Vol. 9, No. 1, 2022, Pages 28-37

JIEMS

Journal of Industrial Engineering and Management Studies

journal homepage: www.jiems.icms.ac.ir

Social media target marketing: Use of social networks data to target marketing

Monireh Hosseini*[1], Zohreh Tammimy[1], Elnaz Galavi[1]

[1] Department of Information Technology, Faculty of Industrial Engineering, K. N. Toosi University of Technology, Tehran, Iran.

Received: Jun 2021-28/ Revised: Nov 2021-06/ Accepted: Nov 2021-16

Abstract

Social networks provide marketing managers and businesses with opportunity to target their customers. By understanding the demographics of users, marketing managers can offer suitable products and services. Although direct questioning can be drawn upon to solicit users' demographics such as age, some customers due to privacy concerns do not like to reveal their personal information and, it cannot come in handy for potential customer identification. The huge amount of data social networks generate can solve this problem. Previous studies in the prediction of demographic characteristics suffer some limitations because they were mainly text based and hence, language-bound. This study investigates how some interactive data can predict users' age. Further, it examines if classification methods can be used for age prediction. The results revealed that the number of friends, number of opposite sex friends, number of comments received, and number of photos which users share can predict users' age. Also, a linear relationship between interactive data and users' age was found.

Keywords: social media target marketing; age identification; social network sites; interactive data and profile information.

Paper Type: Original Research

1. Introduction

Marketing is a well-developed methodological science whose rules change continuously according to the needs and the developments (Saravanakumar and SuganthaLakshmi, 2012). The role of marketing in business development is clear, but the ways through which it is implemented changes radically.

In regular marketing, products or services are offered to all possible customers, and business owners try to sell their products or services to all people, regardless of their gender, age, location and income level (Friedmann, 2009). Today, in view of the increase of competition and globalization, people are inundated with huge number of offers, but they are only interested in what is relevant to them (Friedmann, 2009). So, target marketing as a way of marketing can be useful to enables business owners to create relevant offers (Friedmann, 2009). Target marketing considers all potential audiences for products or services and then breaks them to the smaller segments (Friedmann, 2009). Accurate identification of those group members enables business owners to create especial marketing campaigns to urge audiences to buy their products or services. This model is more effective and profitable than regular marketing (Friedmann, 2009).

In the other hand, due to the increasing popularity of social media such as social network sites (Kemp, 2019; Lanteri, 2019) , many different companies nowadays have started to use social media to promote their products and services which is called social media marketing. The social media brings all people the opportunity to create and distribute their contents easily (Zarrella, 2009). Whether a business is small, medium or large, its customers use social media; hence, there is no reason for business owners not to use social media (Zarrella, 2009). Social media marketing is almost free, easy to use (Zarrella, 2009), lower costs, and improved brand awareness (Dwivedi and et al., 2020). It also can have considerable financial effects on businesses (Zarrella, 2009). For example, Pepsi co. formulated new Mountain Dew, selling more than 36 million of it since 2008, based on customers' opinions collected from social media. In addition, the increasing use of social media generates a huge volume of data. Marketing managers and researchers believe that analyzing social media data and devising marketing strategies based on it can have positive effects on customers buying decisions and in turn, can create value for them (Hinz et al., 2014). However, significant challenges exist from negative electronic word-of-mouth as well as intrusive and irritating online brand presence (Dwivedi et al., 2020). By considering the features of mentioned marketing way and their advantages, providing a method which can be a useful solution for increasing the profitability of businesses and combat the challenges is critical. In order to create social media target marketing, marketing managers should understand who their customers are, what values and beliefs they have, what they think about services or products (Friedmann, 2009). So, it goes without saying that demographic characteristics are constructive and should be given attention to (Friedmann, 2009). Demographic characteristics are quantitative,

*Corresponding Author: monhosseini@gmail.com

DOI: 10.22116/jiems.2022.143488

and measurable entities like age, gender, and income level (Friedmann, 2009). Although direct questioning can be used to obtain demographic data, it suffers some limitations especially with the new social media environment. First, some customers, due to privacy concerns, do not like to reveal their personal information (Kumar et al., 2014; Zheleva and Gettor, 2009). Second, it is not possible to use direct questioning with potential customers. Previous studies in the prediction of demographic characteristics suffer some limitations because they were mainly text based and hence, language-bound. As a result, in this study we investigate if interactive data and profile information can predict users' age. In addition, if supervised learning methods can be used for age identification in social network sites. The study attempts to predict Facebook users' age using language-independent profile and interactive informations such as number of total friends, number of self-postings, number of opposite sex friends, number of received comments, number of photo-postings, and number of music and video postings. We use these interactive features because now we live in a multilingual world wherein linguistic data might not be readily available or, even if available, it might not accurately reflect the customers' behavior because of proficiency hindrances or cultural complexities (Kramsch, 2015). Additionally, the present paper preferred to study Facebook because it is the most popular among the social networking sites (Statista, 2021).
The reminder of this paper is organized as follows. A literature survey of the related works is provided in section 2. Research hypotheses are brought up in Section 3 and section 4 introduces our proposed framework for investigation the relationship between intended features. The empirical testing of framework and its results are presented in section 5.

2. Related works

Social media is an environment for social interaction in which people, i.e. web users, create, share, and/or exchange information in virtual communities and networks (Ahlquist et al., 2008). Social media can be classified into seven types including collaborative projects, blogs, content communities, social networking sites, virtual game world, virtual social world, and mobile social media (Zhang et al., 2015).
Among all kind of social media, the growing popularity of social networks provides businesses with the opportunities to identify their target community or potential customers (Appel et al., 2020). They are provided with the potentiality to better understand the audiences' personal information such as age to adopt their services or products, and to offer their products or services mire intelligently. According to Dwivedi et al. (2020), important topics in social media marketing are artificial intelligence, augmented reality marketing, digital content management, mobile marketing and advertising, B2B marketing, electronic word of mouth and ethical issues.
Several studies have examined the predictability of users' personal information using available social media data. Table 1 provides a brief overview of some of these studies.
Text- based features used as common features to predict the users' genuine information in previous studies (Argamon et al., 2007; Bamman et al., 2014; Hosseini and Tammimy, 2016; Kucukyilmaz et al., 2006, 2008; Marquadt et al, 2014; Mukherjee and Liu, 2010; Nguyen et al., 2011). For example, Hosseini and Tammimy (2016) used some linguistic features such as the number of pronouns, number of verbs, and number of articles to predict users' gender in social media. Nguyen et al. (2011) used content and stylistic features such as percentage of words longer than 6 letters for age prediction. Wang et al. (2019) used profile image, username, screen name, and biography which these are text-based and profile-based features for age, gender, and organization status. Morgan-Lopez et al. (2017) predicted Twitter users' age group using both metadata and language features.These studies notwithstanding their accuracy in the prediction of demographic characteristics, suffer some limitations because they were mainly text based and hence, language-bound. To compensate for this deficit, our research based on our knowledge, pioneers a new strand of research through which age is predicted using global features such as profile and interactive information. This research is placed on the artificial intelligence, digital content management and electronic word of mouth. Because proper and suitable advertising and massages can improve content and word of mouth marketing.

Table 1. Literature review of age and gender predictions

Author(s), date	Research focus	Used feature(s)	Finding
Argamon et al. (2007)	Age and gender differences in content and function words Age and gender prediction by using MBR, WIN	Text-based	• By using function words and 1000 words with highest information- gain predict age and gender with good accuracy • Age and gender are closely related
Kucukyilmaz et al. (2008)	Possibility of predicting several user and message attributes like age, gender, and receiver, so on and evaluation of applicability of various supervised classification techniques	Text-based	• Personal and environmental characteristics have significant impact on ones' vocabulary use and writing style in peer-to-peer communication • By using the word selection patterns and stylistics preferences of chat users, it is possible to predict their sociolinguistic characteristics by employing classification techniques
Peersman, et al. (2011)	Prediction of age and gender on a corpus of chat texts	Text-based	• It is feasible to improve upon random baseline performance for age classification using highly limited data sets • Gender could be a helpful information source in constructing a more accurate classifier for age
Nguyen, et al. (2011)	Author age prediction from text	Text-based and gender	• Content features and stylistic features such as percentage of words longer than 6 letters to be strong indicators of a person's age
Marquardt et al. (2014)	Predictive quality in terms of age and gender of several sets of features extracted from various genres of online social media	Text-based, HTML Tags and total number of posts by a user)	• In difference genres requires multiple categories of features • The accuracy of predicting age gained by using a more complicated classification scheme such as chained classifiers to be negligible
Rangel and Rosso (2016)	Investigation the impact of emotions on author profiling and age and gender identification	Text-based	• most discriminative features for age and gender identification were recognized • modeling the language based on proposed features in article and identified age and gender

3.Research hypotheses

Pfeil et al. (2009) have analyzed the differences between users in social network sites and have noted that teenagers tend to have more friends than older people on Myspace. In addition, they found that teenagers use social networking site to connect to more opposite sex friends and acquaintances. Thus, we hypothesize that:

H1. Number of friends can predict user' age group in social network site.

H2. Number of friends with opposite sex can predict user' age group in social network site.

In addition, Pfeil et al. (2009) showed that teenagers use more music and videos on their social pages compared to older people and teenagers received on average almost ten times as many comments as do older people and. Therefore, we hypothesize:

H3. Number of music and video posting can predict user' age group in social network site.

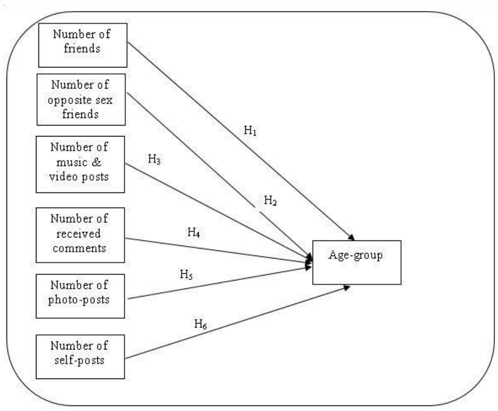
H4. Number of received comments can predict user' age group in social network site. According to Madden et al. (2013), teenagers share more photo of themselves in social media. Thus, we hypothesize:

H5. Number of photo-posting which users have shared can predict age-group of users in social network site.

Moreover, Chang et al. (2015) have noted that Facebook users age have negative related with checking the information about others and self-posting behavior. Thus, we hypothesize that:

H6. Number of posts which users have shared about themselves can predict age group of users in social network site.

Figure 1 represents the research model of this study.



4. Research design

As depicted in figure 2, a framework was postulated to investigate the relationship between intended features which are tabulated in table 2, and users' age in social network site. The framework consists of two phases namely hypothesis testing and classifiers evaluation. In the following subsections, we provide more description about the phases.

4.1. Hypothesis testing

In the hypothesis testing phase, initially, data was pre-processed to find outliers and influential cases. Then, in order to test our hypotheses, logistic regression was employed to test hypothesis.

$$p(y) = \frac{1}{1 + e^{-(b_0 + b_1 X_{1i})}} \quad (1)$$

Based on Field (2009), a great deal of care should be taken in selecting independent variablesfor a model because the values of the regression coefficients depend upon the variables in the model. Therefore, the independent variables included and the way in which they are entered the model can have a great impact (field, 2009). There are several ways in which variables can be entered into the model. Backward method was used in this study to test our hypotheses. In backward method computer begins by placing all variables in the model and then if an independent variable meets the removal criterion (i.e. if it is not making a statistically significant contribution to how well the model predicts the dependent variable) it is removed from the model and the model is re-estimated for the remaining variables (Field, 2009).

4.2 Classifiers evaluation

In classifiers evaluation phase, the effective features which had been identified in the previous phase were used to investigate the performance of some the most influential supervised learning methods. Support Vector Machine (SVM), k-nearest neighbor, logistic regression, and decision tree (C 4.5) were used to build classification models. According to Wu et al. (2008), the aforementioned algorithms are among the most influential data mining algorithms. In order to evaluate classifiers, 10-fold cross validation method was employed to split the dataset into train and test sets. 10-fold-cross validation randomly divides the full data set into 10 parts, in which the class represented in approximately the same proportions as in the full data set. The model trained on nine parts and evaluate on one part. The procedure executed 10 times so, each instance has been used exactly once for testing (Witten et al., 2016).
In order to evaluate the different classifiers, accuracy, precision, recall, and F1-measure were used. These evaluation measures have been defined in formulas 2 to 5 and based on the confusion matrix which is shown in figure3:

$$accuracy = \frac{TP+TN}{P+N} \quad (2)$$

$$precision = \frac{TP}{TP+FP} \quad (3)$$

$$recall = \frac{TP}{TP+FN} \quad (4)$$

$$f1 - measure = \frac{2 \times precision \times recall}{precision + recall} \quad (5)$$

	Predicted class		
Actual Class		Class = Yes	Class = No
	Class = Yes	True Positive	False Negative
	Class = No	False Positive	True Negative

Figure 3. Confusion matrix

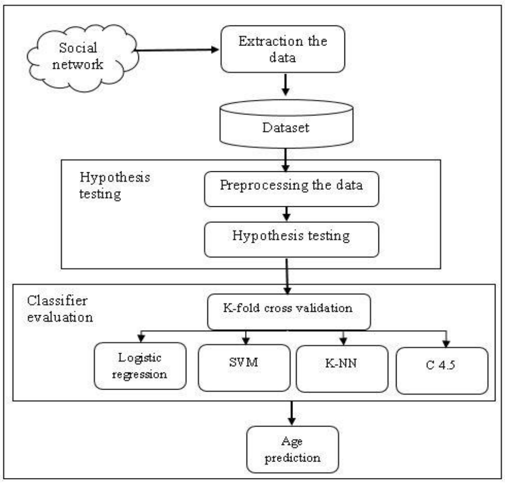


Figure 3. Proposed framework for age identification

5. Experimental study

5.1. Dataset

We conducted an empirical study on Facebook which is one of the most popular social network sites.
Since national culture has effect on the individuals' behavior (Erumban and De Jong, 2006), we have chosen Iranian Facebook users to minimize the effect of culture differences on the results.
The users' interactive data and profile information was collected by convenience sampling method. In order to protect the privacy of users, public data of users has been collected and the collected data has been saved without any identification information.
In total, the profile and interactive information of 103 Iranian Facebook users were collected. The data set includes the data of 62 women and 41 men in the age range of 17-53. As shown in the table 2, the dataset includes the users' age, number of friends, number of opposite sex friends, number of received comments, number of posts which user has shared about himself, number of images which user has shared, and the number of music and video content which user has shared. Based on Miles and Shevlin (2001), a sample size of 100 will be fine to find a medium effect with 6 predictors. However, the users' age is a continuous feature, but based on our aims in this research, we considered it as a categorical feature. According to Hu et al. (2007), the users' age can be classified into 5 classes which are listed in table 3.

Table 2. Features in the dataset

#	features	Description
1	Number of friends	The total number of users' friend
2	Number of opposite sex friends	The total number of users' friend with opposite sex
3	Number of received comment	For 10 current posts of user
4	Number of posts which user has shared about himself	During 2 months
5	Number of images which user has shared	During 2 months
6	Number of music and video content which user has shared	During 2 months

Table 3. Age classes

#	Age	Class label
1	<18	Teenage
2	18-24	Young
3	25-34	Young-adult
4	35-49	Middle-age
5	>49	Senior

5.2. Results

5.2.1. Age identification hypothesis testing

As mentioned above, multinomial logistic regression was employed to test hypotheses. Logistic regression has some assumptions which should be met. Linearity have checked the linear relationship between continuous independent variables and logit of outcome variable which can be tested by defining the interaction term and its log transformation. Independence of errors has noted that the cases of data should not be related and since the data of different individuals were collected, this assumption has met. In order to check multicollinearity, VIF and

Tolerance statistics were obtained. According to Menard (2002), a tolerance value less than 0.1 indicates a serious collinearity problem. In addition, Mayers and Mayers (1990) suggested that a VIF value greater than 10 is cause of concerns for collinearity. As shown in table 4, multicollinearity did not occur in our dataset.

As shown in the table 4, the factors which can predict the user's age in social network site were the number of friends ($p < .05$), the number of opposite sex friends ($p < .05$), the number of comments which user has received ($p < .05$), and the number of photo-posting which user has shared ($p < .05$). According to the obtained results, the hypotheses H1, H2, H4, and H5 were supported so the research model was transformed to the model which represents in figure 4.

Table 4. Regression coefficients and significance level of intended features

Class label	Chi-square	df	Sig.	Collinearity Statistics	
				Tolerance	VIF
friends	37.155	4	.000	.382	2.621
Opposite sex friends	36.52	4	.000	.299	3.343
Self- posting	6.43	4	.169	.566	1.765
Music and video posting	4.494	4	.293	.927	1.078
Received comments	22.92	4	.000	.502	1.991
Photo-posting	23.778	4	.000	.918	1.089

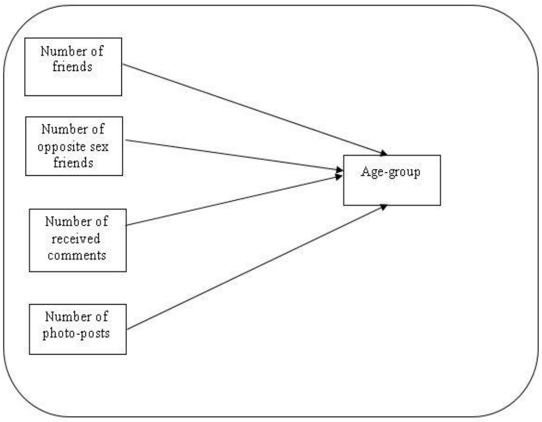


Figure 4. Edited research model

5.2.2. Age identification classifier evaluation phase

Based on the effective features which were identified in the previous phase, four different supervised learning algorithms such as SVM, logistic regression, C 4.5, and k-nearest neighbor were investigated.

Since it is possible to use different kernels in SVM, all of the kernels were investigated to find out which one outperform the others. The results of using different kernels in SVM method are shown in table 6.

Table 5. The performance measures of different kernels in SVM

Kernel	Accuracy	Precision	Recall	F1-measure
Linear	**74.76%**	**76%**	**75%**	**74%**
Multinomial	38.84%	43%	39%	40%
RBF	48.54%	24%	49%	32%
Sigmoid	48.54%	24%	49%	32%

According to the obtained results of using different kernels in SVM, the model showed that linear kernel performs better than others.

In the other hand, since the value of parameter K in the k-nearest neighbor method has impact on the model (Kahn and Johnson, 2013), different values of K were tried to find the optimal value. As shown in the figure 5, the best performance of k- nearest neighbor method was obtained by the value of five for parameter k.

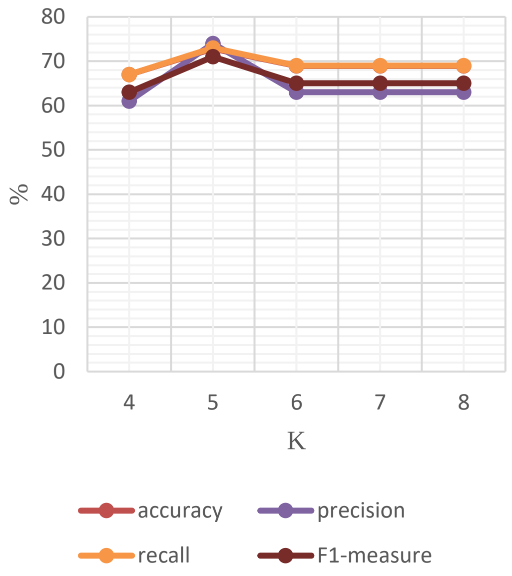


Figure 5. Performance measures of K-nearest neighbor method by different k values

Totally the performance measures of SVM method with linear kernel, k-nearest neighbor method by k value equal 5, C 4.5 method, and logistic regression method were compared which the results have been represented in table 6. Based on the obtained results, logistic regression method outperforms the others so it can conclude that there is a linear relationship between aforementioned interactive features and user's age.

Table 6. The performance measures of used supervised learning methods

Method	Accuracy	Precision	Recall	F1-measure
Logistic regression	77.67%	78%	78%	76%
SVM	74.76%	76%	75%	74%
K-nearest neighbor	72.82%	74%	73%	70%
C 4.5	57.28%	54%	57%	56%

In addition, different interactive features have different affect in the different age groups in logistic regression which this phenomenon has reflected in the regression coefficients. Table 7 has shown these values.

Table 7. The regression coefficients of age prediction

	Number of friends	Number of opposite sex friends	Number of received comments	Number of photo posts
Teenage	0	0.01	0	0.01
Young	-0.01	0.02	-0.02	0.16
Young-adult	0.01	-0.02	0.01	-0.19
Middle-age	-0.01	0.01	0	-0.18
Senior	-0.06	-0.31	0	-13.74

The number of photo-posting which users have shared in social network site was the most important feature among others for age identification problem based on the regression coefficients.
Based on the results, the logistic regression equations for users' age prediction for each age group are defined as below:

$$P(teenage) = \frac{1}{1+e^{-(b_0+0.01X_2+0.01X_4)}} \quad (6)$$

$$P(young) = \frac{1}{1+e^{-(b_0-0.01X_1+0.02X_2-0.02X_3+0.16X_4)}} \quad (7)$$

$$P(young - adult) = \frac{1}{1+e^{-(b_0+0.01X_1-0.02X_2+0.01X_3-0.19X_4)}} \quad (8)$$

$$P(middle - age) = \frac{1}{1+e^{-(b_0-0.01X_1+0.01X_2-0.18X_4)}} \quad (9)$$

$$P(senior) = \frac{1}{1+e^{-(b_0-0.06X_1-0.31X_2-13.74X_4)}} \quad (10)$$

6. Discussion

Our proposed method achieved 77.76 percent accuracy which in comparison to the other studies performs better. The comparison of accuracies and precision which were achieved by other studies and our proposed method is shown in Table 8. Kucukyilmaz et al. (2008) could predict users' age with 75.4 percent for two classes of ages (before 1976 and after that) and with an increase in the number of classes to four, the resulted accuracy was 37.4 percent.

Peersman et al. (2011) reached the accuracy of 88.8 percent but they only considered two classes for age (adolescents, adult) which cannot describe the users' age with enough detail. As mentioned, before we used five age classes for describing with more detail users ages.

Table 8. Prediction accuracies and precision of studies for age prediction

Authors. date	Accuracies and precision
Argamon et al., 2007	Accuracy =77.4%
Kucukyilmaz et al., 2008	Accuracy 2-class: 75.4% Accuracy 4-class: 37.4%
Peersman et al., 2011	Accuracy= 88.8%
Marquardt et al., 2014	Accuracy English: 46.89% Accuracy Spanish: 48.31%
Rangel and Rosso, 2016	Accuracy= 66.24%
Morgan- Lopez et al., 2017	Precision of Text-bead features and metadata= 74% Precision of Only metadata= 58%
Our work	Accuracy= 77.76% Precision= 78%

7. Conclusion

Social media and social network sites provide some opportunities for businesses. Social media marketing is one of the opportunities which is free, easy to use, and financially effective. In addition, social network sites with great number of users are data sources which can help marketing managers and businesses to target their customers. However direct questioning about users' personal information and preferences can be used to obtain users' personal data and preferences, but some users due to privacy concerns do not like to reveal their right personal information and also direct questioning is not applicable for targeting potential customers. Hence in this study, we firstly investigated if users' interactive data and profile information can predict users' age. Secondly, if classification methods can predict users' age accurately. The results revealed that users' age can be identified by some interactive data and profile information such as number of friends, number of opposite sex friends, number of received comments, and the number of photos which users shared in their social pages. In addition, the results of evaluation of different classification methods have shown that there is a linear relationship between aforementioned data and users' age. Logistic regression as a technique at the core of variable rule analysis performed better than other methods. Based on some research culture has effect on the behavior, as a result we restricted our dataset to the data of users with same national culture.

Limitation

The proposed method was tested using limited dataset, and as result it tested medium effect.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Ahlqvist, T., Bäck, A., Halonen, M., and Heinonen, S., (2008). "Social media road maps exploring the futures triggered by social media", VTT Tiedotteita-Valtion Teknillinen Tutkimuskeskus.

Appel, G., Grewal, L., Hadi, R., and Stephen, A. T., (2020). "The future of social media in marketing", Journal of the Academy of Marketing Science, Vol. 48, No. 1, pp. 79-95.

Argamon, S., Koppel, M., Pennebaker, J. W., and Schler, J., (2007). "Mining the blogosphere: Age, gender and the varieties of self-expression", First Monday, Vol. 12, No. 9.

Bamman, D., Eisenstein, J., and Schnoebelen, T., (2014). "Gender identity and lexical variation in social media", Journal of Sociolinguistics, Vol. 18, No. 2, pp. 135-160.

Chang, P. F., Choi, Y. H., Bazarova, N. N., and Löckenhoff, C. E., (2015). "Age differences in online social networking: Extending socioemotional selectivity theory to social network sites", Journal of Broadcasting & Electronic Media, Vol. 59, No. 2, pp. 221-239.

Dwivedi, Y.K., Ismagilova, E., Hughes, D.L., Carlson, J., Filieri, R., Jacobson, J., Jain, V., Karjaluoto, H., Kefi, H., Krishen, A.S. and Kumar, V., (2020). "Setting the future of digital and social media marketing research: Perspectives and research propositions", International Journal of Information Management, p.102168.

Erumban, A. A., and De Jong, S. B., (2006). "Cross-country differences in ICT adoption: A consequence of Culture?", Journal of world business, Vol. 41, No. 4, pp. 302-314.

Field, A., (2009). Discovering statistics using SPSS, Sage publications.

Friedmann, S., (2009). The complete idiot's guide to target marketing, Penguin.

Hinz, O., Schulze, C., and Takac, C., (2014). "New product adoption in social networks: Why direction matters", Journal of Business Research, Vol. 67, No. 1, pp. 2836-2844.

Hosseini, M., and Tammimy, Z., (2016). "Recognizing users' gender in social media using linguistic features", Computers in Human Behavior, Vol. 56, pp. 192-197.

Hu, J., Zeng, H. J., Li, H., Niu, C., & Chen, Z. (2007, May). Demographic prediction based on user's browsing behavior. In Proceedings of the 16th international conference on World Wide Web (pp. 151-160).

Kemp, S. (2019). Digital 2019: Global social media users pass 3.5 billion.

Kramsch, C., (2015). "Applied linguistics: A theory of the practice", Applied linguistics, Vol. 36, No. 4, pp. 454-465.

Kucukyilmaz, T., Cambazoglu, B. B., Aykanat, C., & Can, F. (2006, October). Chat mining for gender prediction. In International conference on advances in information systems (pp. 274-283). Springer, Berlin, Heidelberg.

Kuhn, M., and Johnson, K., (2013). Applied predictive modeling (Vol. 26), Springer.

Kumar, S., Morstatter, F., and Liu, H., (2014). Twitter data analytics, Springer.

Lanteri, S., (2019). Persona Spotlight: The Online Shopper. Global web index. Retrieved from https://blog.globalwebindex.com/marketing/persona-spotlight-the-online-shopper/

Madden, M., Lenhart, A., Cortesi, S., Gasser, U., Duggan, M., Smith, A., and Beaton, M., (2013). "Teens, social media, and privacy", Pew Research Center, Vol. 21, pp. 2-86.

Marquardt, J., Farnadi, G., Vasudevan, G., Moens, M. F., Davalos, S., Teredesai, A., & De Cock, M. (2014). Age and gender identification in social media. Proceedings of CLEF 2014 Evaluation Labs, 1180, 1129-1136.

Menard, S. (2002). Applied logistic regression analysis (Vol. 106). Sage.

Shevlin, M., & Miles, J. (2000). Applying regression and correlation: A guide for students and researchers. Applying Regression and Correlation, 1-272.

Morgan-Lopez, A. A., Kim, A. E., Chew, R. F., and Ruddle, P., (2017). "Predicting age groups of Twitter users based on language and metadata features", PloS one, Vol. 12, No. 8, e0183537.

Mukherjee, A., & Liu, B. (2010, October). Improving gender classification of blog authors. In Proceedings of the 2010 conference on Empirical Methods in natural Language Processing (pp. 207-217).

Myers, R. H., and Myers, R. H., (1990). Classical and modern regression with applications (Vol. 2), Duxbury press Belmont, CA.

Nguyen, D., Smith, N. A., & Rose, C. (2011, June). Author age prediction from text using linear regression. In Proceedings of the 5th ACL-HLT workshop on language technology for cultural heritage, social sciences, and humanities (pp. 115-123).

Peersman, C., Daelemans, W., & Van Vaerenbergh, L. (2011, October). Predicting age and gender in online social networks. In Proceedings of the 3rd international workshop on Search and mining user-generated contents (pp. 37-44).

Pfeil, U., Arjan, R., and Zaphiris, P., (2009). "Age differences in online social networking–A study of user profiles and the social capital divide among teenagers and older users in MySpace", Computers in Human Behavior, Vol. 25, No. 3, pp. 643-654.

Rangel, F., and Rosso, P., (2016). "On the impact of emotions on author profiling", Information processing & management, Vol. 52, No. 1, pp. 73-92.

Saravanakumar, M., and SuganthaLakshmi, T., (2012). "Social media marketing", Life Science Journal, Vol. 9, No. 4, pp. 4444-4451.

Clement, J. (2020). Global social networks ranked by number of users 2020. Statista. com.

Tagliamonte, S. A. (2006). Analysing sociolinguistic variation. Cambridge University Press.

Wang, Z., Hale, S., Adelani, D. I., Grabowicz, P., Hartman, T., Flöck, F., & Jurgens, D. (2019, May). Demographic inference and representative population estimate from multilingual social media data. In The world wide web conference (pp. 2056-2067).

Witten, I. H., & Frank, E. (2002). Data mining: practical machine learning tools and techniques with Java implementations. Acm Sigmod Record, 31(1), 76-77.

Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., and Philip, S. Y., (2008). "Top 10 algorithms in data mining", Knowledge and information systems, Vol. 14, No. 1, pp. 1-37.

Zarrella, D. (2009). The social media marketing book. " O'Reilly Media, Inc.".

Zhang, X., Wang, W., de Pablos, P. O., Tang, J., and Yan, X., (2015). "Mapping development of social media research through different disciplines: Collaborative learning in management and computer science", Computers in Human Behavior, Vol. 51, pp. 1142-1153.

Zheleva, E., & Getoor, L. (2009, April). To join or not to join: the illusion of privacy in social networks with mixed public and private user profiles. In Proceedings of the 18th international conference on World wide web (pp. 531-540).

This article can be cited: Hosseini, M., Tammimy, Z., (2022). Social media target marketing: Use of social networks data to target marketing, Journal of Industrial Engineering and Management Studies, Vol.9, No. 1, pp. 28-37.