Tech Science Press

Analyzing Customer Reviews on Social Media via Applying Association Rule

Nancy Awadallah Awad[1,*] and Amena Mahmoud[2]

[1]Department of Computer and Information Systems, Sadat Academy for Management Sciences, Cairo, 11742, Egypt
[2]Faculty of computers and Information, Department of Computer Science,
Kafr el Sheikh University, Kafr El Sheikh, 33511, Egypt
[*]Corresponding Author: Nancy Awadallah Awad. Email: rarecore2002@yahoo.com
Received: 15 January 2021; Accepted: 17 February 2021

Abstract: The rapid growth of the use of social media opens up new challenges and opportunities to analyze various aspects and patterns in communication. In-text mining, several techniques are available such as information clustering, extraction, summarization, classification. In this study, a text mining framework was presented which consists of 4 phases retrieving, processing, indexing, and mine association rule phase. It is applied by using the association rule mining technique to check the associated term with the Huawei P30 Pro phone. Customer reviews are extracted from many websites and Facebook groups, such as re-view.cnet.com, CNET. Facebook and amazon.com technology, where customers from all over the world placed their notes on cell phones. In this analysis, a total of 192 reviews of Huawei P30 Pro were collected to evaluate them by text mining techniques. The findings demonstrate that Huawei P30 Pro, has strong points such as the best safety, high-quality camera, battery that lasts more than 24 hours, and the processor is very fast. This paper aims to prove that text mining decreases human efforts by recognizing significant documents. This will lead to improving the awareness of customers to choose their products and at the same time sales managers also get to know what their products were accepted by customers suspended.

Keywords: Machine learning; text mining; social media; big data; association rule; document clustering

1 Introduction

Since the rise in social media usage in the last decade, as an additional source to traditional media, individuals have been looking to gain information from the crowd. Social media data can be analyzed to gain insights into issues, trends, influential actors, and other kinds of information [1]. Through the use of an assessment of the social network, influencers or opinion pioneers can be distinguished and the scope of such a person can be disclosed by analyzing their follow-up network.

Text Mining (TM) is defined as a process to extract meaningful information from the collected text data. Before applying any data mining techniques, it should take into consideration the important process of TM which is preprocessing operations [2].

In the field of text data analysis, there are several applications used similar to information extraction, summarization, and document classification, clustering. The vast amount of textual documentation becomes more intensive study with the rise of web technology and needs to be perfectly processed to help researchers get meaningful information data mining techniques [3].

The objective of this study is to discuss how the applying of the association rule with text mining helps researchers to know the importance of an item (product) in social media without referring to review a huge amount of data from customers. In this study, the researcher analyzed the Huawei P30 Pro phone's consumer feedback, and text mining techniques are used to evaluate how various words are used for other words and what responses from the customer based mainly on this phone.

This research organization is as follows, Section 2 discusses the background of the text mining process, Section 3 presents the previous studies which focused on text mining techniques, Section 4 researchers present the proposed framework for text mining, Section 5 discusses the methodology of this study and Section 6 discusses the analysis and results for applying the proposed framework for text mining the customers' reviews on social media.

2 Background

2.1 Text Mining Process

The pattern is extracted from the unstructured data or natural language text as the input, as TM is the extraction of meaningful information from the text and then processed to obtain structured text [2]. TM includes five key steps for the text to be processed as:

— Document Gathering: In the first step, the text documents are collected in different formats which be in form of HTML doc, pdf, word [4].
— Document Pre-Processing:

In the second process, removing redundancies, separate words, inconsistencies, and stemming hence documents are prepared for the 3 next stages, as follows [5,6].

(a) Tokenization:
    The document string given is split into a single unit or token [5].
(b) Removal of Stop word:
    The removal of usual words like a, an, but, and, of, the, etc., in this step.
(c) Stemming:
    A stem is a group of words with equal significance that are very similar. The basis of a specific word [4] is described in this method. Porter's algorithm is one of the prevalent stemming algorithms.

Text Transformation: since the text document is a collection of words and their occurrences [7], the Vector Space Model and Bag of Words are the two main ways in which such documents are represented.

Feature Selection: this method retrieves an irrelevant feature from input [7,8]. Two techniques in the selection of features are filtering and wrapping techniques.

—Pattern Selection: the conventional process of data mining is combined with the process of text mining in this stage [8,9]. (Tab. 1) illustrated text mining techniques.

Table 1: Text Mining techniques

Technique	Description
Information extraction (stemming)	It is the task of automatically extracting structured information from unstructured and semi-structured machine-readable documents [10]. This activity concerns processing human language texts by using natural language processing (NLP).
Summarization	Reduces a text document to create a summary of the most important points of the original document.
Topic tracking	Keeps the user profile based on previous searches and very efficiently guess other documents based on the user profile [8,9].
Classification	Detects counts of words and decides the subject of the document from that count [11]. Text documents are classified into predefined class labels in this process.
Categorization	This is the task of assigning free-text documents to predefined categories [12]. It can provide conceptual views of collections of documents and has significant real-world applications.
Clustering	The most significant unsupervised learning issue is the finding of a structure in a collection of unlabeled data [11,12].
Concept linkage	To find related documents, text mining uses the linkage of the technique concept. Instead of searching, this mechanism navigates documents. It provides the facility for related documents to be linked.
Natural language processing	Design and create such a computer system to examine, understand, and generate NLP [13].
Stop-word list	Remove unimportant words such as "a", "the," "so" and so on.

3 Literature Review

The Internet is an environment to collect a huge expanded amount of data. Whereas data can be extensively ordered into two types, qualitative and quantitative data [13,14].

Social network investment is a form of consumption and the various types of returns on social capital, such as economic returns [15], political, social, and cultural gains, are the drivers of utility consumption. [16] However, to generate social capital, not all social media have equal political significance, as this depends on the types of platforms and the types of activities [17,18].

In this section, previous studies of text mining on social media will be represented, such as social media effects on customer's procurement via the internet, also text analysis through machine learning will be introduced.

3.1 Social Media Analysis

Authors in [19] have identified that the volume of data has been most often cited as a challenge addressed by social media analysis researchers. They presented a summary of the main challenges and challenges facing researchers in the (discovery, collection, and preparation) steps of the research process of social media analytics that come before the data is analyzed.

Authors in [20,21] indicated that the use of social media exposes individuals to information on political issues or current events, raising their awareness and understanding of these issues and increasing their likelihood of engaging in civic and political life.

While authors described in [22,23] that text mining tools help analyze and determine what the posts like and determining what the posts like, the individuals on the social media network (behavior and reaction) refer to or meaning.

3.2 Text Analysis

Authors in [2] integrated one of the association rule mining algorithms namely Apriori into text mining to find interesting patterns and it can easily understand by visualization techniques.

They indicated that the analysis of sentiment is a specific form of text analysis for the identification of valence and the analysis of subjectivity of user-generated content (UGC).

Authors in [24] applied sentiment analysis methods, the overall Web-based textual data, and various forms of UGC, whether positive, neutral, or negative, can be measured. Authors in [25] said that getting a structured form of text is the most important task in text mining, while authors in [26] discussed that dealing with structured data using mining tools is easy as compared to unstructured data. They also presented text mining applications in the national security systems, bioinformatics, and business intelligence.

Authors in [27] described the text mining process steps and presented a survey in text mining techniques which are used in research fields such as decision tree, clustering, categorization.

Authors in [28] applied the k-mean clustering technique to combine similar text documents through a web-based text mining process, and to identify information occurrence in documents, they used TF-IDF (Term Frequency-Inverse Document Frequency) algorithm.

3.3 Supervised and Unsupervised Machine Learning

Supervised learning refers to a classification technique for machine learning that uses a set of labeled training data to determine class labels for unnoticed instances. One of the common algorithms for classification (K-Nearest Neighbors, Vector Machine Support [SVM], Logistic Regression, Naive Bay [NB]) [29,30]. Authors in [29] illustrated that for sentiment analysis, the classic lexicon-based approach ("unsupervised technique") and both methods have been used.

The lexicon-based approach compares the characteristics of the text with pre-defined positive and negative sentiment lexicons and determines whether the document has a more positive or negative tone. For UGC valence classification, the supervised classification method exists. But the restriction of the lexicon-based approach to online review sentiment detection is that this method is highly domain-dependent.

Authors in [31] noted that supervised machine learning techniques have shown relatively better performance than unsupervised methods, but the need for large expert annotated training data to be generated from scratch could be one limitation, as the technique may fail when there is insufficient training data.

By comparing the performance of various classification techniques ('helpfulness analysis'), authors in [32] tried to identify the most helpful TripAdvisor hotel reviews. Authors discussed in [33] that the lexicon-based approach is a common unsupervised method of determining the polarity and semantic orientation of SM statements that involve predefining positive and negative word and phrase lexicons.

The process of Social Media Analytics (SMA) proposed by [34] was used to inform the framework for classification. There were three steps involved in the process of analyzing SM content: tracking, preparation, and analysis.

The following SM analytical methods are applied after a thorough review of the SMA methods used to accomplish these processes, such as text analysis; sentiment analysis; content analysis; trend analysis; predictive analytics; social network analysis; spatial analysis; and comparative analysis.

4 Proposed Framework for Text Mining System

In this section, the proposed framework for text mining will be presented with 4 main stages (retrieving data, processing, indexing, and association rule) phase as illustrated in (Fig. 1).

4.1 Retrieving Data

In this study, the researcher gathered data on review.cnet.com, CNET.technology on Facebook, and amazon.com from consumer feedback. File formats (RTF, txt, doc, etc.) are approved at this stage and will be translated into XML format at the processing stage.

4.2 Processing Phase

The processing phase has some sub-steps (transformation, filtration, and stemming of the documents). In this phase firstly text gathers from different sources for transformation. After that, unimportant words such as grammatical words (common adverbs, articles, determiners, pronouns, prepositions, and non-informative verbs (be)) are removed from documents content by the filtering process. Checking the content of the documents and eliminate all the unimportant words that are listed in stop words and also, after that, the special characters, parentheses, commas will be replaced with the spaces among words in the converted document. After completion of the categorization process, the process of word stemming will be started, which removes the word's prefixes and suffixes. A stemming dictionary (lexicon) will be used as a stemming algorithm.

4.3 Indexing Phase

The techniques for automated production of indexes associated with documents usually rely on frequency-based weighting schemes. The weighting scheme TF-IDF (Term Frequency, Inverse Document Frequency) is used to assign higher weights to distinguished terms in a document, and it is the most widely used weighting scheme [35] which is defined as Eq. (1) illustrated, where

$w\,(i,j) \ge 0$:

$$w\,(i,j) = \textit{tfidf}\,(d_i, t_j) = \begin{cases} Nd_i, t_j * log_2 \dfrac{|C|}{Nt_j} & \textit{if } Nd_i, t_{j\ge1} \\ 0 & \textit{if } Nd_i, t_{j=0} \end{cases} \tag{1}$$

N tj refers to the no. of documents in collection C

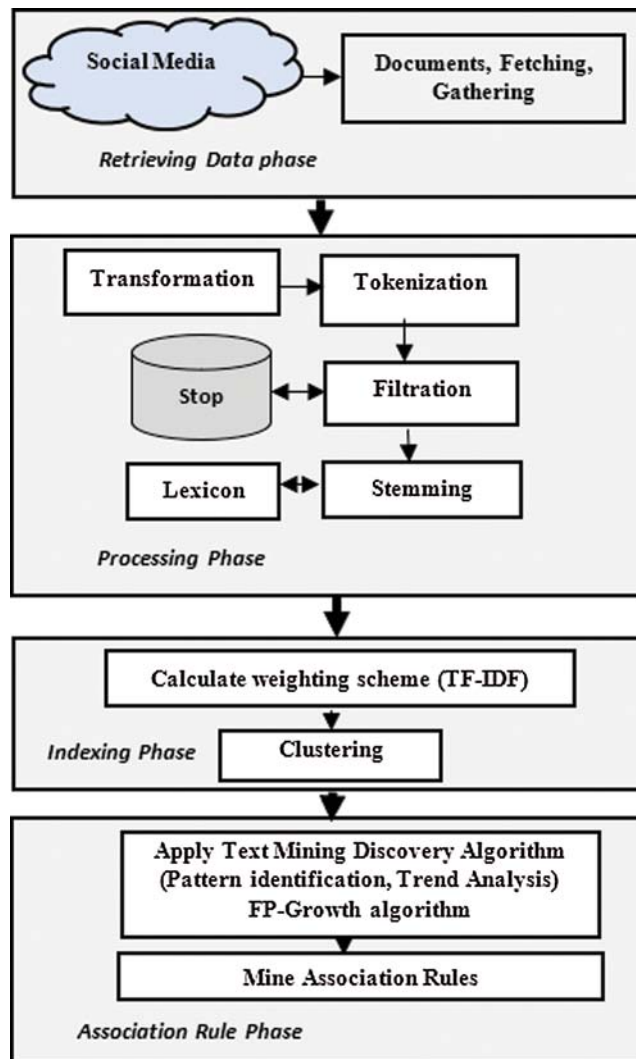Where the second clause, the value of w(i, j) = 0, when words that do not occur (Ndi, $t_j = 0$).

Figure 1: Framework for text mining system

Document frequency formula as Eq. (2) illustrated is:

$$log_2 \frac{|C|}{Nt_j} = logC - logNtj \tag{2}$$

$(\log\ C - \log N t_j = \log C - \log 1 = \log C)$ gives full weight to words that occur in one document [36]. To perform easily indexing, for each document select the keywords that achieve the given weight constraints, this will be done when a weighting scheme has been selected After this step, the X-mean clustering of documents was done.

4.4 Association Rule Mining

In this phase, an algorithm is used to find out the related words that are frequently used and to generate the confidence and lift factors on these words that will be helpful to make association rules. For text mining using the association rule, the Frequent Pattern Growth (FP-Growth) algorithm is used.

The FP-Growth algorithm is more applicable than the Apriori algorithm. It represents the database in the form of a frequent pattern tree or FP tree whose purpose is to mine the most frequent pattern [37]. The database is fragmented into "pattern fragment," each item of these fragmented patterns is analyzed. The lower nodes of the FP tree represent the item sets while the root node represents null [38].

(Fig. 2) illustrated pseudo code to mine the frequent pattern using the FP-Growth algorithm.

```
Algorithm FP_Growth
// Problem Description: Frequent Pattern mining using FP Growth
algorithm.
//Inputs: FP Tree, Transaction database (TD), Minimum Support (∂).
//Output: Frequent Patterns.
{
  If FP tree contains single path then
  {
      For each combination of the nodes in the path.
        {
            Generate patterns from the combination that satisfy the minimum
support.
        }
  }
Else
  {
    For each item in the header of the FP Tree.
    {
      Generate patterns from those combinations that satisfy the minimum
support.
        Construct conditional pattern base.
        Construct conditional FP Tree.
    }
  }
}
```

Figure 2: Pseudocode for FP growth algorithm [39]

5 Methodology

Customer reviews are collected from several sites and Facebook groups such as review.cnet.com, CNET. Technology on Facebook and amazon.com, where customers from everywhere put their notes about mobile phones. In this study, a total of 192 reviews of Huawei P30 Pro were collected from previously mentioned Facebook groups and amazon sites to analyze them through text mining techniques.

Next, stop words were removed which have no significant information and occur very frequently such as the words 'a', 'an,' 'is,' 'are,' this will be done through the stop-words process. After that, unimportant words such as grammatical words (common adverbs, articles, determiners, pronouns, prepositions, and non-informative verbs (be)) are removed from documents content by the filtering process. Next, a stemming dictionary (lexicon) will be used as a stemming algorithm.

Afterward, the indexing phase will be started, with the TF–IDF value of each word in each document was weighed. Each word existing in the matrix was created with TF–IDF scores.

Next step, the X-mean [40] clustering algorithm will be applied after forming the document term matrix. X-mean clustering produced three clusters. After that, the association rule is applied. RapidMiner tool is used for processing collected.

6 Analysis Results

6.1 Analysis of Clusters

X-mean clustering is applied for collecting data and produced 3 clusters which were identified as technical feedback, emotional feedback, and smartphone brands feedback. Some of the words are given in (Tab. 2).

Table 2: Important words in three clusters

Cluster_0	Cluster_1	Cluster_2
Charger	love	Galaxy
Android	Fast	Mate
Face	Awesome	Pro
Recognition	Overpriced	iPhone
Camera	Work	Huawei
Buy	Cheap	Apple
Screen	Worth	Samsung
Lightning	Great	Product
Device	Quality	Device
Features	Waste	Phone
USB	Life	Note
Headphone	Look	Mobile
Jack	Lot	Model
Lasts	Old	S
Display	Perfect	P
Adapter	Price	
Battery	Simple	
Plug	Smaller	
Sound	Excellent	
Service	Fact	
Cable	Amazing	
Build	Available	
	Disappointed	
	Big	
	Simple	
	Outstanding	

Cluster_0 represents customer's remarks which focused on the technical aspects of Huawei pro, whereas Cluster_1 is represented customer's remarks which focused the emotional feedback, but Cluster_2 has represented customer's comparison between several products and brands (Huawei, iPhone, Samsung) concerning the features of Huawei P30 Pro, iPhone 11 and Samsung Galaxy note 10 plus.

6.2 Analysis of Word Counts

Word counts can be used to determine what are the most words which should be meaningful in the output. Hence, all reviews of Huawei pro, Huawei, and mate have occurred very frequently. (Tab. 3) represents counts of the word for the most word repeated in customer's reviews.

Table 3: Counts of word

Word	Freq.	Word	Freq.	Word	Freq.	Word.	Word
phone	151	S	216	Mate	42	Lasts	4
iPhone	80	Samsung	44	Charge	13	LCD	6
Face	4	Scratches	4	Android	6	Life	6
Recognition	4	Se	54	Awesome	19	Look	10
Fast	10	Service	13	Overpriced	4	Lot	10
Love	6	Simple	17	Screen	16	Mobile	4
Features	12	Smaller	4	Lightning	4	model	5
Green	4	Sound	9	Device	6	Nice	8
Pro	107	T	287	Beautiful	4	Note	42
Huawei	224	Thanks	4	Camera	30	Old	6
Buy	12	Think	4	Waste	9	OLED	4
Use	20	Time	8	Water	11	Outstanding	4
Money	8	USB	8	Work	14	p	316
Great	24	Verizon	4	World	4	Get	10
Quality	21	Price	13	Worth	6	Good	26
i	361	Problem	6	Perfect	12	Got	6
m	204	Product	8	Phones	8	Headphone	6
Speed	4	Review	8	Plug	9	Jack	10
Fact	10	Display	4	Connect	6	Charger	4
Feature	15	Example	4	Day	6	Cheap	6
Galaxy	36	Excellent	6	Disappointed	4	Compare	4
Bluetooth	11	Available	15	Actually	4	Amazing	18
Build	4	Average	4	Adapter	4	Apple	24
c	170	Battery	18	Cable	6	Big	6

6.3 Word Association Analysis

Association rule mining presents the relation to other words and their occurrences in the document. In this phase, the FP-Growth algorithm is used to extract the related words that are repeatedly used and to generate the confidence and lifting factors on these words that will be helpful to make association rules.

In this study, (Tab. 4) illustrated important rules extracted from applying association rule, higher strength of the rule related to lifting value. The rules determined the association between

smartphones brands (Apple, Huawei, Samsung,) and products (iPhone 11, Pro, mate, Galaxy S, note 10 plus), it can be said that the word note associated with Samsung and also galaxy word with Samsung. Also, it can be said that at least some people have associated the words good, camera, and great with either Huawei (P20 Pro, P30 Pro, mate 20) or the product especially Huawei p30 pro.

Table 4: Text mining results via applying association rules

Premises	Conclusion	Support	Confidence	Lift
Phone, pro, camera	Huawei	0.052083333	1	2.042553
Phone, pro, great	Huawei	0.052083333	1	2.042553
Phone, p, camera	Huawei	0.052083333	1	2.042553
Phone, p, great	Huawei	0.052083333	1	2.042553
Pro, camera	Huawei, p	0.09375	1	3.310345
p, camera	Huawei, pro	0.09375	1	3.310345
Pro, p, camera	Huawei	0.09375	1	2.042553
p, good	Huawei, pro	0.0625	1	3.310345
Pro, p, good	Huawei	0.0625	1	2.042553
Pro, great	Huawei, p	0.0625	1	3.310345
p, great	Huawei, pro	0.0625	1	3.310345
Pro, p, great	Huawei	0.0625	1	2.042553
Phone, pro, camera	Huawei, p	0.052083333	1	3.310345
Phone, p, camera	Huawei, pro	0.052083333	1	3.310345
Phone, pro, p, camera	Huawei	0.052083333	1	2.042553
Phone, pro, great	Huawei, p	0.052083333	1	3.310345
Phone, p, great	Huawei, pro	0.052083333	1	3.310345
Phone, pro, p, great	Huawei	0.052083333	1	2.042553
Phone, note	Samsung, galaxy	0.052083333	1	6.857143
Phone, note, galaxy	Samsung	0.052083333	1	5.647059
Apple	iPhone	0.114583333	1	3.428571
Mate	Huawei	0.1875	1	2.042553
Phone, p	Huawei	0.145833333	1	2.042553
Phone, mate	Huawei	0.052083333	1	2.042553

7 Conclusion

Text mining decreases human efforts by recognizing significant documents. So, not all 192 (Customer's reviews) were important to be read to understand what customers opinions about Huawei P30 Pro which has been a large portion by most of the reviewers. The loyalty to the iPhone was re-presented by some user feedback and compared with the Samsung Galaxy note 10 plus by some others.

Customers who love Samsung claim that it was easy to use and nice in price, but others assume that the charger of Samsung is poor. In specific, Huawei lovers (Huawei P30 Pro) say that it has strong points such as best safe, high camera quality, a battery that lasts more than 24 h, and a very good processor.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1]   M. Milad and E. Zapatka, "Sensemaking in social media crisis communication–a case study on the Brussels bombings in 2016," in *Proc. of the 25th European Conf. on Information Systems*, Guimarães, Portugal, pp. 2169–2186, 2017.

[2]   S. Sheela and T. Bharathi, "Analyzing different approaches of text mining techniques and applications," *International Journal of Computer Science Trends and Technology*, vol. 6, no. 4, pp. 23–29, 2018.

[3]   I. M. Nasir, M. A. Khan, M. Yasmin, J. H. Shah, M. Gabryel *et al.,* "Pearson correlation-based feature selection for document classification using balanced training," *Sensors*, vol. 20, no. 23, pp. 1–18, 2020.

[4]   S. M. Inzalkar and J. Sharma, "A Survey on text mining-techniques and application," *International Journal of Research in Science & Engineering*, vol. 24, pp. 1–14, 2015.

[5]   D. S. Dang and P. H. Ahmad, "A review of text mining techniques associated with various application areas," *International Journal of Science and Research*, vol. 4, no. 2, pp. 2461–2466, 2015.

[6]   F. N. Patel and N. R. Soni, "Text mining: A brief survey," *International Journal of Advanced Computer Research*, vol. 2, no. 4, pp. 243–248, 2012.

[7]   L. Kumar and P. k. Bhatia, "Text mining: Concepts, process and applications," *Journal of Global Research in Computer Sciences*, vol. 4, no. 3, pp. 36–39, 2013.

[8]   N. Zhong, Y. Li and S. Wu, "Effective pattern discovery for text mining," *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 1, pp. 30–44, 2012.

[9]   R. Patel and G. Sharma, "A survey on text mining techniques," *International Journal of Engineering Computer Science*, vol. 2, no. 5, pp. 5621–5625, 2014.

[10] R. Balamurugan and Dr S. Pushpa, "A review on various text mining techniques and algorithms," in *2nd Int. Conf. on Recent Innovation in Science, Engineering and Management*, New Delhi, pp. 837–848, 2015.

[11] Dr. G. Rasitha Banu and V. K. Chitra, "A survey of text mining concepts," *International Journal of Innovations in Engineering and Technology*, vol. 5, no. 2, pp. 121–127, 2015.

[12] S. Kumar and R. Karthika, "A survey on text mining process and techniques," *International Journal of Advanced Research in Computer Engineering & Technology*, vol. 3, no. 7, pp. 2279–2284, 2014.

[13] J. Elgot, "From relationships to revolutions: Seven ways facebook has changed the world," *Guardian*, 2015. [Online]. Available: https://www.theguardian.com/technology/2015/aug/28/from-relationships-to-revolutions-seven-ways-facebook-has-changed-the-world.

[14] A. Basalingappa, M. Subhas and R. Tapariya, "Understanding likes on facebook: An exploratory study," *Online Journal of Communication and Media Technologies*, vol. 6, no. 3, pp. 234–249, 2016.

[15] T. Goh, Z. Xin and D. Jin, "Habit formation in social media consumption: A case of political engagement," *Behaviour & Information Technology*, vol. 38, no. 3, pp. 1–16, 2019.

[16] X. Chen, D. Tao and Z. Zhou, "Factors affecting reposting behaviour using a mobile phone-based user-generated-content online community application among Chinese young adults," *Behaviour & Information Technology*, vol. 38, no. 2, pp. 1–12, 2019.

[17] T. A. Engbers, M. F. Thompson and T. F. Slaper, "Theory and measurement in social capital research," *Social Indicators Research*, vol. 132, no. 2, pp. 537–558, 2017.

[18] J. Kahne and B. Bowyer, "The political significance of social media activity and social networks," *Political Communication*, vol. 35, no. 3, pp. 1–24, 2018.

[19] S. Stieglitz, M. Mirbabaie, B. Ross and C. Neuberger, "Social media analytics–challenges in topic discovery, data collection, and data preparation," *International Journal of Information Management*, vol. 39, no. 1, pp. 156–168, 2018.

[20] T. L. Towner and C. L. Muñoz, "Baby boom or bust? The new media effect on political participation," *Journal of Political Marketing. Advance Online Publication*, vol. 17, no. 1, pp. 32–61, 2018.

[21] G. Wolfsfeld, M. Yarchi and T. Samuel-Azran, "Political information repertoires and political participation," *New Media & Society*, vol. 18, no. 9, pp. 2096–2115, 2016.

[22] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data," *Journal of Bioinformatics and Computational Biology*, vol. 3, no. 2, pp. 185–205, 2005.

[23] Y. Zhao, "Analysing twitter data with text mining and social network analysis," in *Proc. of the 11th Australasian Data Mining and Analytics Conf.*, Canberra, Australia, pp. 41–47, 2013.

[24] A. R. Alaei, S. Becken and B. Stantic, "Sentiment analysis in tourism: Capitalizing on big data," *Journal of Travel Research*, vol. 58, no. 2, pp. 175–191, 2017.

[25] S. Dasgupta and K. Sengupta, "Analyzing consumer reviews with text mining approach: A case study on Samsung galaxy s3," *Paradigm*, vol. 20, no. 1, pp. 1–13, 2016.

[26] S. H. Liao, P. H. Chu and P. Y. Hsiao, "Data mining techniques and applications—A decade review from 2000 to 2011," *Expert Systems with Applications*, vol. 39, no. 12, pp. 11303–11311, 2012.

[27] N. Zhong, Y. Li and S. T. Wu, "Effective pattern discovery for text mining," *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 1, pp. 30–44, 2012.

[28] R. Rajendra and V. Saransh, "A novel modified apriori approach for web document clustering," *International Journal of Computer Applications*, vol. 33, pp. 159–171, 2013.

[29] M. Farshid and H. Elizabeth, "Social media analytics in hospitality and tourism a systematic literature review and future trends," *Journal of Hospitality and Tourism Technology*, vol. 10, no. 4, pp. 764–790, 2019.

[30] T. Akram, M. A. Khan, M. Sharif and M. Yasmin, "Skin lesion segmentation and recognition using multichannel saliency estimation and M-SVM on selected serially fused features," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 2, pp. 1–20, 2018.

[31] A. P. Kirilenko, S. O. Stepchenkova, H. Kim and X. Li, "Automated sentiment analysis in tourism: Comparison of approaches," *Journal of Travel Research*, vol. 57, no. 8, pp. 1012–1025, 2017.

[32] M. P. O'Mahony and B. Smyth, "A classification-based review recommender," *Knowledge-Based Systems*, vol. 23, no. 4, pp. 323–329, 2010.

[33] M. Taboada, J. Brooke, M. Tofiloski, K. Voll and M. Stede, "Lexicon-based methods for sentiment analysis," *Computational Linguistics*, vol. 37, no. 2, pp. 267–307, 2011.

[34] S. Stieglitz, L. Dang-Xuan, A. Bruns and C. Neuberger, "Social media analytics," *WIRTSCHAFTSINFORMATIK*, vol. 56, no. 2, pp. 101–109, 2014.

[35] J. Paralic and P. Bednar, "Text mining for documents annotation and ontology support," in *Intelligent Systems at Service of Mankind*. Germany: UBooks Verlag, Augsburg, Deutschland, 2003, ISBN 3-935798-25-3.

[36] H. Mahgoub, "Mining association rules from unstructured documents," *International Journal of Applied Mathematics and Computer Sciences*, vol. 1, no. 4, pp. 201–206, 2006.

[37] B. Christian, "An implementation of the FP-growth algorithm," in *Proc. of the 1st Int. Workshop on Open Source Data Mining: Frequent Pattern Mining Implementations*, Chicago, Illinois, pp. 1–5, 2005.

[38] S. Dasgupta and K. Sengupta, "Analyzing consumer reviews with text mining approach: A case study on Samsung galaxy S3," *Paradigm*, vol. 20, no. 1, pp. 1–13, 2016.

[39] S. Sundaramoorthy, "Comparison of hybrid approaches with traditional algorithms for improving scalability of frequent pattern mining," *Journal of Web Engineering*, vol. 14, no. 3&4, pp. 286–300, 2015.

[40] D. Pelleg and A. Moore, "X-means: Extending k-means with efficient estimation of the number of clusters," in *Proc. of the 17th Int. Conf. on Machine Learning*, San Francisco, CA, Morgan Kaufmann, pp. 727–734, 2000.