ISSN: 2278-4632
Vol-11 Issue-02 2021

PAID ADVERTISEMENT ON FACEBOOK AN EVALUATION USING A DATA MINING APPROACH

Pappu Sharada, Dr. Anil Kumar Mishra[2]
[1](*Computer Science & Engineering, Gandhi Engineering College, Bhubaneswar*)
[2](*Computer Science & Engineering, Gandhi Engineering College, Bhubaneswar*)

ABSTRACT

This paper focuses on evaluating the performance of paid publications (paid ads) on Facebook and proposes a managerial implication to maximize the paid publications' performance in reaching as many people as possible with the greatest possible engagement. Artificial neural networks, Garson's algorithm, and support vectormachine weighting were used to analyze publication characteristics. The results demonstrate that in terms of number of people reached, paid publications are only 9.52% relevant, comparedwith24.35%forposthour,16.85%fortotalpagelikes,and16.44%fortypeof publication. With regards to number of comments, likes, and shares, paid advertisement contributes 10.46%, 19.51%, and 15.55% relevance, respectively. Additionally, the real numbersofpublicationimprovementsfromusingpaidadsandnonpaidadsarecompared and calculated. This paper provides a managerial implication for marketing managers to improve the paid ad performance of company Facebookpages.

Keywords: Social media, Data mining, Paid ads, Artificial Neural Networks.

1. INTRODUCTION

In recent years, social media has been one of the most famous communication platforms among students (Pelling and White, 2009). Social media provides various opportunities such as reaching out to new customers, interacting with current costumers, promoting, and introducing special products or services (Curran et al., 2011). Use of social media as a marketing and promotion tool has been widely recognized as a proven, powerful tool to gain millions of customers (Hanna et al., 2011). Online advertising revenue increased from $36 million in 2002 to $6 billion in 2012 and is expected to increase further in the future (Yuan et al., 2013).

In early 2018, there were 132.7 million Internet users in Indonesia, and more than 90% of those Internet users were active on Facebook (wearesocial, 2018). This information reveals huge opportunities for online advertisement through Facebook in Indonesia. For companies that want to promote their products to Indonesia, using Facebook as an online publisher is probably a good strategy. Moreover, 27% of the surveyed respondents in recent studies indicate that Facebook is the most efficient marketing tool for companies with fewer than 500 workers (Needleman, 2014).

Furthermore, most private universities in Malaysia rely on social media to attract new students (Krishnan and Sajilan, 2014).

Facebook as an advertisement tool has two major methods to promote products or services: organic promotion and paid promotion. Through organic promotion, users can create free company pages or individual fan pages, group discussions, or even personal accounts without any charge. However, if companies decide to use organic promotion (without any payment), then the company fan pages must have numerous followers and Facebook mandates some specific publishing polices. By contrast, if companies are willing to spend money on Facebook paid advertisements (paid ads), the reach of the company to potential customers is wider and does not rely solely on the followers of company fan pages. Paid ads benefit the company because potential customers can be targeted without limitation. However, several researchers have discovered thatFacebook paid ads do not reach new customers effectively (e.g. Margarida,2013).

Because paid advertisements are not cheap, each decision must be controllable. The results of each decision should be easily measured. Data mining provides someexcellent research methods for analyzing various data sets (Turban et al., 2011). Moreover, data mining is regarded as a powerful tool for extracting information from complex and copious social media data (Barbier and Liu, 2011). Machine learning is a set of data mining methods for analyzing such data sets. Artificial neural networks (ANNs) have been widely used as data mining tools for social media data (e.g Krebs,2017).

ThisresearchaimedtoreviewtherelevanceofFacebookpaidadsintermsofnumber ofreachedpeopleandtogaugethepublicationengagementofeachpostbyusingANNs. Garson's algorithm was used to interpret the ANN results. In addition, to helpmarketing managers in optimizing the performance of paid publications, a managerial implication regardingpublicationwasproposedbyusingsupportvectormachineweight(SVM).The

ISSN: 2278-4632
Vol-11 Issue-02 2021

managerial implications consist of a rank of the input variables based on their relevance to reach numerous people and boost engagement of each paidpublication.

The results of the study reveal that paid ads accounts for only 9.52% of relevance to total lifetime post reach. Compared with other inputs, such as post hour, total page likes, and types of publications, the reach of paid ads in the post's lifetime was less effective. Paid ads have 10.46%, 15.55%, and 19.51% of the relevance of comments, shares, and likes, respectively. In this paper, a managerial implication strategy based on the importance of each subfeature (Table 1) is proposed to improve paid publication performance. The most relevant subfeatures are then considered as the most important factors in improving total lifetime reach and total engagement. Thus, this relevance rank of subfeatures would be considered when making a publication decision for publishing through the company's Facebook page. The publication decision would specify the post time,posttype,postcategory,andotherrelatedstrategies.Comprehensiveexplanationis provided in the managerial implicationssection.

2. RELATEDWORK

Theoretically,acomputershouldworkbetterthanahumanbrain.Componentsinside the computer enable remarkable processing ability and memory. However, the human brain is adaptive and continues to learn and thus can manage itself during its long life (Kriesel, 2007). These versatile characteristics were implemented in ANNs for data mining operations (e.g., classification and prediction). ANNs consist of three types of layers: input layers, hidden layers, and output layers. These interconnected groups of artificial neurons propagate information through variousconnections.

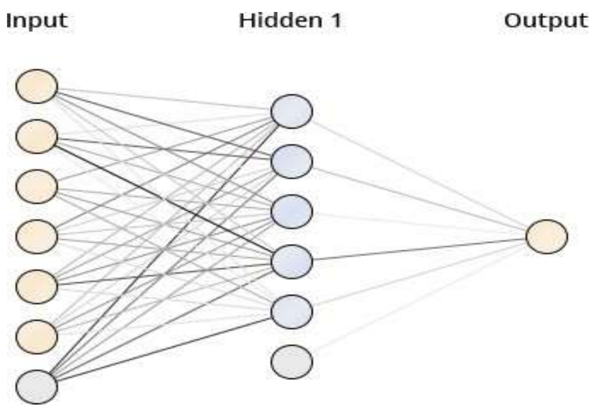

Figure 1 ANNs results using RapidMiner

The effectiveness of ANNs for the analysis of social media data has been proven by some researchers (e.g. Bollen et al., 2011). ANNs proved to be powerful tools in predictingusers'emotionsthroughsocialmediapostreactions(Krebs,2017).Inthedata miningfield,thenonlineartrainingcapabilitiesofANNshavebeenwidelyrecognizedas an excellent tool especially in large and complicated data set operations (Misra and Dehuri,2007).Therefore,inthisresearchweusedANNstoanalyzetherelevanceofpaid advertisements on Facebook. Because ANNs generate complex hidden layers andoutput weights, Garson's algorithm was used to interpret the hidden layers and output weights. Garson's algorithm was invented by Garson (1991) and then modified by Goh (1995). The present study used Garson's algorithm to obtain the relative importance of each network input. This relative importance is calculated by three major steps. The first step is to multiply hidden neuron weights and output weights. The second step is to calculate the relative contribution of each hidden neuron. The third step is to calculate the relative importanceoftheinputsbytherelativecontributionresults.Inthisstudy,fordetermining theimportanceofeachinputandsubinputofpublicationcharacteristics,SVMweighting was employed. SVM considers a data set as a two-group classification problem and determines the best hyperplane (Cortes and Vapnik,1995).

ISSN: 2278-4632
Vol-11 Issue-02 2021

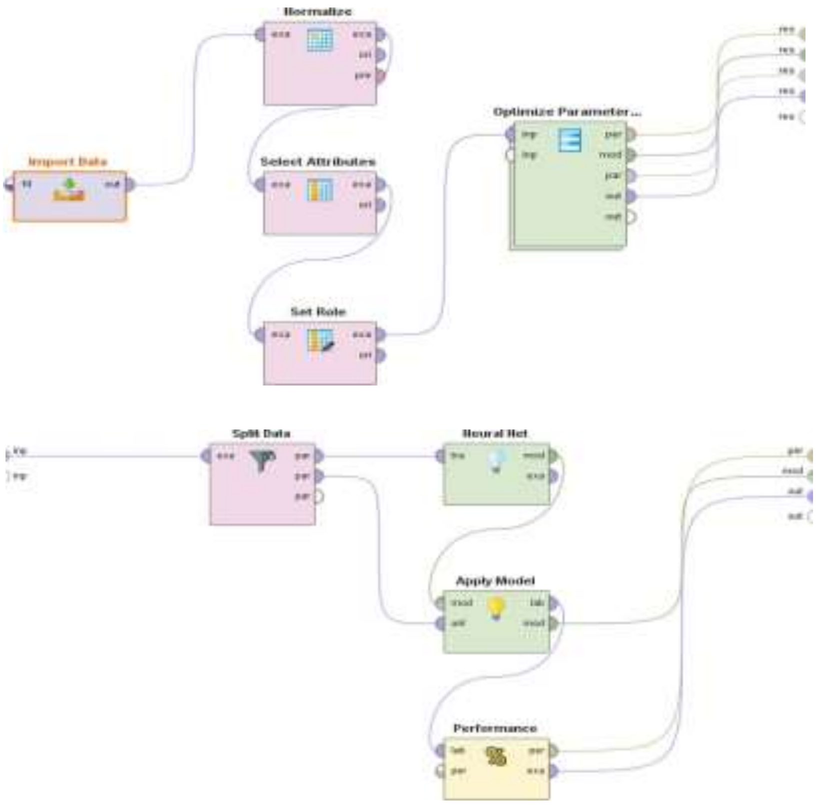

Figure 2. Data mining and optimize parameter processes Using RapidMiner (Grid)

3. DATA ANDMETHODOLOGY

ThedatawasobtainedfromaUCIdatasetandconsistedof500postspublishedfrom a famous cosmetic company (Moro et al., 2016). This data was processed with ANNs to measure the performance of paid ads. To interpret the results of ANNs, Garson's algorithm was used. To guarantee the accuracy of the ANNs, a grid method was used to determine the number of training cycles. Moreover, before the grid method was run, the datawassplitintotwoparts;thefirstpartwasusedastrainingdata(70%)andthesecond part was used as test data (30%). A complete data mining process using RapidMiner is shown in Fig. 2. Input variables consisted of seven characteristics of each publishedpost and four label variables of company Facebook posts. Details are described in Table 1. First, seven input variables were standardized by using Z-transformation and the dependent variable (Y) was preprocessed into the range {0, 1} using theformula:

$$r_n = \frac{y_n - \min(Y)}{\max(Y) - \min(Y)}$$

Table 1. Features

Features	Role
Pages Total Likes	Input Variables
Type	Input Variables (Photo, Status, link, and Video)
Category	Input Variables (Action, Product, and Inspiration)
Post Month	Input Variables (January-December)
Post Weekday	Input Variables (Sunday-Saturday)
Post Hour	Input Variables (1-24)
Paid	Input Variables (1,0)

Total lifetime post reach	Output or label variable
Comments	Output or label variable
Likes	Output or label variable
Shares	Output or label variable
Total Interactions	Comments + likes + shares

Then, 500 posts' data were used to obtain the hidden layers and output weights. To further analyze the paid ads' effectiveness, we used a sampling method to select and compare 139 paid and 139 nonpaid ads. At the conclusion of this research, a managerial strategy on how to publish a publication was obtained by feeding the seven inputs into the system and analyzing them in terms of SVM weight. This managerial implication is aimed to help marketing managers in selecting ad type, ad category, and ad time, when publishing their company Facebook pages. Because the purpose was to find the best strategy for paid publications, 139 publications that used paid ads were considered.

Table 2. Hidden-output ANNs results

Variable	HN A	HN B	HN C	HN D	HN E
Pages Total Likes	1.567	5.285	-2.702	0.808	-1.696
Type	1.923	4.876	0.373	3.192	-1.809
Category	1.118	1.778	-0.157	2.779	-0.679
Post Month	-1.005	-2.93	3.275	1.324	1.675
Post Weekday	-1.617	-0.026	-0.713	3.272	0.97
Post Hour	-3.627	-1.017	3.061	5.216	-3.438
Paid	-0.065	-1.514	-2.702	0.387	-1.846
Output	-0.477	1.265	-0.726	-0.271	-0.967

4. DISCUSSIONS ANDRESULTS

Paid Ads and Total lifetime postreach

We analyzed the weight of each input variable by using ANNs. To achieve the best possibleresults,gridsearchwasperformedtofindtheoptimalparameters.Ingridsearch analysis,thenumberoftrainingcycleswassetintherange500–10,000cycles.Theresult revealed that the configuration with 1070 (Table 3) training cycles had the smallest root meansquarederrors.Thisobservationwasusedasthespecifiednumberoftrainingcycles for analyzing the weights of the seven inputs in relation to the total reach of a post over its entire lifetime. The results are shown in Table2.

Table 3 Optimized parameters of input variables with respect to total reach (Grid)

Iterations	Training cycles	Root Mean Squared error
1	500	0.15
2	548	0.15
.	.	.
.	.	.
.	.	.
12	1023	0.17
13	1070	0.12

ISSN: 2278-4632
Vol-11 Issue-02 2021

14	1118	0.28
.	.	.
.	.	.
.	.	.
.	.	.
200	9953	0.22

Because the results of ANNs only consist of hidden neuron weights and output weights, we applied Garson's algorithm to interpret the hidden neuron weights. The resultsofGarson'salgorithmareshowninFigure3.Overallweightingresultsrevealthat the hour of the publication is more important than other inputs (Fig. 3). Post hour makes a 24.35% contribution to total lifetime post reach; total page likes of publications is in secondplacewith16.85%relevance;typeofpostisinthirdplacewith16.44%relevance; andpostmonthisinfourthplacewith14.57%.Basedontheseresults,wediscoveredthat paid ads through company Facebook pages only have 9.52% relevance to lifetime total reach of a post. Post hour, page total likes, type, post month, and post weekday were determinedtohavemorerelevancethanpaidadstototallifetimepostreach.Theseresults demonstrate that paid ads contribute little to lifetime total reach and post hour was the mostrelevantfactorthatenabledpublicationstoreachnumerouspeople.Theimplication is that a company should pay attention to timely publishing; a company should not randomly publish a post and promote it through paidads.

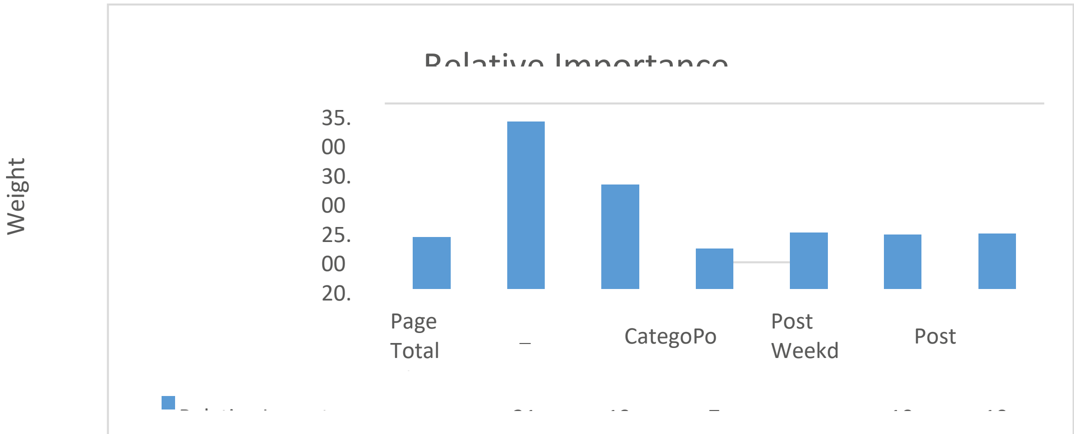

Figure 5 Relative importance of input variables with respect to number of comments

Paid ads andComments

To calculate the weights of seven inputs with respect to number of comments, first we determined the number of training cycles using a grid search method. The results proved that 690 training cycles had the lowest error. Therefore, we implemented 690 training cycles to calculate weight with ANNs. Based on the ANNs weighting and Garson's algorithm results, we understand that type of post has the highest relevance of allinputs(Fig.5).Thetypeofpublicationhas31%relevancetothenumberofcomments. This insight provides valuable information. To improve the number of comments, publications rely on the type of the publications. Photo, link, video, and status were the types in our research. Regarding the types of publications that contribute to number of comments,theresultsarediscussedinthemanagerialimplicationsection(section5).The second most important factor is the category of the post. In our research, post categories were divided into action (special offer), product, and inspiration. Thus, based on these results, we conclude that to generate numerous comments, the type and category are essential. To understand which type and category contribute the most the number of comments, the weight of each subinputs was analyzed using SVM. Those subinput variables are discussed in the managerial implicationsection.

Table 4 Optimal parameters of input variable with respect to number of comments (Grid)

Iterations	Training Cycles	Root mean squared error

ISSN: 2278-4632
Vol-11 Issue-02 2021

1	500	0.041026
2	548	0.053121
3	595	0.052383
4	643	0.039198
5	690	0.027391
6	738	0.049775
.	.	.
200	9953	0.091451

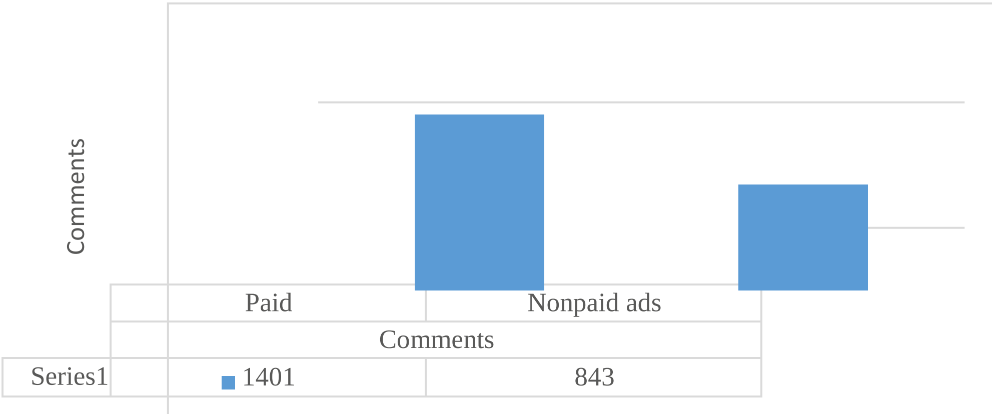

Figure 6 Contribution of paid ads to number of comments

Furthermore, comparing the number of comments between publications with and without paid ads, we observe that publications that used paid ads generated 1,401 comments from 139 publications, compared with 843 comments without paid ads (Fig. 6). Each of the posts that used paid ads generated 10.10 comments on average, and each nonpaidpostgenerated6.01onaverage.Thisresultindicatesthatpaidadscontributelittle to comment activity. This calculation is confirmed by the weight of each input variable in the context of the number of comments. Use of paid ads only has a relevance of 10.46 to encourage comments. Type and category are more effective factors that contribute more to encouragingcomments.

Table 5 Optimal parameters of inputs with respect to number of likes (Grid)

Iterations	Training Cycles	Root mean squared error
1	500	0.041068
2	548	0.051143
3	595	0.053334
.	.	.
.	.	.
.	.	.
15	1165	0.06102
16	1213	0.038819
17	1260	0.069318
.	.	.
.	.	.
.	.	.
200	9953	0.091182

This result shows that in supporting number of likes, paid ads function better than supportingcomments.Moreover,basedoncomparisonresultsfrom139publicationswith paid ads and 139 publications

ISSN: 2278-4632
Vol-11 Issue-02 2021

without paid ads, publications with paid ads generated 32,755 total likes and earned 235 likes for each publication on average, compared with 20,458 total likes and 147 likes for each publication on average for publications without paid ads (Figure 8). This implies that paid ads provide 59% more likes. Category of publication provides 17.99% relevance for number of likes. Thus we can conclude that paid ads, type, and category of publication dominate the contributions to number oflikes for each publication. Therefore, we conclude that using paid ads and selecting the best type and category of publication would considerably increase the number oflikes.

Paid ads andShares

Similartothepreviousexperiments,gridanalysiswasperformedtofindtheoptimal parameters. Thus, 1735 training cycles with 0.029 root mean squared error were determined to be optimal (Table 6). The weights of the seven inputs are shown inFigure
9. Category of publication had a vital role in supporting the number of shares of each publication. The relevance of category attained 21% importance; paid ads provided a 15.55% contribution to the number of shares.

Table 6 Optimal parameters of input variable respect to number of shares (Grid)

Iterations	Training Cycles	Root mean squared error
1	500	0.030315
2	548	0.039944
.	.	.
.	.	.
27	1735	0.029484
28	1783	0.058806
.	.	.
.	.	.
200	9953	0.087521

Weconsidered139publicationswithpaidadsand139publicationswithoutpaidads. The139postswithpaidadsresultin4,517shares,anaverageof32.49shares.Fornonpaid publications, 3,489 shares were accumulated as organic share reach, an average of 25.01 shares (Figure 10). The percentages of paid and nonpaid publications were quite close. The difference between paid and nonpaid publications in terms of share numbers was 8 sharesonaverage.Theimprovementofpublicationwithpaidadswas29.46%relativeto nonpaid publication. Paid ads had a 15.55% percentage of importance with respect to number ofshares.

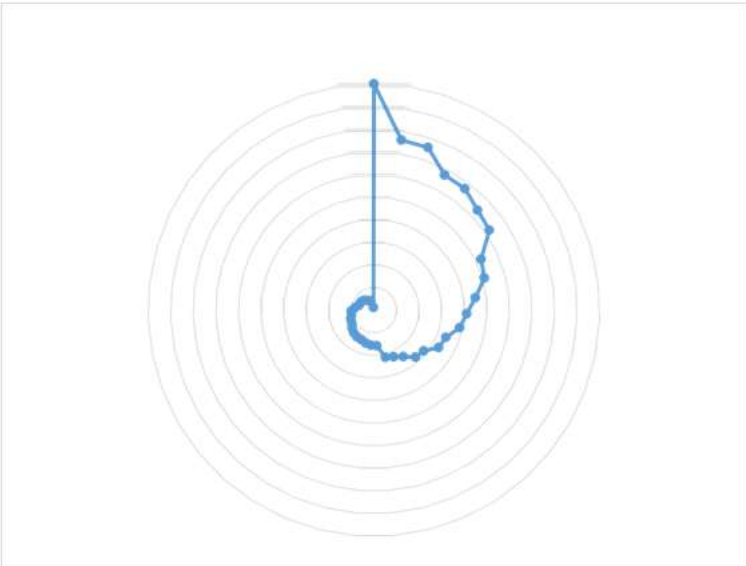

Figure 11 Relative importance of inputs toward total lifetime post reach (only for paid publications)

5. CONCLUSIONS

1. Paid ads contribute 9.52% of total relevance for total lifetime post reach. Post hour, type of publications, and post month have higher contributions to total lifetime post reach. This result provides useful information; instead of solely relying on paid ads, marketers should post the right material at the righttime.
2. Publications that used paid ads as a strategy do not have any perceptible impact on number of comments. Paid ads produce an average of 10.10 comments to each post and posts without paid ads average 6.01 comments per post. Type and category of publications have more significant effects on comments. Thus, understanding these aspects may increasecomments.
3. In terms of relevance, paid ads support the number of likes better than total lifetime post reach, comments, and shares and improve likes by 59% onaverage.
4. The importance of paid ads toward number of shares is 15.55%; shares are improved 29.46%ifpaidadsareused.Ifwefocusonthepercentageofrelevantinputstonumber oflikes,thencategoryofpublicationsisthemostimportantfactortonumberofshares.

   Therefore,tomaximizeshares,companiesshouldunderstandthecategoryofeachpost instead of only using paid ads as a publishingstrategy.
5. This research results suggest that companies need to understand their publication. Having paid ads is not a bad strategy for reaching potential customers, but combinations of paid ad strategies and other content strategies (time of publishing, type, category, having more pages likes) are highlyrecommended.

Even if company Facebook pages reach many people and generate excellent engagement,suchsuccessesdonotguaranteeanyrealsalesorrevenue.Difficultiesexists in measuring performance of such publications in terms of real sales because marketing touch points that affect customers' decisions are vague and hard to predict. The touch points that convert various customer behaviors into real sales should be studied. Based on such research, the allocation of budget to particular marketing strategies could be determined.Thisprocessiscalledmarketingattribution.Marketingattributionforonline advertising through company Facebook pages is a potential topic for futureresearch.

REFERENCES

[1] Anand, B.N. and Shachar, R., 2009. Targeted advertising as a signal. *QME*, *7*(3), pp.237-266.

[2] Bačík, R., Fedorko, R., Kakalejčík, L. and Pudło, P., 2015. The importance of Facebook ads in terms of online promotion. *Journal of Applied Economic Sciences*, 10(5),p.35.

[3] Barbier, G. and Liu, H., 2011. Data mining in social media. In *Social Networks data analytics* (pp. 327-352). Springer, Boston,MA.

[4] Bollen, J., Mao, H. and Zeng, X., 2011. Twitter mood predicts the stock market. *Journal of computational science*, *2*(1),pp.1-8.

[5] Cortes, C. and Vapnik, V., 1995. Support-vector networks. *Machine learning*, *20*(3),pp.273-297.

[6] Curran,K.,Graham,S.andTemple,C.,2011.AdvertisingonFacebook. *International Journal of E-business Development*, 1(1), pp.26-33.

[7] Garson, G.D., 1991. Interpreting neural-network connection weights. *AI expert*, *6*(4),pp.46-51.

[8] Goh, A.T., 1995. Back-propagation neural networks for modeling complex systems. *Artificial Intelligence in Engineering*, *9*(3),pp.143-151.

[9] Hanna,R.,Rohm,A.andCrittenden,V.L.,2011.We'reallconnected:Thepower of the social media ecosystem. *Business Horizons*, 54(3),pp.265-273.

[10] Krebs, F., Lubascher, B., Moers, T., Schaap, P. and Spanakis, G., 2017. Social Emotion Mining Techniques for Facebook Posts Reaction Prediction. *arXiv preprintarXiv:1712.03249*.

[11] Kriesel, D., 2007. *A Brief Introduction on NeuralNetworks*.

[12] Krishnan, K.S.T. and Sajilan, S., 2014. The Effects of Social Media on Gen Z's Intention to Select Private Universities in Malaysia. *Review of Integrative Business and Economics Research*, *3*(2),p.466.

[13] Lee, D., Hosanagar, K. and Nair, H.S., 2018. Advertising content and consumer engagement on social media: evidence from Facebook. *ManagementScience*.

[14] Margarida Barreto, A., 2013. Do users look at banner ads on Facebook?. Journal of Research in Interactive Marketing, 7(2),pp.119-139.

[15] Misra, B.B. and Dehuri, S., 2007. Functional link artificial neural Networks for classification task in datamining.

[16] Moro,S.,Rita,P.andVala,B.,2016.Predictingsocialmediaperformancemetrics and evaluation of the impact on brand building: A data mining approach.*Journal of Business Research*, *69*(9),pp.3341-3351.

[17] Needleman,S.E.andMarshall,J.,2014.Facebookadsbecomecostlierchoicefor small businesses. *The Wall StreetJournal*.
[18] Park,J.Y.,Sohn,Y.andMoon,S.,2016,March.PowerofEarnedAdvertisingon Social Networks Services: A Case Study of Friend Tagging on Facebook. In ICWSM (pp.299-308).
[19] Pelling, E.L. and White, K.M., 2009. The theory of planned behavior applied to young people's use of social Networksing web sites. *CyberPsychology & Behavior*, 12(6),pp.755-759.
[20] Stone,B.,2010.AdspostedonFacebookstrikesomeasoff-key.*NewYorkTimes,* 3.
[21] Turban, E., Sharda, R., Delen, D., & Efraim, T. (2011). *Decision support and business intelligence systems* (9th ed.).Pearson.
[22] We Are Social, 2018, "Digital in 2018: World's Internet Users Pass The 4 Billion Mark", https://wearesocial.com/blog/2018/01/global-digital-report-2018 (June 2018).
[23] Yuan, S., Wang, J. and Zhao, X., 2013, August. Real-time bidding for online advertising: measurement and analysis. In Proceedings of the Seventh International Workshop on Data Mining for Online Advertising (p. 3).ACM.