ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

# AN INTELLIGENT APPROACH FOR PREDICTING SOCIAL MEDIA IMPACT ON BRAND BUILDING

**[1]ALTYEB ALTAHER, [2]AHMED HAMZA OSMAN**

[1,2]Faculty of Computing and Information Technology in Rabigh, King Abdulaziz University, Jeddah, Saudi Arabia

E-mail: [1]aaataha@kau.edu.sa, [2] ahoahmad@kau.edu.sa

**ABSTRACT**

Social media networks such as Twitter and Facebook plays important roles in many aspects of our lives and affects many of our decisions. This paper presents a data mining model consists of different five classification and regression algorithms to predict the significant performance metrics of posts announced in the Facebook pages of the brands. The algorithms utilized in the model include the Generalized Liner Regression (GLR), Normal Regression (NR), support Vector Machine, Neural Network, and CHAID decision tree classifier. Using a dataset contained a 790 published posts in the cosmetic brand, the Lifetime post consumers achieved the best posts performance metrics with an average accuracy of 0.82 among all the algorithms in the proposed model, followed by the Lifetime post total reach performance metric with an average accuracy of 0.79. The findings of this research potentially help the manager's in making the right decisions regarding whether to publish a post.

**Keywords:** *Data Mining, Classification, Social Media, Brand Building, Performance Metrics*

## 1. INTRODUCTION

The exponential increasing in the number of internet users coupled with the new enhances in the telecommunication technologies, have made the social media as suitable place for customers to discuss ideas and opinions about services and products [1]. According to Statista, the number of social network users is estimated to reach 2.95 billion users in 2020, with more than 1.86 billion users active each month, Facebook is considered as a market pioneer in terms of scope and reach [2]. Considering the fast development in the social media, it could become the potential media channel for companies to contact their customers [3].

Companies started to depend on the social media networks in their strategies to attract the customers [4]. Many Research efforts concentrated on exploring the relations between publishing on social networks and customers' interactions [5]. However, few researchers devoted their efforts to develop predictive approaches able to expect the impact of a post before the publication. An approach with ability to expect the influence of announced posts can offer a significant benefits, related to the suitability of the social media based promotion of products and services. The social media publications have potential impact on the brand building [6]. Therefore, the findings of this research could help in making the right decisions related to improving trademark credit. Data mining techniques offer an important methodology for getting knowledge based on the available data [7]. The data mining techniques have been used to explore the market trends based on the inputs of the users [8]. However, most of the researches concentrated on evaluating the users opinions based on information gathered from different sources [9]. Our research focused on the prediction of the influence of announcing posts on the social media pages of the firm. The main contributions of this paper as follows:

- Developing a model for predicting the influence of posts based on their features
- Evaluating the developed model by measuring the dissimilarity between the predicted value and the real value.
- Exploring the potential of data mining techniques for predicting the influence of social media posts on trademark credit.

The rest of this paper is structured as follows: section 2 presented the related work, the proposed model is explained in section 3. Experimental Results and discussion presented in section 4. Section 5 concludes our paper.

## 2. RELATED WORK

The impact of brand advertisements in social media on the brand trust. According to [11], digital interactions are probably affecting 64 cents of each dollar paid in shopping starting from 2015, based on a survey of over 3000 US customers, showing the significant influence of social media on the incomes of the firms.

The establishment of online customer media may be initiated by social media networks such as Facebook and Twitter, offering an evolving concern about specific products, corporations and trademarks. Therefore, to build commercial importance, corporations should integrate the community development as portion of the of social media implementation [12]. Trademark societies based on social media improve spirits of community in the members and support making significance for firms and participants [13]

Hudson et al.[14] performed three studies to investigate the correlation between customers behavior and social media practice, the results indicated that attracting customers through social media was related to achieving high consumer–brand relationships. In [15], Griffiths and McLean assessed the adoption of social media in UK-based firm, they found that social media promotion had a optimistic influence on trademark reputation.

Most of the researches in the literature strongly signifies the impact of social media on trademark building. However, research in this area is still limited. Research studies are expected to be conducted in future, presenting innovative studies for satisfying the need for more investigation [16].

## 3. THE PROPOSED MODEL

The proposed model for consists of different five classification and regression algorithms to predict the significant performance metrics of posts announced in the Facebook pages of the brands. The algorithms utilized in the model include the Generalized Linear Regression (GLR), Normal Regression (NR), support Vector Machine, Neural Network, and CHAID decision tree classifier. The Generalized Linear model enlarges the general linear model, thus the dependent variable is linearly associated to the elements and covariates through identified connection function. The Support Vector Machine (SVM) node allows the classification of data into two sets without over fitting. SVM performs better with large data sets. Neural network is a useful technique to cluster the data set into different sets. It consist of a big number of interconnected neurons working together to find solutions to certain problems. Neural networks need to be trained to learn. The CHAID method creates decision trees based on chi-square statistics to find best splits. In addition, CHAID has the ability to create non-binary trees [17].

## 4. CORPUS AND EXPERIMENTAL RESULTS

This section presents the used corpus and the experimental results.

### 4.1 Corpus

The introduced method comprises a different classification and machine learning experimental techniques at its principal, resulting in a UCI Facebook dataset process. The UCI Facebook dataset [18] has been collected from the internet Facebook's side of the renowned cosmetic brand in 2014. The dataset contained a 790 published posts in the cosmetic brand. It is categorized into four main types includes; Identification features (the attributes that defining each single post); content features (represents the word-based part of the published post); Classification features (the attributes that can describe the published post); Performance features (the evaluation parameters that influence of the published post such as 'Page total likes' feature). Table 1 demonstrates the UCI Facebook Corpus that was used in this study.
Table 1 shows the attributes gathered in the dataset. The data mainly was extracted from the Facebook page of the firm. The exclusions were "category" and "total interactions". The table explains the descriptions and domain values of these features as well.

Journal of Theoretical and Applied Information Technology
15th September 2017. Vol.95. No.17

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

*Table 1: The features and features types in the UCI Facebook Corpus .*

| No | Feature | Feature Type | Source | Data type and Domain |
|---|---|---|---|---|
| 1 | Posted | Identification | Facebook | Date/time |
| 2 | Permanent link | Identification | Facebook | Text |
| 3 | Post ID | | | |
| 4 | Post message | Content | Facebook | Text |
| 5 | Type | Categorization | Facebook | Factor: {Link, Photo, Status, Video } |
| 6 | Category | Categorization | Facebook | page managers Factor: {action, product, inspiration } |
| 7 | Paid | Categorization | Facebook | Factor: {yes, no } |
| 8 | Page total likes | Performance | Facebook | Numeric |
| 9 | Lifetime post total reach | | | |
| 10 | Lifetime post total impressions | | | |
| 11 | Lifetime engaged users | | | |
| 12 | Lifetime post consumers | | | |
| 13 | Lifetime post consumptions | | | |
| 14 | Lifetime post impressions by people who have liked your page | | | |
| 15 | Lifetime post reach by people who like your page | | | |
| 16 | Lifetime people who have liked your page and engaged with your post | | | |
| 17 | Comments | Performance | Facebook | Numeric |
| 18 | Likes | | | |
| 19 | Shares | | | |
| 20 | Total interactions | Performance | Computed | Numeric |

## 4.2 The experimental results and discussion

Our experiments models were conducted and performed in two main phases; training and testing stages as shown in figure 1 and figure 2.

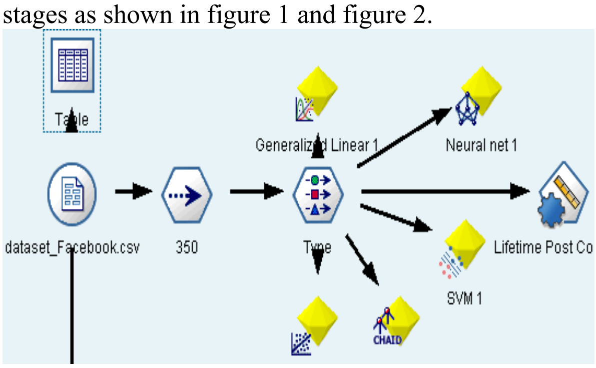



*Figure 1: Training Model*

The dataset divided into 80% for training phase and 20 for the testing phase. In the training stage, all features are selected as an input of the suggested model except the target feature. The target features were selected based on the discretion of the dataset that was explained by Sérgio Moro and et.al [18]. Their proposed model can be selected one of the twelve attributes from feature 9 to 20 as ordered in Table 1. Actually, these features used for computing the total interactions on the Facebook plus cosmetic brand site. In our proposed model, we used the important features that can affect the accuracy performance including feature number 9, 10, 12 and 14.

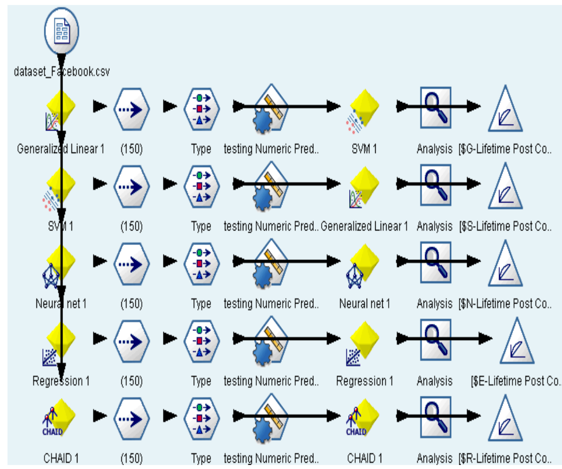



*Figure 2: Testing Model*

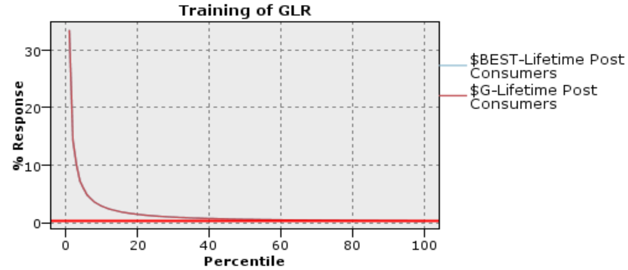
On the other hand, the testing phase was performed based on the output of the training phase by including the newly generated features such as predicted value and predicted label. The suggested model used different five classification and regression algorithms to examine the proposed model. These algorithms are Generalized Liner Regression (GLR), Normal Regression (NR), support Vector Machine, Neural Network, and CHAID decision tree classifier. The model has been build starting by selecting the training dataset and ending by extracting results using the five classifiers method separately. Table 2 shows the extracted results of the proposed model using different classifier algorithms.

Journal of Theoretical and Applied Information Technology
15th September 2017. Vol.95. No.17

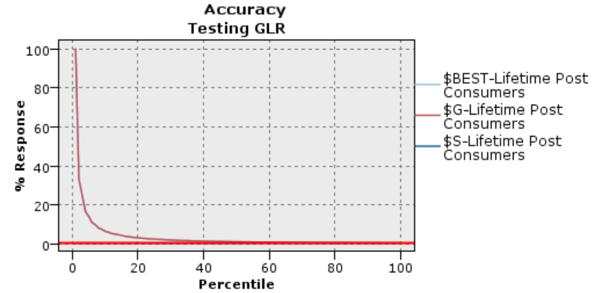
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

| Algorithm | Training Accuracy/Error | | Testing Accuracy/Error | | No of feature used | Target |
|---|---|---|---|---|---|---|
| | Accu | Error | Accu | Error | | Output Feature |
| GLR | 0.99 | 0.1 | 1 | 0 | 18 | Lifetime post consumers |
| Neural Network | 0.96 | 0.4 | 0.80 | 0.2 | 18 | Lifetime post consumers |
| SVM | 0.77 | 0.33 | 0.58 | 0.42 | 18 | Lifetime post consumers |
| Regression | 0.99 | 0.1 | 1 | 0 | 16 | Lifetime post consumers |
| CHAID | 0.95 | 0.5 | 0.72 | 0.28 | 13 | Lifetime post consumers |
| | | | | | | |
| GLR | 0.99 | 0.001 | 1 | 0 | 18 | Lifetime post total reach |
| Neural Network | 0.82 | 0.18 | 0.70 | 0.3 | 18 | Lifetime post total reach |
| SVM | 0.33 | 0.67 | 0.59 | 0.41 | 18 | Lifetime post total reach |
| Regression | 0.99 | 0.01 | 1 | 0 | 16 | Lifetime post total reach |
| CHAID | 0.90 | 0.1 | 0.69 | 0.31 | 13 | Lifetime post total reach |
| | | | | | | |
| GLR | 0.99 | 0.01 | 1 | 0 | 18 | Lifetime post total impressions |
| Neural Network | 0.91 | 0.09 | 0.63 | 0.37 | 18 | Lifetime post total impressions |
| SVM | 0.36 | 0.64 | 0.60 | 0.4 | 18 | Lifetime post total impressions |
| Regression | 0.99 | 0.01 | 1 | 0 | 16 | Lifetime post total impressions |
| CHAID | 0.88 | 0.12 | 0.69 | 0.31 | 13 | Lifetime post total impressions |
| | | | | | | |
| GLR | 0.98 | 0.02 | 1 | 0.2 | 18 | Total no of impressions who have liked a page. |
| Neural Network | 0.88 | 0.12 | 0.67 | 0.42 | 18 | Total no of impressions who have liked a page. |
| SVM | 0.45 | 0.55 | 0.59 | 0 | 18 | Total no of impressions who have liked a page. |
| Regression | 0.98383 | 0.01617 | 1 | 0.28 | 16 | Total no of impressions who have liked a page. |
| CHAID | 0.90475 | 0.09525 | 0.69 | 1 | 8 | Total no of impressions who have liked a page. |

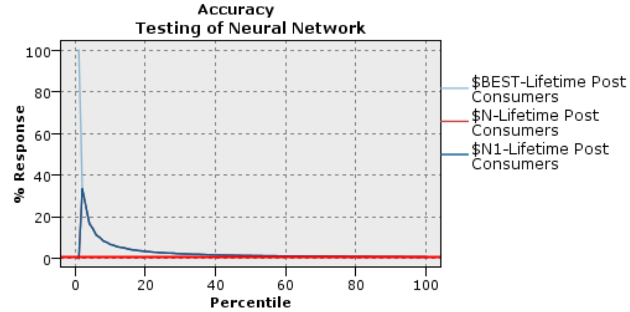
Table 2 illustrates the performance results of training and testing phases. The results include prediction accuracy and misclassification error for each classifier algorithm with different target field such as Lifetime post consumers, Lifetime post total impressions, Lifetime post total reach, and the total number of the persons who put like on the firm's page. We noted that the regression algorithms (GLR and NR) achieved better results in the training accuracy of 99% and 0.1 miss-classification error. However, the other examined classifier methods obtained unsatisfying prediction results in term of training and testing accuracy. We also noted that the number of selected input by each classifier is different from the prediction algorithm to another and this due to the variation of the input features and the inner characteristic of each classifier method itself. We found out that the GLR obtained optimal results in the training and testing phases using all the 18 input features. Whereas the other techniques used less than 18 features and even the other classifier method that used all features obtained fewer accuracy results compared with other examined techniques.



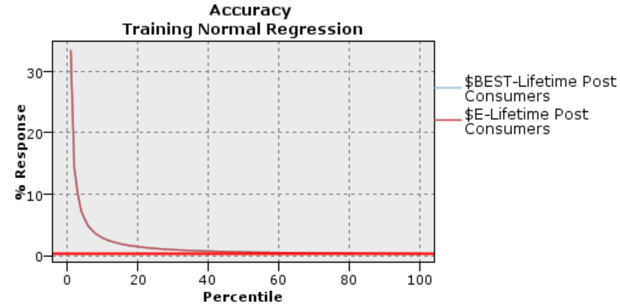

Accuracy



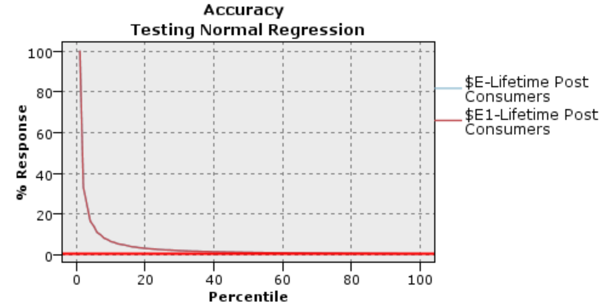

Accuracy




Accuracy




Accuracy




Accuracy

**Journal of Theoretical and Applied Information Technology**
15th September 2017. Vol.95. No.17


ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

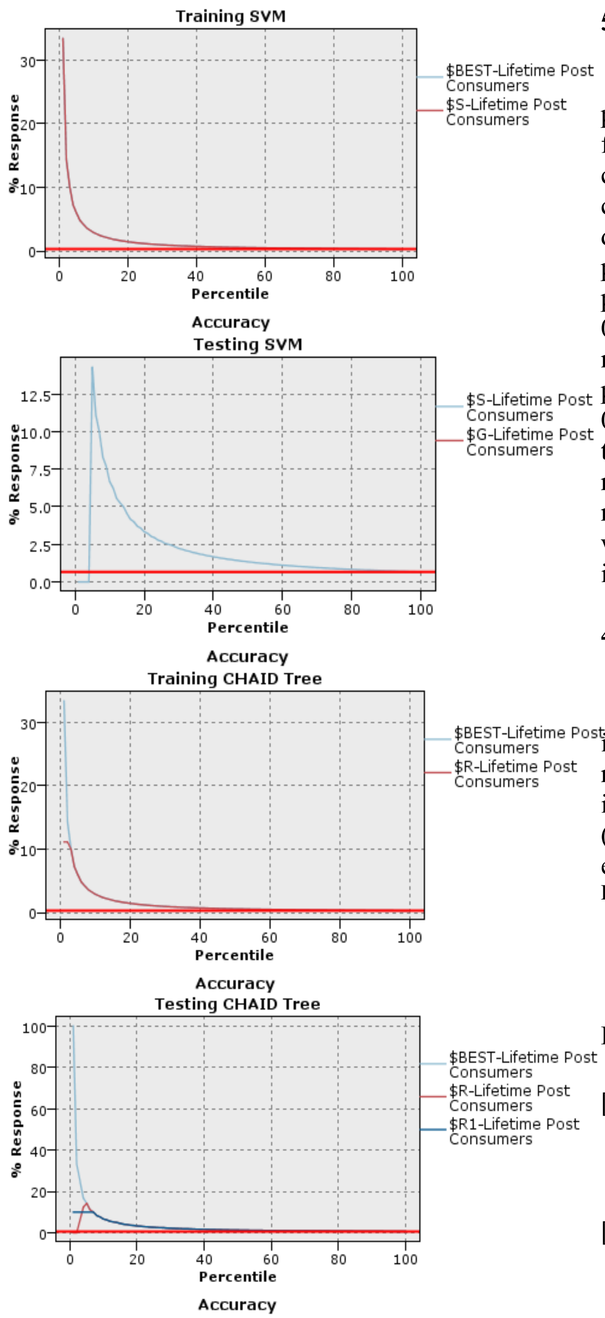



*Figure 3:Training and testing results visualization.*

Figure 3 demonstrates the training and testing results visualization. The Lifetime post consumers achieved the best posts performance metrics with an average accuracy of 0.82 among all the algorithms in the proposed model, followed by the Lifetime post total reach performance metric with an average accuracy of 0.79. The findings of this research potentially help the manager's in making the right decisions regarding whether to publish a post.

## 5. CONCLUSION

This paper presented data mining model to predict the impact of performance metrics obtained from posts announced in the Facebook page of the company. The proposed model used different five classification and regression algorithms, which uses different number of input features. The Lifetime post consumers achieved the best posts performance metrics with an average accuracy of 0.82 among all the algorithms in the proposed model, followed by the Lifetime post total reach performance metric with an average accuracy of 0.79. The findings of this research potentially help the manager's in making the right decisions regarding whether to publish a post. For future research, more advanced data mining techniques with be considered to predict the social media influence on trademark building.

## 4. EQUATIONS

When numbering equations, enclose numbers in parentheses and place flush with right-hand margin of the column. Equations must be typed, not inserted.
(If nonstandard fonts are used its better to put equations as images instead of text)
Example:

$$\mathrm{Net}_j = w_0 + \sum_{i=1}^{n} x_i w_{ij} \qquad (1)$$

## REFRENCES:

[1] Kaplan, A. M., & Haenlein, M. Users of the world, unite! The challenges and opportunities of Social Media. Business Horizons, 53(1), 2010, pp. 59–68.
[2] Statista, Number of social network users worldwide from 2010 to 2020 (in billions). Retrieved May 3, 2017, From https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/
[3] Korschun, D., & Du, S. How virtual corporate social responsibility dialogs generate value: A framework and propositions. Journal of Business Research, 66(9), 2013, pp. 1494–1504.
[4] Lariscy, R.W., Avery, E. J., Sweetser, K. D., & Howes, P. Monitoring public opinion in cyberspace: How corporate public relations is facing the challenge. Public Relations Journal, 3(4), 2009, pp. 1–17

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[5] Cvijikj, I. P., Spiegler, E. D., & Michahelles, F. The effect of post type, category and posting day on user interaction level on Facebook. *In Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on*, 2011, pp. 810–813

[6] Edosomwan, S., Prakasan, S. K., Kouame, D.,Watson, J., & Seymour, T. The history of social media and its impact on business. Journal of Applied Management and Entrepreneurship, 16(3), 2011, pp. 79–91.

[7] Turban, E., Sharda, R., Delen, D., & Efraim, T. Decision support and business intelligence systems (9th ed.), 2011. Pearson.

[8] Trainor, K. J., Andzulis, J. M., Rapp, A., & Agnihotri, R. Social media technology usage and customer relationship performance: A capabilities-based examination of social CRM. Journal of Business Research, 67(6), 2014, pp. 1201–1208.

[9] Bianchi, C., & Andrews, L. Investigating marketing managers' perspectives on social media in Chile. Journal of Business Research, 68(12), 2015,pp. 2552–2559.

[10] Laroche, M., Habibi, M. R., Richard, M. O., & Sankaranarayanan, R. The effects of social media based brand communities on brand community markers, value creation practices, brand trust and brand loyalty. Computers in Human Behavior, 28(5), 2012, pp. 1755–1767.

[11] Digital, Deloitte .Navigating the new digital divide—capitalizing on digital influence in retail, 2015. Retrieved September 12, 2015, from http://www2.deloitte.com/content/dam/Deloitte/us/Documents/consumer-business/us-cb-navigating-thenew-digital-divide-v2-051315.pdf

[12] Culnan,M. J.,McHugh, P. J., & Zubillaga, J. I. Howlarge US companies can use Twitter and other social media to gain business value. MIS Quarterly Executive, 9(4), 2010, pp. 243–259.

[13] Laroche, M., Habibi, M. R., Richard, M. O., & Sankaranarayanan, R. The effects of social media based brand communities on brand community markers, value creation practices, brand trust and brand loyalty. Computers in Human Behavior, 28(5), 2012, pp. 1755–1767.

[14] Hudson, S., Huang, L., Roth, M. S., & Madden, T. J. The influence of social media interactions on consumer–brand relationships: A three-country study of brand perceptions and marketing behaviors. International Journal of Research in Marketing, 2015.

[15] Griffiths, M., & McLean, R. Unleashing corporate communications via social media: A UK study of brand management and conversations with customers. Journal of Customer Behaviour, 14(2), 2015, pp. 147–162.

[16] Okazaki, S., & Taylor, C. R. Social media and international advertising: Theoretical challenges and future directions. International Marketing Review, 30(1), 2013,pp. 56–71.

[17] Cortez, P. Data mining with neural networks and support vector machines using the R/rminer tool. Advances in data mining. Applications and theoretical aspects, 2010, pp. 572–583.

[18] Moro, S., Rita, P., & Vala, B.. Predicting social media performance metrics and evaluation of the impact on brand building: A data mining approach. Journal of Business Research, 69(9), 2016, pp. 3341-3351.