5

An Analysis of Demographic and Behavior Trends Using Social Media: Facebook, Twitter, and Instagram

Amandeep Singh, Malka N. Halgamuge, Beulah Moses
School of Computing and Mathematics, Charles Sturt University, Melbourne, VIC, Australia

1 Introduction

Technology has become a very important part of everyone's life. Everyone from the age of 5 to 65 years is on social media every day with billions of users sending messages, sharing information, comments, and the like [1]. With the advancement in information technology, social networking sites such as Facebook, Twitter, Instagram, and LinkedIn are available for the users to interact with families, colleagues, and friends. As a result, social activities are shifting from real things to virtual machines [2]. Analyzing the behavior of individuals from social networking sites is a complex task because there are several methods used. By gathering information from different resources and then analyzing that information, the behavior of the users can be examined. In this research, we have collected different studies about assessing human behavior with the help of social media and compared them according to the different methods used by different authors [3].

To know the personal preference of the users on the social media is a very important task for businesses [4]. Companies can then target those interested customers who are active on the social media in related areas. By gathering information about user behavior pattern, the preferences of the individuals can be identified [5]. Different researchers have found various methods to collect information about human intentions. In this research, our main aim is to analyze how information is analyzed in social media and how this information is useful. This research is very useful as methods to detect human behavior that has been analyzed on different social media [6].

In this research, 30 research papers have been collected from different social media providers such as Facebook, Instagram, and Twitter. After analyzing the data given in these papers, the different methods used were examined. In particular, the behavior of users was analyzed from aspects such as likes, comment, and shares from Facebook, Instagram, and Twitter [7].

The first section provides the material and methods used by the 30 authors to predict the behavior of social media users. This section included data collection, data inclusion

Social Network Analytics. https://doi.org/10.1016/B978-0-12-815458-8.00005-0

criteria, and data analysis [8]. The next section is the result section which provides the statistical analysis and the percentage of research completed on different social media. The result section includes a table which provides the research paper analysis according to the year along with pie chart figures, data collection, and behavior analysis methods and classifications based on different methods with line graphs [9]. The next section is a discussion on the given topic and the last section is the conclusion of this research work.

2 Material and Methods

Data were collected from different conference papers published in the IEEE. From these papers, different methods of analyzing the user behavior [10] was assessed. This report is based on a review of the published articles and analyzes the methods they have used. The data are given in a tabular form.

2.1 Data Collection

Data were collected from 30 various journal papers from the IEEE library regarding the analysis of the user behavior using social media from 2015 to 2017. The collected data were related to Facebook, Twitter, and Instagram in different countries [11]. The attributes that were used for data collection were: applications, methods used, description of the method, number of users, limitations, and results. This raw data is presented in Table 1 [32].

2.2 Data Inclusion Criteria

The different data attributes used to analyze the papers are given in Table 1. This included the following: author name, applications, methods used, detail of methods, number of users, limitations, and results. Data were gathered relating to different social networking sites [17]. In our analysis, the different methods that have been used by researchers to analyze the user's behavior are explored. In this research, three different social media datasets have been collected, which represents the methods and technologies used to understand the behavior of the users.

2.3 Analysis of Raw Data

The raw data presented in Table 1 specifies the attributes that were used to conduct this research. We pooled and analyzed 30 studies based on the impact of variables used in their studies. The descriptive details of the study based on the publication year were then analyzed to observe the behavior of the social media user from 2015 to 2017. A comparison of the methods they used to investigate the behavior of users was then done.

This research included papers from the last 3 years from 2015 to 2017. All papers used data from Facebook, Instagram, and Twitter.

Table 1 Behavior Analysis Using Social Media Data Extracted From 30 Scientific Research Papers During 2015–17

No.	Study	Social Media	Methods/Technologies Used	Description	Users	Limitations	Results
1.	Park et al. [9]	Instagram	Snowballing method Coding rules: binary coding Regression analysis	Quantitative method is used to analyze the relationship between sexual images and social engagement Number of likes were used Snowballing method was used to collect people's image data Binary coding rule was used to self-code the images collected Regression analysis was used to analysis the behavior of 200 users	200	Data does not show that who and why people get more interested in sexual images Causal relationship cannot use to prove the relationship between sexual images and number of likes	With number of likes degree of sexuality is known in the given images Results show that men and women get more like when they upload their selfies
2.	Farahani et al. [12]	Twitter	Regression and correlation methods were used Mean absolute percentage error (MAPE) and mean absolute error (MAE) Gaussian mixture model	Metrics were used to analyze the different behavior in different dimensions Data are collected and filtered on the basis of Iran election with maximum tweets Gaussian mixture model (GMM) was used to detect influential users	Top 20 users with 148,713 tweets	Correlation of other influence measures were not evaluated No prediction algorithm No weighted measure technique	Original Tweets (OTI) is very important Results shows that OTI and metrics play very crucial role Retweet impact has 3.9 MAE and 0.12 MAPE RT2 has 7.12 MPE and 0.0 MAPE RT3 has 4.5 MAE and 0.13 MAPE
3.	Castro et al. [13]	Twitter	Social network analysis techniques Machine learning Partition clustering	Methods were used to detect the political behavior of Venezuelan election	60,000	Political alignment of entire state was not determined Tweets were not	Average score of discriminative political features in both clusters were compared

Continued

Table 1 Behavior Analysis Using Social Media Data Extracted From 30 Scientific Research Papers During 2015–17—cont'd

No.	Study	Social Media	Methods/Technologies Used	Description	Users	Limitations	Results
			algorithm Text processing Term frequency-inverse document frequency (TF-IDF)	Clustering algorithm was used to analyze to citizen public speech Twitter communication pattern and linguistic dimension used For each tweet, unique identifier, publication date, geographical location, and tweet contents were used for analysis TF-IDF provides the score that gives how words are relevant to texts		analyzed in different time windows Different weighing alternatives to TF-IDF and geographical subdivision were not considered	Cluster state 1 represent opposition state and 0 represent government state TF-IDF represents 79.17% accuracy in election outcomes
4.	Mungen et al. [14]	Instagram	Fuse-motif analysis Mungan and Kaya's network motif Combination of Triad FG, motif-bases social position and quad closure methods	Motif-based analysis is based on the posts by the Instagram users Most influenced posts were used instead of most influenced person System calculates all language pairs Unique model was split in three different models such as creating graph, find most influenced people and influencing users only	20,000	Other factors related to Instagram are not considered such as images data, shared, and comments Models are complicated and hard to understand by the normal users	Result shows that four normal motifs have largest impact of 3.9 among others and with 22% frequency 1 norm-2 mid-1 pop motif has lowest impact of 1.2 with 2% frequency

5.	Wiradinata et al. [15]	Instagram	Path diagram model analysis Technology acceptance model (TAM)	TAM is used to know the factors for the acceptance of any system and data collection method used for sampling Path diagram model is used with statistical software AMOS 20 to know the behavior of consumers in small medium enterprise (SME) those using Instagram	200	Complex path analysis model as it uses normality testing, validity testing, reliability testing	Exogenous variables: technology-specific valuation (TSV), number of users (NOU), and perceived ease of use (PEOU) have influence (direct or indirect) on the endogenous variables Intervening variable perceived usefulness (PU) have influence on endogenous variable to behavioral influence (BI)
6.	Jiang et al. [11]	Twitter	CrossSpot algorithm Suspiciousness metric. Multimode Erdos-Renyi model KL-divergence principle Minimum Description Length (MDL) Principle	Fraud detection in twitter is main aim of this research Tensors was used to represents counts of events Suspicious metric is derived based on ERP model Five axioms were used to predict the behavior of different users	225	Metric based on more sophisticated model is not included	CrossSpot has more suspiciousness score than HOSVD in case of retweeting and hashing data
7.	Nasim et al. [16]	Facebook	Binary classification problem Simmel's theory Sociological theory "foci" behind friendship formation Facebook data was provided with Algopol project	Impact of additional interaction information were studied Binary classification problem was used for the link interface problem Third-party apps have access to the user profile and these can be used to	586 users and comments posted by 6400 users	Privacy is major concern Algorithm for news feed is not known Filtering is not done properly [16]	It has been observed that individuals who are friends with each others have similar interests Two evaluation metrics were used to judge the performance of classifier ROC and PR used to

Continued

Table 1 Behavior Analysis Using Social Media Data Extracted From 30 Scientific Research Papers During 2015–17—cont'd

No.	Study	Social Media	Methods/Technologies Used	Description	Users	Limitations	Results
			Tenfold cross validation method LR, linear discriminant analysis (LDA) and support vector machines (SVM)	access the information of users Theories suggest that interaction will take place among users who are connected			calculate different measures
8.	Jarvinen et al. [17]	Instagram	Partial least square (PLS) SmartPlus3 Software Path-weighing scheme	It is an extended version of UTAUT2 Model Out of 199 responses 187 respondents were used Given conceptual model was tested with SmartPlus3 software Hedonic motivation is important to derive consumer's intension to continue using SNS 86.6% users were used from Europe to analysis [17]	187	Generalizability is major limitation It is not sure that the sample shows the number of SNS users Data is not correct in terms of origin of users Results cannot be applied to global Instagram users	Variances explained in behavioral intension is 67% and use behavior is 46% which is higher compared to UTAUT2 model
9.	Geeta et al. [10]	Twitter	Demographic analysis	Data are collected from the tweets by users Opinions of different users have been analyzed and then sentiment analysis is performed and at the end demographic analysis is achieved to get the required data	30 million	Current location of users is not identified. So, it is not clear that user tweets from the real location or not	Result shows the opinions of users in five different countries United States has high percentage of tweets done in Oscar event, India has high percentage of tweets in T20 event, France user tweet more on Paris attack and Australian users tweets high on formula 1

10.	Dalton et al. [18]	Twitter	JavaFX application	JavaFX is the improvement in Maltego in which automation of entity is not possible JavaFX automation uses text file and MySQL database as input and produce results	5000	Reliability of IP	Automation is possible in terms of more flexibility and speed
11.	Hosseinmardi et al. [6]	Instagram	Fivefold cross validation Logistic regression classifier with forward feature selection approach	Data are extracted from the initial posts LRC was used to train a predictor To analyze the behavior comment, images and followers on Instagram was used Focus was on unigram and bigrams	25,000	Performance in classifier needs to be improved Deep learning and neutral learning was not used Less input features Comments on previous shares was not used	This method achieves high performance in predicting behavior 80% data used for training and 20% for testing 0.68 recall and 0.50 precision was used to detect behavior
12.	Chinchilla et al. [19]	Instagram	Cross industry standard process for data mining (CRISP-DM) Clustering and association rules	CRISP-DM is designed for hierarchical process model Data are collected from the 1435 records and after analyzing the data behavior of customers is evaluated	1435	Data mining models are limited to only Instagram and Facebook and other companies cannot use this data for behavior analysis	According to different clusters, attribute like has high number then others and TIPO_MODA is last in the numbers
13.	Dewan et al. [20]	Facebook pages	Supervised learning algorithms. Bag of words Crowdsourcing technique: web of trust (WOT)	Like, comment and share are analyzed, and textual contents was collected from three sources: message, name, and link Bag of words produced sparse vector and this vector used for classification	627 FB Pages and most recent 100 posts	Large group and events were not covered Bag of words is based on limited history of 100 posts Pages can change behavior over time	Results are based on the different classifier and it is concluded that Neural Network classifier of Trigram feature set has high rate of accuracy of 84.13%

Continued

Table 1 Behavior Analysis Using Social Media Data Extracted From 30 Scientific Research Papers During 2015–17—cont'd

No.	Study	Social Media	Methods/Technologies Used	Description	Users	Limitations	Results
14.	Toujani et al. [21]	Facebook	Fuzzy sentiment classification Fuzzy SVM I told you application to get old FB messages	Opinion mining was performed on the FB users Given system is consist of four phases: input (I), natural language processing (NLP), machine learning process (MLP), output (O) to produce the desired results Investigation is performed on the basis of coordination between machine learning and NLP	260	This system can only be used where users know Arabic and French/ English language No mobility-based machine learning	In basic SVM, 74% of precision and F-measure for positive opinion, whereas in Fuzzy SVM, 88.2%
15.	Lukito et al. [2]	Twitter	Analyze and comparison of three different statistical models Questionnaire based on big five personality inventory Java program was developed to answer the questions	Machine learning, lexicon based, grammatical rule approach models were used for analysis with comparison Data were collected from Facebook profile and then used twitter based on this data to predict the behavior	142 users with 2,00,000 tweets	Accuracy is low in term of IE, SN, and TF personality treats Variable accuracy	Based on the bar graph machine learning approach has high accuracy in IE personality factor, grammatical rule has high accuracy in SN factor, etc.

16.	Santos et al. [22]	Twitter	Visualization Different computational techniques Set of keywords was used for data collection	Data collected based on the tweets on World Cup 2014 which was held in 2014 in Brazil Visual system used to recognize patterns, spot trends and identify outliers Data and text mining and natural processing computational techniques were used After data pre-process, visualization was designed and handed over to journalists	851,292 tweets	Manual data process Emotions were not included This method has not included different versions Analyze data is complex process	Analysis was based on the focus group discussions on two major aspects such as journalism criteria and visualization techniques Graph A to Graph D was used to visualization
17.	Rabab'ah et al. [23]	Twitter	Twitter tweepy API tools was used for data collection NetworkX-METIS package was used to partition the retweet graph	Controversy level is identified with the help of tweets on social contents of Arabic language Data are collected from Twitter from September to October 2015 with hashtags on the trending topics Retweet graph was designed based on tweets and then retweet graph is portioned by removing noisy nodes Controversy measures RW, EC, and GMCK are applied	1.5 million tweets	This method is dependent on the structure of interaction between participants in the conversation only Retweet activity and retweet graph are only focus areas Other ways can be considered than these graphs	Controversy level was measures using random walk (RW) and embedded controversy (EC) Figures elaborate that in RW, most controversial topic is T6 with 0.822 value In GMCK, T6 is most controversial and in EC T6 is most controversial

Continued

Table 1 Behavior Analysis Using Social Media Data Extracted From 30 Scientific Research Papers During 2015–17—cont'd

No.	Study	Social Media	Methods/Technologies Used	Description	Users	Limitations	Results
18.	Lima et al. [24]	Twitter	Machine learning and text mining techniques Sentiment analysis Personality prediction David Keirsey classifications Myers Briggs type indicator test	Temperament predictor was designed to assess the individual behavior Message from twitter was captured using MBTI test 16 types of messages were monitored with Briggs and Myers words	29,200	Search for meta attributes is not available Collection of data for different classifier is a complex task [24]	Two hypotheses were tested named: single multiclass classifier and classification into binary problem Results show that there is best accuracy of 34.35% for NB Classifier Artisan and Guardian have high accuracy of an average of 87%
19.	Do et al. [5]	Twitter	Emotion analysis method Machine learning classifications State-of-the-art method Feature vector Classifier using support vector machine (SVM)	Middle East respiratory syndrome (MERS) case study was used for analysis Emotions expresses in twitter messages were exploited Public responses were analyzed using twitter messages Korean twitter messages were classified in seven categories which include neutral and Ekman's six basic emotions Messages were categorized in feature vector	5706 tweets	Complicated method Can only be used on twitter accounts Cannot apply on other social networking sites (SNS)	Figures show that 80% of the tweets is neutral and fear and anger dominates Trends of emotions over time were analyzed and shows that number of anger increase over time that result increase in public anger and fear and sadness decrease

20.	Li et al. [25]	Twitter	Classic sentiment analysis Granular partitioning method Data mining algorithm Jtwitter.jar libraries getFriendTimeline() method. REST API Pearson product moment correlation coefficient	Relationship between twitter users with stock market was analyzed to know behavior Jtwitter.jar was used to get friends status on twitter REST API was used to know home timeline or own status Four types of data return formats were used such as XML, JSON, RSS, and Atom	30,000	Timeliness of data requirement is very high Low processing speed Time of data is not improved with the improvement in accuracy	Result show that 2807 happy modes users were on 11/12, sad modes were on 11/19, anger were on 11/16, fear on 11/17, disgust was on 11/20 and most surprised mode was on 11/19
21.	Rao et al. [26]	Twitter	SocialKB framework Closes world assumption (CWA) and open world assumption (OWA) Apache Spark's stream processing API and Twitter 4J Spark SQL to process collected tweets	SocialKB model was used for modeling and reasoning about twitter posts and to discover suspicious users User and nature of their post was analyzed SocialKB relies on KB to know behavior of users and their posts KB will have entities, relationships, facts and rules Tweets were collected using Apache Spark's API and Twitter 4J	20,000	Different attack models were not analyzed First-order formulas used in KB is the biggest barrier ScocialKB framework does not know that data is independently and identically distributed	Each tweet has over 100 attributes Predicate tweeted (userID, tweetID) has more than 27,000 counts Only 16. 6% of URLs output by SocialKB were incorrectly detected as malicious

Continued

Table 1 Behavior Analysis Using Social Media Data Extracted From 30 Scientific Research Papers During 2015–17—cont'd

No.	Study	Social Media	Methods/Technologies Used	Description	Users	Limitations	Results
22.	Modoni et al. [27]	Twitter	Psycholinguistic analysis REST API Index analyzer	Psycholinguistic analysis is done on the Twitter contents which are written in Italian language Main aim is to analyze index and automatic analysis of the Twitter social posts REST API used to get the twitter data Information is gathered on the basis of location and time zone Correlation between weather and health is performed	3000 posts per day	Lack of interoperability Linked data facility is not available to integrate data from different other social media platforms	Calculation is performed on the basis of temperature, humidity and depression Calculation of correlation between temperature and depression provides the result as −0.8, which indicates high negative correlation
23.	Maruf et al. [8]	Twitter	Textalytics Media Analysis API Sentiment analysis Linguistic-based analysis with LIWC tool Personality analysis	Category scores from tweets used to analysis the behavior of twitter users By combining information from different social media comprehensive virtual profile can build Response prediction, news feed prediction, advertisement research can be done with this method	105	Complex process Not suitable to detect individual behavior on different subjects	Results show that users with high conscientiousness interested in human rights, crime, law and justice, etc. Achievement, humans, perceptual process have high score in comments on political and social issues

24.	Ghavami et al. [1]	Facebook	Classification algorithm and Pearson correlation coefficient formula were used for personality treat based on the user likes	Online questionnaire was designed 65 user's public posts were collected Comment-like-graph and post-like-graph was built Investigation on the correlation between each personality treat and 17 features from these two graphs	65	Small group of people participated in online survey and some did not show their trust to participate in the research	Correlation score table shows that N (Neuroticism) has weak correlation in CLI and CLP, whereas agreeableness and extroversion personality type have strong correlation
25.	Peng et al. [28]	Facebook	Jieba as a text segmentation tool for Chinese language Support vector machine (SVM) algorithm	Textual data was collected of FB posts of 222 users Feature extraction and feature selection were used for data processing Document matrix was designed SVM learning algorithm was used	222	Best accuracy is only 73.5% Experiment was conducted only on extraversion; not on other 4 factors	Table on average score of each personality shows that Openness to experience personality has high average score than others All user with more than 900 friends score high in extraversion personality factor
26.	Mihaltz et al. [3]	Facebook	TrendMiner Natural language processing (NLP) method	Collected data was processed using NLP tools such as segmentation, tokening with huntoken tool Sentiment analysis was performed to evaluate the behavior of users	14,000 public posts, 2 million comments, 1300 pages	Limited for Hungarian users only Limited users	Custom tool was evaluated for identification of psychological phenomena against human judgment
27.	Pang et al. [29]	Instagram	Demographic analysis Text analysis Image analysis Age detection process	Demographic was analyzed by photo with face detection and face analysis tools Tags associated with the pictures were analyzed Penetration is done by analyzing followers of the brand Drinking behavior is analyzed	600	Media data mining is not involved Fake information can be collected as some accounts are fake	Results shows that Heineken brand has high number of followers and 51.91% male above 18 drink this particular brand

Continued

Table 1 Behavior Analysis Using Social Media Data Extracted From 30 Scientific Research Papers During 2015–17—cont'd

No.	Study	Social Media	Methods/Technologies Used	Description	Users	Limitations	Results
28.	Bhagat et al. [4]	Twitter	Cut-based classification approach using messaged exchanged on social media	SetiStrength and Treebank analysis is evaluated and limitation in these methods are evaluated New cut-based classification architectural approach is used to analysis the text documents with classification method	7 million	Graphical user interface (GUI) can be used for better understanding of users	Polarity, subjectively, and hierarchical polarity is used which shows that subjectively polarity analysis has more accuracy than other two
29.	Tsai et al. [30]	Facebook	Distributed data collection module Social personal analysis using user operation complexity analysis Personal preference analysis	Social personal analysis is designed using Facebook personal information. Data is collected on the basis of how many users like, share and join the pages on the Facebook With the help of this data personal preference analysis is done	10,000	With the analysis of personal preference interest in the different innovative application services will be big issue of social analysis in the future Not able to reach different application services areas in social analysis	The result show the data on the basis of different tests which shows: Page viewed by users, bounce rate, and Click rate. Table shows that Test_A has high PV, Test_B has high bounce rate, and Test_C has high click rate
30.	Ray et al. [31]	Facebook	Empirical analyze Mathematical and empirical model Inverse Gaussian distribution Binomial distribution	Mathematical and empirical model used to analyze the behavior of Facebook users Mathematical model will help to know the likeability of users from the point of view and with probability of viewing posts IGD to know the viewing probability	1200	Effectiveness of mobile learning is not included Software implementation is hard	Result shows that photos posted by users gets 39% more interaction than links, videos or any other text-based updates Textual posts, liking and comments were used to analyze 59% of users are those who are daily active and 96% are monthly users

3 Results

The aim of this research is to know the methods used by researchers to predict the behavior of social media users. In this research, data were collected based on the use of three different social networking sites such as Facebook, Instagram, and Twitter. A random user list was used to analyze the behavior. In our final analysis, we pooled the data, which showed a statistically significant difference in various parameters (published year, methods, results, and limitations) for different social media sites. The results section includes the percentage of research on the three social networking sites, research papers according to year with bar graph representations, data collection and behavior analysis methods and classification based on the different methods with line graph representations.

3.1 Statistical Analysis

We performed statistical analysis to organize the data and predict the trends based on the analysis. This showed the different social media sites used based on the data given in Table 1.

As shown in Fig. 1 and Table 2, 27% of data was based on Facebook users, 23% of data was based on Instagram users, and 50% of data was based on Twitter users. As such, it is clear that Twitter is used more than other two social media sites for the analysis of the behavior of users.

3.2 Research Papers According to Year

Table 3 represents the data based on the year published. This indicates that most of research was completed on Twitter in 2016 and there was no research done in 2017 on Facebook regarding the behavior of users.

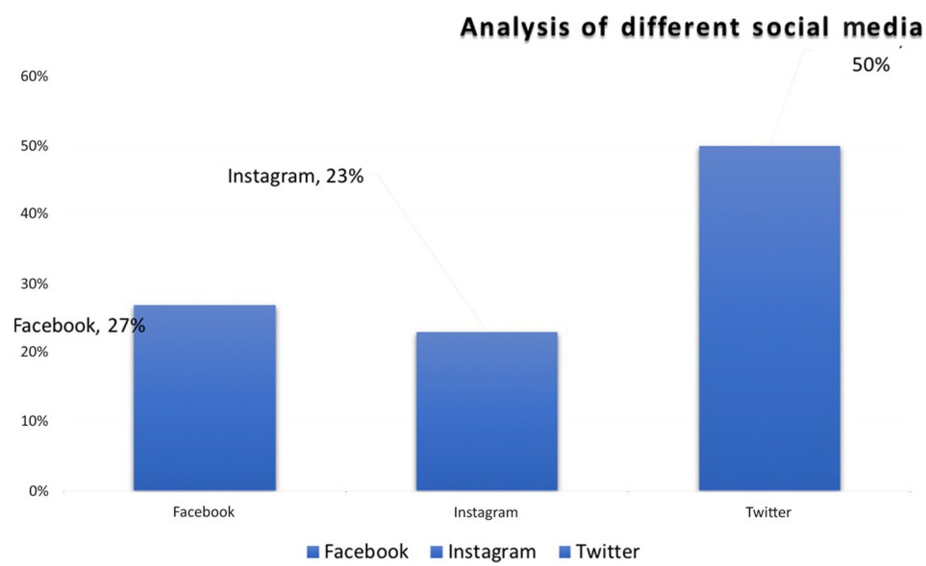

FIG. 1 Analysis of different social media to predict the behavior of users.

Table 2 Percentage of Number of Researches Completed in Three Different Social Media

Application	Number of Studies	Percentage
Facebook	8	27
Instagram	7	23
Twitter	15	50

Table 3 Number of Researches According to Year

Year	Social Media	Number of Studies
2015	Facebook	5
	Twitter	2
	Instagram	1
2016	Facebook	3
	Twitter	11
	Instagram	4
2017	Facebook	–
	Twitter	2
	Instagram	2

Fig. 2 shows that most of the research studies have been completed on Twitter in 2016. There was one research on the behavior analysis topic on Facebook in 2017.

3.3 Data Collection Method and Behavior Analysis Methods Used

Data collection techniques and behavior analysis methods used by different studies are shown in Table 4.

3.4 Classification Based on Different Methods

The behavior of users can be analyzed using different methods as shown in Table 5.

Fig. 3 is based on the classification of papers based on the different methods used and it is clear that the researchers have used analysis techniques more than others and they have rarely used coding rules.

4 Discussion

In this analysis, we observed that the amount of studies on Facebook and Instagram in the period from 2015 to 2017 was low, so there is a need of more research in these important areas.

This review study will help the readers to understand the different methods that the authors have used in their research studies on behavior analysis in social media.

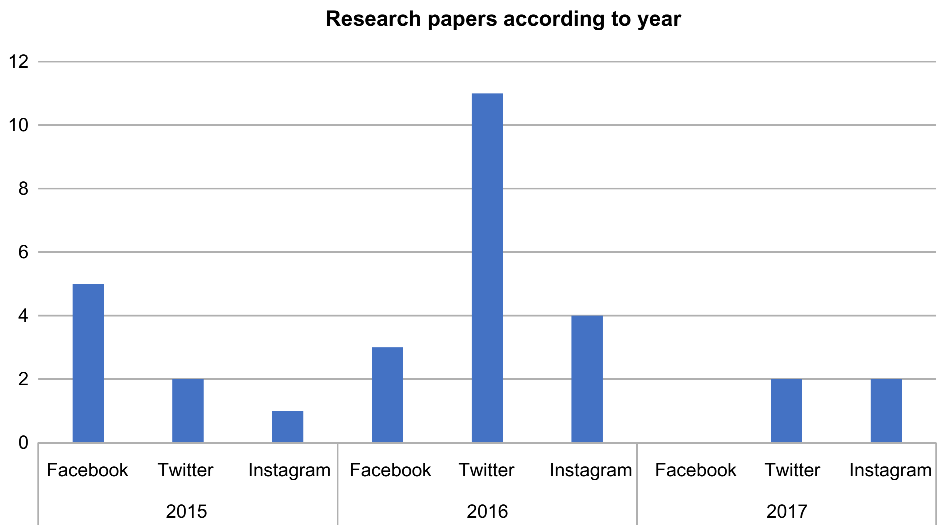

FIG. 2 Research papers according to 2015–17.

Table 4 Data Collection and Behavior Analysis Methods Used by Different Authors

Data Collection Techniques	Behavior Analysis Methods
Snowballing method, Gaussian mixture model (GMM), Quad closure methods, Binary classification problem, third-party apps, SmartPlus3 Software, Bag of words, Opinion mining, Questionnaire, Set of keywords, Twitter tweepy API, MBTI test	Regression analysis, quantitative method, correlation methods, machine learning, partition clustering algorithm, text processing, term frequency-inverse document frequency (TF-IDF), fuse-motif, CrossSpot algorithm, tenfold cross validation method, partial least square (PLS), JavaFX application, fivefold cross validation, cross industry standard process for data mining (CRISP-DM), supervised learning algorithms, fuzzy sentiment classification, NetworkX-METIS, Myers Briggs type indicator test, Jtwitter.jar libraries, REST API, index analyzer, LIWC tool, Jieba tool

An examination of the different methods of behavior analysis carried out with the help of social media is the main aim of this research. Thirty research studies were collected and analyzed to understand the personality of individuals who use social media such as Facebook, Twitter, and Instagram. Only three types of social network sites were included in this research. This analysis from the reported studies gives an overview of methods used to predict the personality of social media users.

As seen from Fig. 1, 50% of research was done on Twitter from 2015 to 2017, whereas as the other two social networking sites, Facebook and Instagram, only had 27% and 23%, respectively. Moreover, some studies [14, 21] proposed more than one method to analyze individuals' behavior.

Table 5 Classification Based on Different Methods to Analysis the Behavior of Users

Analysis techniques	Regression analysis, social network analysis, fuse-motif analysis, demographic analysis, fuzzy sentiment classification, sentiment analysis, classic sentiment analysis, psycholinguistic analysis, index analyzer, personality analysis, text/image analysis.
Coding rules	Binary coding
Models	Gaussian mixture model, path diagram model, technology acceptance model (TAM), multimode Erdos–Renyi model, machine learning, lexicon based, grammatical rule approach models, Mathematical, and empirical model
Algorithms	Machine learning, partition clustering algorithm, CrossSpot algorithm, supervised learning algorithms, feature vector, data mining algorithm, classification algorithm and Pearson correlation coefficient formula.
Principle	KL-divergence principle, minimum description length (MDL) principle.
API	Twitter tweepy API tools, REST API, Apache Spark's stream processing API, Textalytics media analysis API

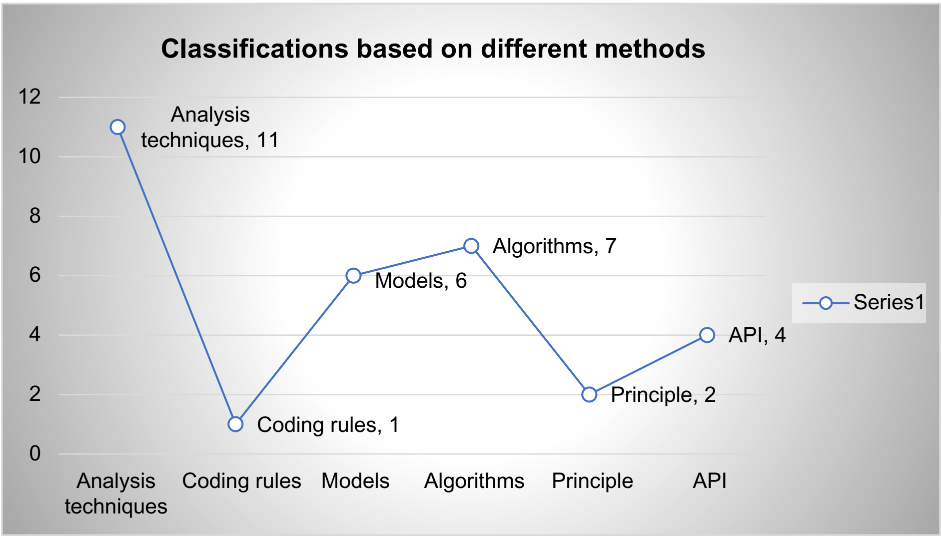

FIG. 3 Classification based on different methods used by 30 different studies.

A major issue in this area is the security and privacy of the information that the users put on the social media. However, some of the studies included in this review provided suggestions and methods to help secure the personal information of users. Many authors also discussed machine learning technique to observe the personality of social networking site users.

The results showed that most of the research completed in 2016 were on Twitter rather than Facebook and Instagram. In 2015, most research was done on Facebook and the least research was done on Instagram. On the other hand, in 2016 Twitter has the highest numbers of research papers and Facebook had the lowest numbers. In 2017, Twitter and Instagram had the highest number of research paper while Facebook had none at all.

Table 6 Demographic and Behaviour Trends From the Different Social Media

According to age: Age group between 45 and 55 use more Facebook than Twitter and Instagram. More than 79% of this age group use Facebook according to current trend
Use of smart phones: Another reason of using social media have been increased in the past year is smart phones. Smart phones have more visual interaction and people can access the social media easily. Advancement in the mobile phones play very important role in the increased users of social media
According to location: More people use the social media while they go out for dinner with family and friends. Other locations where people like to use social media is gym, cinema and home specially in lounge room area more than other rooms
According to time: More than 70% people use internet in the evening and 57% people use as a first thing in the morning. There is minimum use of social media during Breakfast, lunch, at work and commuting
Frequency of using social networking sites: More than 35% people use social media more than five times a day as compared to 20% people who never use social networking site in a day. There are only 3% people who use once a week
APPS: More than 68% use apps to access the social media and fewer people use websites to access the social media

Data collection and behavior analysis methods provided by authors were collected as raw data and analyzed. A classification based on the methods used by the authors for analysis was created.

Previous review studies did not include the limitations and number of users' attribute in their analysis. We have included these two attributes in Table 1 to make the research more specific and easy to understand for the readers [13].

The analysis of the papers indicated that Twitter has been the most used to predict the personality of social media users. Considering Table 1, there is a need for more variety in research methods on Instagram to understand the behavior of users.

A cut-based classification method was used to analyze the behavior of Twitter users by Bhagat et al. [4]. From the analysis done by these authors, they have concluded that cut-based classification method can be extended in the future to provide GUI for users for polarity classifications and subjectivity classifications. Real-time user messaging can also be analyzed in the future [18].

This review study is based on the analysis of behavior of individuals, who use social network in their daily life. This study benefits readers as it helps to identify the methods used by different researchers and the number of researchers that applied these methods. This review study provides a clear description of the methods, limitations, and results that have been used by previous researches in studies during 2015–17.

More than 37% people of the world use social media; however, the way social media users interact with each other vary greatly. There are demographic and behavioral trends from the Facebook, Twitter, and Instagram that are discussed in Table 6.

5 Conclusion

In this review paper, we have reviewed and analyzed data collected from 30 different published articles from 2015 to 2017 on the topic of behavior analysis using social media. It is found that there were 69 different methods used by the researchers to analyze their data. From these methods, the most common technique to analyze the behavior of individuals was analysis techniques. From this study, it is clear that there is need for more research to predict the personality and behavior of individuals on the Instagram. This study found that 50% of research was done on Twitter and 11 different analysis techniques were sued. While reviewing the research articles, it was clear that the researchers have used more than one method for data collection and behavior analysis. Table 1 has all the data analysis of the paper reviewed in the study. Furthermore, unlike past research papers, this chapter included the attributes of the number of users and the limitations of the work done. These studies mostly focused on Twitter with some research on Facebook and Instagram. In this research paper, we have attempted to fill the gap by including the number of users and limitation attributes. There are some challenges to find the solutions to the issues that have been discussed, but these require urgent attention. This study should be useful as a reference for researchers interested in the analysis of the behavior of social media users.

Author Contribution

A.S. and M.N.H. conceived the study idea and developed the analysis plan. A.S. analyzed the data and wrote the initial paper. M.N.H. helped in preparing the figures and tables, and in finalizing the manuscript. All authors read the manuscript.

References

[1] S.M. Ghavami, M. Asadpour, J. Hatami, M. Mahdavi, in: Facebook user's like behavior can reveal personality, International Conference on Information and Knowledge Technology, Tehran, Iran, 2015.

[2] L.C. Lukito, A. Erwin, J. Purnama, W. Danoekoesoemo, in: Social media user personality classification using computational linguistic, International Conference on Information Technology and Electrical Engineering (ICITEE), Tangerang, Indonesia, 2016.

[3] M. Miháltz, T. Váradi, in: TrendMiner: large-scale analysis of political attitudes in public facebook messages, IEEE International Conference on Congnitive Infocommunications, Budapest, Hungary, 2015.

[4] A.P. Bhagat, K.A. Dongre, P.A. Khodke, in: Cut-based classification for user behavioral analysis on social websites, Green Computing and Internet of Things (ICGCIoT), Noida, India, 2015.

[5] H.J. Do, C.-G. Lim, Y.K. Jin, H.-J. Choi, in: Analyzing emotions in twitter during a crisis: a case study of the 2015 Middle East respiratory syndrome outbreak in Korea, Big Data and Smart Computing (BigComp), Hong Kong, China, 2016.

[6] H. Hosseinmardi, R.I. Rafiq, Q. Lv, S. Mishra, in: Prediction of cyberbullying incidents in a media-based social network, International Conference on Advances in Social Networks Analysis and Mining (ASONAM), San Francisco, CA, USA, 2016.

[7] J. Kheokao, W. Siriwanij, in: Media use of nursing students in thailand, International Symposium on Emerging Trends and Technologies in Libraries and Information Services, Noida, India, 2015.

[8] A.H. Maruf, N. Meshkat, M.E. Ali, J. Mahmud, in: Human behaviour in different social medias: a case study of twitter and disqus, International Conference on Advances in Social Network Analysis and Mining, San Jose, CA, USA, 2015.

[9] H. Park, J. Lee, Do private and sexual pictures receive more likes on instagram? Research and Innovation in Information Systems (ICRIIS), Langkawi, Malaysia, 2017.

[10] G. Geeta, R. Niyogi, in: Demographic analysis of twitter users, Advances in Computing, Communications and Informatics (ICACCI), Jaipur, India, 2016.

[11] M. Jiang, A. Beutel, P. Cui, B. Hooi, Spotting suspicious behaviors in multimodal data: a general metric and algorithms, IEEE Trans. Knowl. Data Eng. 28 (8) (2016) 2187–2200.

[12] H.S. Farahani, A. Bagheri, E.H.K. Mirzaye Saraf, in: Characterizing behavior of topical authorities in twitter, International Conference on Innovative Mechanisms for Industry Applications, Tehran, Iran, 2017.

[13] R. Castro, L. Kuf'fo, C. Vaca, in: Back to #6D: predicting venezuelan states political election results through twitter, eDemocracy & eGovernment (ICEDEG), Quito, Ecuador, 2017.

[14] A.A. Mungen, M. Kaya, in: Quad motif-based influence analyse of posts in Instagram, Advanced Information and Communication Technologies (AICT), Lviv, Ukraine, 2017.

[15] T. Wiradinata, B. Iswandi, in: The analysis of instagram technology adoption as marketing tools by small medium enterprise, Information Technology, Computer, and Electrical Engineering (ICITACEE), Semarang, Indonesia, 2016.

[16] M. Nasim, R. Charbey, C. Prieur, U. Brandes, Investigating Link Inference in Partially Observable Networks: Friendship Ties and Interaction, Trans. Comput. Social Syst. 3 (3) (2016) 113–119.

[17] J. Järvinen, R. Ohtonen, H. Karjaluoto, in: Consumer acceptance and use of instagram, System Sciences (HICSS), 2016.

[18] B. Dalton, N. Aggarwal, in: Analyzing deviant behaviors on social media using cyber forensics-based methodologies, Communications and Network Security (CNS), Philadelphia, PA, USA, 2016.

[19] L.D.C.C. Chinchilla, K.A.R. Ferreira, in: Analysis of the behavior of customers in the social networks using data mining techniques, International Conference on Advances in Social Networks Analysis and Mining (ASONAM), San Francisco, CA, USA, 2016.

[20] P. Dewan, S. Bagroy, P. Kumaraguru, in: Hiding in plain sight: characterizing and detecting malicious facebook pages, IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), San Francisco, CA, USA, 2016.

[21] R. Toujani, J. Akaichi, in: Fuzzy Sentiment Classification in Social Network Facebook' Statuses Mining, International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT), Hammamet, Tunisia, 2016.

[22] C.Q. Santos, R. Tietzmann, M. Trasel, S.M. Moraes, I.H. Manssour, M.S. Silveira, in: Can visualization techniques help journalists to deepen analysis of Twitter data? Exploring the "Germany 7 x 1 Brazil" case, Hawaii International Conference on System Sciences, Porto Alegre, Brazil, 2016.

[23] A. Rabab'ah, M. Al-Ayyoub, Y. Jararweh, M.N. Al-Kabi, in: Measuring the controversy level of arabic trending topics on twitter, International Conference on Information and Communication Systems (ICICS), Irbid, Jordan, 2016.

[24] A.C.E.S. Lima, L.N. de Castro, in: Predicting temperament from twitter data, International Congress on Advanced Applied Informatics, São Paulo, Brazil, 2016.

[25] Q. Li, B. Zhou, Q. Liu, in: Can twitter posts predict stock behavior, Cloud Computing and Big Data Analysis (ICCCBDA), Chengdu, China, 2016.

[26] P. Rao, A. Katib, C. Kamhoua, K. Kwiat, L. Njilla, in: Probabilistic inference on twitter data to discover suspicious users and malicious content, Computer and Information Technology (CIT), Nadi, Fiji, 2016.

[27] G.E. Modoni, D. Tosi, in: Correlation of weather and moods of the Italy residents through an analysis of their tweets, International Conference on Future Internet of Things and Cloud Workshop, Varese, Italy, 2016.

[28] K.-H. Peng, L.-H. Liou, C.-S. Chang, D.-S. Lee, in: Predicting personality traits of chinese users based on facebook wall posts, Wireless and Optical Communication Conference (WOCC), Taipei, Taiwan, 2015.

[29] R. Pang, A. Baretto, H. Kautz, J. Luo, in: Monitoring adolescent alcohol use via multimodal analysis in social multimedia, Big Data (Big Data), Santa Clara, CA, USA, 2015.

[30] C.-H. Tsai, H.-W. Liu, T. Ku, W.-F. Chien, in: Personal preferences analysis of user interaction based on social networks, Computing, Communication and Security (ICCCS), Pamplemousses, Mauritius, 2015.

[31] K.S. Ray, M. Saeed, S. Subrahmaniam, in: Empirical analysis of user behavior in social media, International Conference on Developments of E-Systems Engineering, Duai, United Arab Emirates, 2015.

[32] S. Kamal, N. Dey, A. Ashour, S. Ripon, E. Balas, M. Kaysar, FbMapping, an Automated System for Monitoring Facebook Data, 3 October, 2017.

Further Reading

[33] Statista 2014 Age Distribution of Active Social Media Users Worldwide as of 3rd Quarter 2014, by Platform https://www.statista.com.