BOLT series, green is their favorite color, and heard about the brand from a friend or a colleague; (b) Consumers in the second cluster are from Tainan City, Sagittarius, aged 26 to 30, purchased the BOLT series, green is their favorite color, and heard about the brand from a friend or a colleague; (c) Consumers in the third cluster are from Kaohsiung City, Gemini, aged 21 to 25, purchased the BOLT series, green is their favorite color, and Instagram is their source of information; and (d) Consumers in the fourth cluster are from New Taipei City, Aries, aged 21 to 25, purchased the BOLT series, yellow is their favorite color, and they heard about the brand from a friend or a colleague.

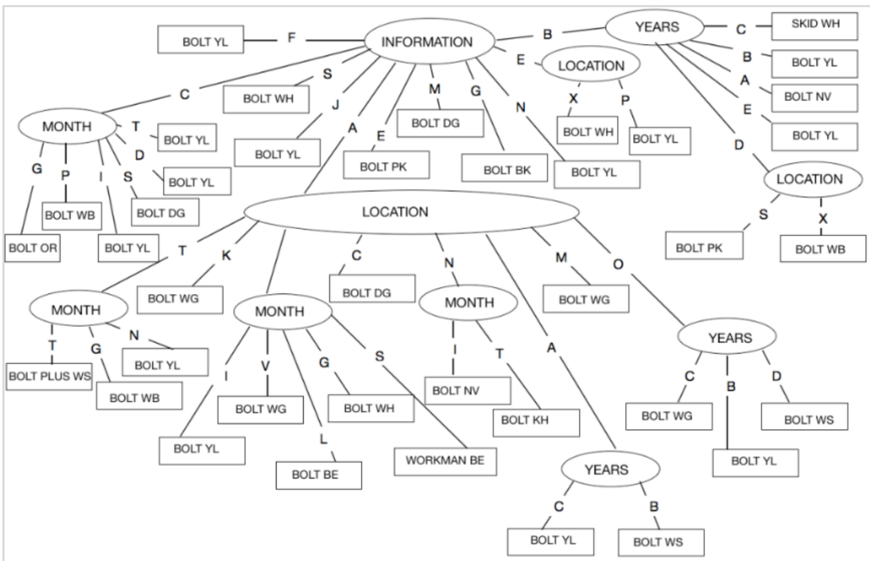
Table 1. Results of clustering analysis

Attribute	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Record(percentage)	188(34%)	114(21%)	162(30%)	81(15%)
City	New Taipei City	Tainan	Kaohsiung	Taipei
Star Sign	Libra	Sagittarius	Gemini	Aries
Age	Under 20	26 to 30	21 to 25	21 to 25
Item(s) of Purchase	BOLT	BOLT	BOLT	BOLT
Color	Green	Green	Green	Yellow
Source of Information	Friend	Friend	Instagram	Friend

Decision model development

WEKA was used as the analytical tool to derive four clusters. We then developed a sales decision model by using a decision tree algorithm based on these 4 clusters. The total amount of data in the first category was 196 with five attributes. As shown in Figure 2, after 10-fold cross-validation, the total rate of accuracy reached 89.29%, the percentage of total error was 10.71%, Kappa statistics 0.864, the true positive rate was 0.893, the false positive rate was 0.023, the true value and model precision rate was 0.889, the recall rate was 0.893, *F*-measure was 0.882, and receiver operating characteristic curve was 0.968.

The following decision rules were derived from the model: (a) When the customer is a Scorpio, and his or her source of information is Instagram, this customer's preference is BOLT DG; (b) Customers aged 31 to 35 from Hsinchu (a city), and source of information came from a friend or colleague, then they are more likely to purchase BOLT PK; and (c) A Taurus from Tainan City and received information from the Facebook, then he or she is more likely to purchase BOLT KH.

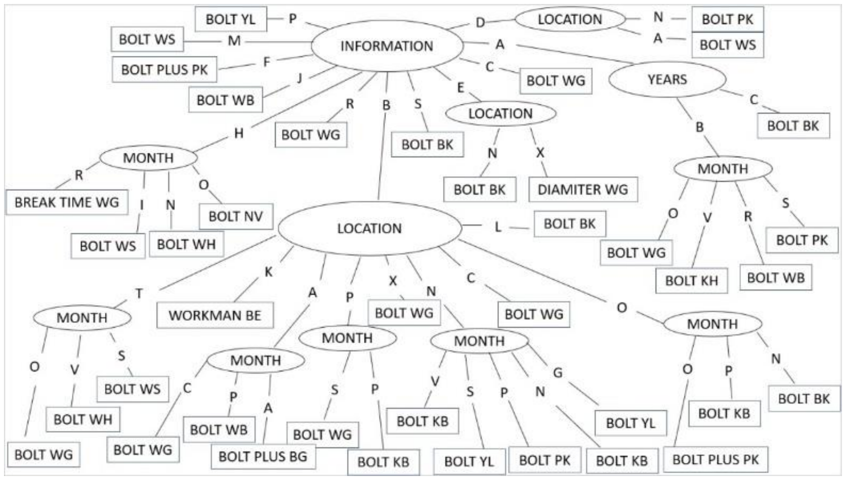


Note: Information = source of information; Location = city; Month = star sign; Years = age.

Figure 2. Decision model based on the first cluster

The total amount of data in the second category is 166, with five attributes. As shown in Figure 3, after 10-fold cross-validation, the total accuracy rate reached 82.53%, the percentage of total error was 17.47%, Kappa statistics was 0.8032, the true positive rate was 0.825, the false positive rate was 0.026, the true value and model precision rate was 0.797, the recall rate was 0.825, *F*-measure was 0.806, and receiver operating characteristic curve was 0.964.

According to this model, here are some examples of the decision rules: (a) When the customer is an Aries, and his or her source of information is from Facebook and Instagram, this customer's preference is likely to be BREAK TIME WG; (b) A Gemini from Tainan City whose source of information is from a friend, then he or she is more likely to buy BOLT WG; and (c) Customers aged 21 to 25 whose star sign is Sagittarius and received information from the Facebook, then they are more likely to purchase BOLT WG.



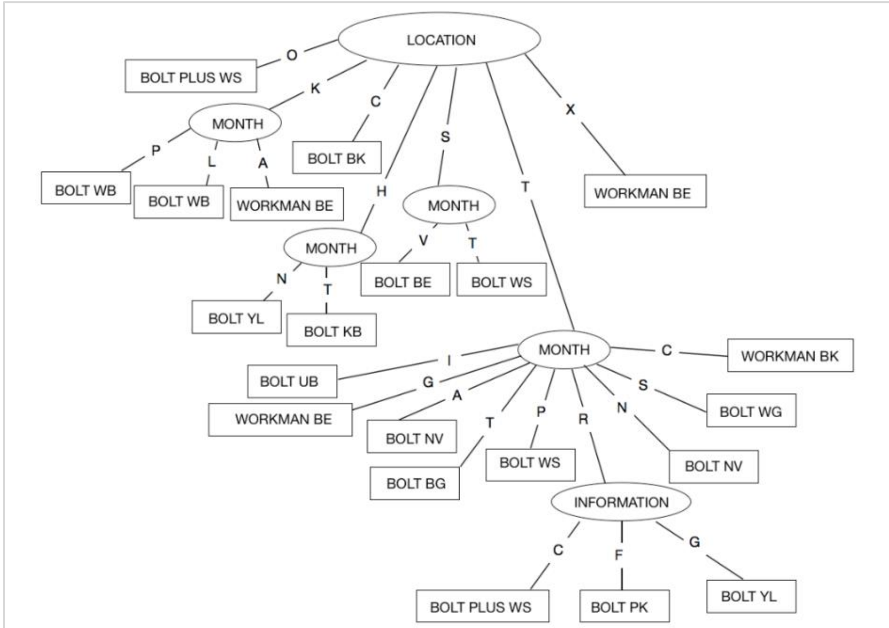
Note: Information = source of information; Location = city; Month = star sign; Years = age.

Figure 3. Decision model based on the second cluster

The total amount of data in the third category is 162, with five attributes. As shown in Figure 4, after 10-fold cross-validation, the total accuracy rate reached 94.90%, the total error rate was 5.10%, Kappa statistics was 0.942, the true positive rate was 0.949, the false positive rate was 0.004, the true value and model precision rate was 0.994, the recall rate was 0.949, *F*-measure was 0.941, and receiver operating characteristic curve was 0.988.

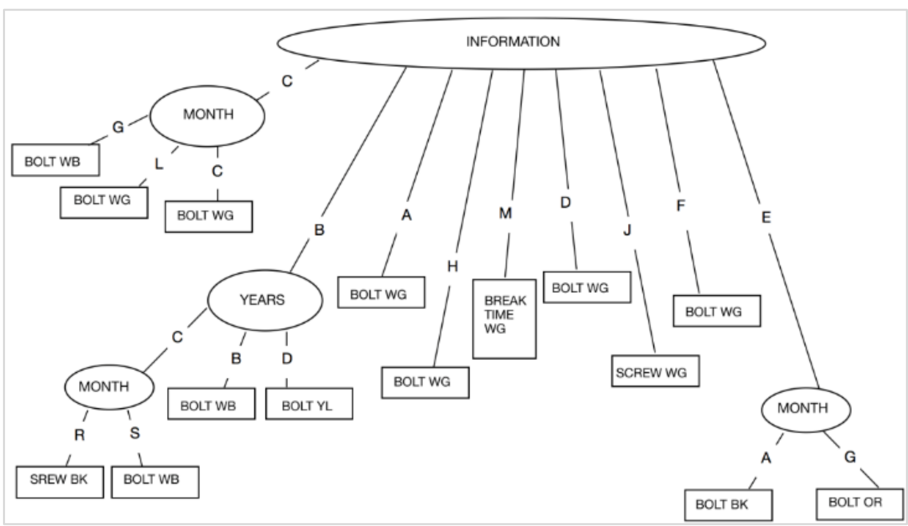
According to this model, here are some examples of the decision rules: (a) When the customers are from New Taipei City, they are more likely to purchase WORKMAN BE; (b) A Pisces from Kaohsiung City is more likely to buy BOLT WB; and (c) Customers from Taipei City whose star sign is Aries and received information from the Instagram, then they are more likely to purchase BOLT PLUS WS.

The total number of data in the fourth category was 115, with five attributes. As shown in Figure 5, After 10-fold cross-validation, the total accuracy rate reached 92.17%, the total error rate was 7.83%, Kappa statistics was 0.902, the true positive rate was 0.922, the false positive rate was 0.012, the true value and model precision rate was 0.898, the recall rate was 0.922, *F*-measure was 0.906, and receiver operating characteristic curve was 0.977.



Note: Location = city; Month = star sign; Information = source of information.

Figure 4. Decision model based on the third cluster



Note: Information = source of information; Month = star sign; Years = age.

Figure 5. Decision model based on the fourth cluster

According to the model, here are some examples of the decision rules: (a) When customers receive information from Facebook, they are more likely to purchase BOLT WG; (b) A Gemini who received information from newspaper and magazines are more likely to buy BOLT OR; and (c) Customers aged 26 to 30 and received information from friends are more likely to purchase SCREW BK.

Model Validation

We made recommendations to Company X on how to customize their target audience settings for their Facebook ads based on the above findings. Then we analyzed the effectiveness of their Facebook ads with the Facebook Relevance Score. It measures the quality of ads and the level of engagement from the target audience. The relevance score ranges from one to ten. A score of one indicates low relevance of the ads to target audiences and a score of ten represents high relevance of the ads to target audiences. Calculation of the score also considers the positive and negative feedback received from audiences. A higher score demonstrates more positive feedback from the audience, and vice versa. The feedback is considered positive when the audience clicks on the ads, clicks on the links, clicks on "liked", shares the ads, watches the video, downloads the app, redeems an offer, or joins an activity. Negative feedback includes inaction on ads, clicks on "hid the ads", clicks on "don't want to see the ads", and other actions showing no interest in the ads (Facebook for Business, 2021a).

The average relevance score of Company X increased from six to eight after following the recommendations from the findings, showing the effectiveness and practicality of our study. Moreover, we adopted the Return on Ad Spending (ROAS; Facebook for Business, 2021b) of Company X to assess the effectiveness of Facebook advertising. ROAS is calculated as the return from Facebook ads divided by total ad spend. Simply put, it is the proportion of revenue generated from a dollar spent on Facebook advertising (Facebook for Business, 2021b). Before the change of audience settings, Company X had a ROAS of 201%. After changing Facebook ads settings as our results suggested, their ROAS surged to 634%. In other words, for every dollar spent on Facebook advertising, the company received 6.34 dollars in return. Again, this is another affirmation of the practical value of our study.

CONCLUSION

Ownership of data is a huge competitive advantage nowadays, not to mention the important role of data mining plays in retailing (Malik, 2013). In the past research, data mining has been used to support decision making in retailing and marketing (Turow *et al.,* 2015). However, the application of such technique was limited to scholarly or theoretical research. Very few have applied the technique on field data, i.e., data taken directly from real transactions (Turow *et al.*, 2015). Even if the data was taken from the industry, no studies have examined the effectiveness of statistical models on actual administration of advertising.

In this study, we analyzed real customer transactional data and found geographical differences in preferences for different color and style. After clustering analysis, we grouped customers into four clusters, and modeled decision trees based on the data to predict future consumptions. In the future, this model can also be applied to make strategic marketing plans.

Furthermore, this study analyzed the off-line transactional sales data and used the results to alter social media advertising audience settings. We then examined the effectiveness of our research model on Facebook advertising. Instead of inventing new ways to run data mining, we applied the conventional technique to the field and received satisfactory feedback. The stronger ROAS value provided preliminary support for our model and paved the way to further studies.

Although this study uncovers fruitful findings about the practical application of data mining techniques in the real business world, there are some limitations. First, we collected data from only one leisure shoe brand in Taiwan, this may limit the generalizability of our results. Future studies may extend this technique to other industries where their businesses operate both online and offline. Second, we used *k*-means algorithm for clustering and C4.5 algorithm to develop decision tree models. There are other methods for clustering or classification and we have not tested our model based on other algorithms. Future studies may adopt other methods to examine their model in real data. Lastly, we have shown the merit of discerning consumption patterns from offline transaction data and applying it to the audience settings of online marketing campaigns, future studies may also want to do the same with online transaction data and find out if the model also gives promising answers.

REFERENCES

Belanche, D., Cenjor, I., Pérez-Rueda, A. Instagram stories versus Facebook wall: an advertising effectiveness analysis. *Spanish Journal of Marketing - ESIC*, 2019, *23*, 69-94.

Bose, I., Mahapatra, R.K. Business data mining – a machine learning perspective. *Information & Management*, 2001, *39*(3), 211-225.

Chen, M.C., Chao, C.M., Wu, K.T. Pattern filtering and classification for market basket analysis with profit-based measures. *Expert Systems*, 2012, *29*(2), 170-182.

Chou, J.S., Hsu, Y.C., Lin, L.T. Smart meter monitoring and data mining techniques for predicting refrigeration system performance. *Expert Systems with Applications*, 2014, *41*(5), 2144-2156.

Chen, M.S., Han, J., Yu, P.S. Data mining: an overview from a database perspective. *IEEE Transactions on Knowledge and data Engineering*, 1996, *8*(6), 866-883.

Cho, Y.H., Kim, J.K., Kim, S.H. A personalized recommender system based on web usage mining and decision tree induction. *Expert systems with Applications*, 2002, *23*(2), 329-342.

Chiclana, F., Kumar, R., Mittal, M., Khari, M., Chatterjee, J.M., Baik, S.W. ARM–AMO: an efficient association rule mining algorithm based on animal migration optimization. *Knowledge-Based Systems*, 2018, *154*, 68-80.

Chan, C., Lewis, B. A basic primer on data mining. *Information Systems Management*, 2002, *19*(4), 56-60.

Chi-Hsien, K., Nagasawa, S. Applying machine learning to market analysis: knowing your luxury consumer. *Journal of Management Analytics*, 2019, *6*(4), 404-419.

Constantinides, E. Foundations of social media marketing. *Procedia – Social and Behavioral Sciences*, 2014, *148*, 40-57.

Deshpandé, R., Stayman, D.M. A tale of two cities: distinctiveness theory and advertising effectiveness. *Journal of Marketing Research*, 1994, *31*, 57-64.

Djenouri, Y., Belhadi, A., Fournier-Viger, P., Fujita, H. Mining diversified association rules in big datasets: a cluster/GPU/genetic approach. *Information Sciences*, 2018, *459*, 117-134.

Daskalaki, S., Kopanas, I., Goudara, M., Avouris, N. Data mining for decision support on customer insolvency in telecommunications business. *European Journal of Operational Research*, 2003, *145*(2), 239-255.

Exenberger, E., Bucko, J. Analysis of online consumer behavior – design of CRISP-DM process model. *Agris On-Line Papers in Economics & Informatics*, 2020, *12*(3), 13-22.

Facebook for Business. (2021a), Relevance score, available at, https://www.facebook.com/business/help/1703421129885535, referred on 21/2/2021.

Facebook for Business. (2021b), Website Purchase ROAS (Return on Ad Spend), available at, https://www.facebook.com/business/help/1283504535023899, referred on 21/2/2021.

Fayyad, U., Piatetsky-Shapiro, G., Smyth, P. The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM*, 1996, 39(11), 27-34.

Fisher, M., Raman, A. Using Data and Big Data in Retailing. *Production and Operations Management*, 2018, *27*(9), 1665-1669.

Ge, Z., Song, Z., Ding, S.X., Huang, B. Data mining and analytics in the process industry: the role of machine learning. *IEEE Access*, 2017, *5*, 20590-20616.

Jarboui, B., Cheikh, M., Siarry, P., Rebai, A. Combinatorial particle swarm optimization (CPSO) for partitional clustering problem. *Applied Mathematics and Computation*, 2007, *192*(2), 337-345.

Kaplan, A.M., Haenlein, M. Users of the world, unite! The challenges and opportunities of social media. *Business Horizons*, 2010, *53*, 59-68.

Kim, J.K., Cho, Y.H., Kim, W.J., Kim, J.R., Suh, J.H. A personalized recommendation procedure for Internet shopping support. *Electronic commerce research and applications*, 2002, *1*(3-4), 301-313.

Kurgan, L.A., Musilek, P. A survey of knowledge discovery and data mining process model. *The Knowledge Engineering Review*, 2006, 21(1), 1-24.

Kaur, H., Wasan, S.K. Empirical study on applications of data mining techniques in healthcare. *Journal of Computer Science*, 2006, *2*(2), 194-200.

Lavidge, R., Steiner, G.A. A model for predictive measurements of advertising effectiveness. *Journal of Marketing*, 1961, *25*, 59-62.

Lee, E.B., Lee, S.G., Yang, C.G. The influences of advertisement attitude and brand attitude on purchase intention of smartphone advertising. *Industrial Management & Data Systems*, 2017, *117*, 1011-1036.

Macdonald, E., Sharp, B. Management perceptions of the importance of brand awareness as an indication of advertising effectiveness. *Marketing Bulletin*, 2003, *14*, 1-11.

Malik, P. Governing big data: principles and practices. *IBM Journal of Research and Development*, 2013, *57*(3/4), 1-13.

Mittal, M., Goyal, L.M., Hemanth, D.J., Sethi, J.K. Clustering approaches for high-dimensional databases: a review. *WIREs Data Mining and Knowledge Discovery*, 2019, *9*(3), e1300.

Madsen, D.Ø., Silva, E.S., Hassani, H. The application of big data in fashion retailing: a narrative review. *International Journal of Management Concepts and Philosophy*, 2020, *13*(4), 247-274.

Ngai, E.W., Xiu, L., Chau, D.C. Application of data mining techniques in customer relationship management: a literature review and classification. *Expert systems with applications*, 2009, *36*(2), 2592-2602.

Peng, C.C., Tsai, C.J., Chang, T.Y., Yeh, J.Y., Hua, P.W. A new approach to generate diversified clusters for small data sets. *Applied Soft Computing*, 2020, *95*, 106564.

Rodrigues, M.W., Isotani, S., Zarate, L.E. Educational data mining: a review of evaluation process in the e-learning. *Telematics and Informatics*, 2018, *35*(6), 1701-1717.

Rubinson, J. Empirical evidence of TV advertising effectiveness. *Journal of Advertising Research*, 2009, *49*, 220-226.

Statista. Most popular social networks worldwide as of July 2021, ranked by number of active users(in millions), available at, https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/, referred on 9/8/2021.

Stewart, K., Kammer-Kerwick, M., Koh, H., Cunningham, I. Examining digital advertising using an affect transfer hypothesis. *Journal of Research in Interactive Marketing*, 2018, *12*, 231-254.

Tseng, W.T., Chiang, W.F., Liu, S.Y., Roan, J., Lin, C.N. The application of data mining techniques to oral cancer prognosis. *Journal of Medical Systems*, 2015, *39*(5), 1-7.

Ting, C.K., Liaw, R.T., Wang, T.C., Hong, T.P. Mining fuzzy association rules using a memetic algorithm based on structure representation. *Memetic Computing*, 2018, *10*(1), 15-28.

Turow, J., McGuigan, L., Maris, E.R. Making data mining a natural part of life: physical retailing, customer surveillance and the 21st century social imaginary. *European Journal of Cultural Studies*, 2015, *18*(4-5), 464-478.

Voorveld, H.A.M., van Noort, G., Muntinga, D.G, Bronner, F. Engagement with social media and social media advertising: the differentiating role of platform type. *Journal of Advertising*, 2018, *47*(1), 38-54.

Wang, X., Ng, C.T. New retail versus traditional retail in e-commerce: channel establishment, price competition, and consumer recognition. *Annals of Operations Research*, 2020, *291*(1), 921-937.

We Are Social. Digital 2021, available at, https://wearesocial.com/digital-2021, referred on 18/8/2021.

Wu, X., Zhu, X., Wu, G.Q., Ding, W. Data mining with big data. *IEEE Transactions on Knowledge and Data Engineering*, 2014, *26*(1), 97-107.

Yin, C., Guo, Y., Yang, J., Ren, X. A new recommendation system on the basis of consumer initiative decision based on an associative classification approach. *Industrial Management & Data Systems,* 2018, *118*(1), 188-203.

Yang, F., Li, F.W. Study on student performance estimation, student progress analysis, and student potential prediction based on data mining. *Computers & Education*, 2018, *123*(1), 97-108.

Yee, O.S., Sagadevan, S., Malim, N.H.A.H. Credit card fraud detection using machine learning as data mining technique. *Journal of Telecommunication, Electronic and Computer Engineering*, 2018, *10*(1-4), 23-27.

Yao, L.Y., Xiong, J.Y. Analysis online shopping behavior of consumer using decision tree. *Advanced Materials Research*, 2011, *271-273*, 891-894.

Yim, M.Y.C., Kim, Y.K., Lee, J.G. How to easily facilitate consumers' mental simulation through advertising: the effectiveness of self-referencing image dynamics on purchase intention. *International Journal of Advertising*, 2020, *40*, 810-834.

Yousef, M., Dietrich, T., Rundle-Thiele, S. Social advertising effectiveness in driving action: a study of positive, negative and coactive appeals on social media. *International Journal of Environmental Research and Public Health*, 2021, *18*(11), 5954.

Zhu, Y.Q., Chen, H.G. Social media and human need satisfaction: Implications for social media marketing. *Business Horizons*, 2015, *58*, 335-345.