Computers and Electrical Engineering 86 (2020) 106723

Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

# A machine learning-based approach to enhancing social media marketing☆

B. Senthil Arasu, B. Jonath Backia Seelan*, N. Thamaraiselvan

*Department of Management Studies, National Institute of Technology, Tiruchirappalli, India*

ARTICLE INFO

*Article history:*
Received 19 December 2019
Revised 28 May 2020
Accepted 28 May 2020

*Keywords:*
Social media marketing
Social analytics
Artificial intelligence
Information sciences
Machine learning and WEKA

ABSTRACT

Social media (SM) represent beneficial channels for marketers, business promoters and consumers. To acquire continuous revenues and more active customers, key business players should understand the behaviour and purchase preferences of buyers. To predict the buying decisions of purchasers, data about purchase intentions and desires have to be extracted with the help of data mining techniques. The purpose of this paper is to examine social media data analytics using machine learning tools; this new approach for developing a social media marketing strategy employs the Waikato Environment for Knowledge Analysis (WEKA). WEKA is compared with other algorithms of interest and found to outperform its peers, especially with regard to parameters such as precision, recall, and F-measure, indicating that WEKA performs better than other approaches.



## 1. Introduction

The internet is a dominant marketing tool, and it can be used to attract customers, build trustworthiness and extend a product or service's brand [1]. SM offer platforms where people communicate and collaborate virtually. Users' thoughts are controlled and influenced by frequent advertisements that they come across on various micro blogging and social media platforms [2]. Business analysts use SM for business exploration, corporate knowledge gathering, and product awareness. The current number of social media users is increasing every day due to their varied browsing interests [3]. Fig. 1 shows how individuals consider social media when making buying decisions. This paper mainly focuses on the possible ways to leverage marketing via SM using various available machine learning techniques to predict customer purchase preferences. The remainder of this paper is organized as follows. Section II presents an overall literature review on the trends in SM platforms, and Section III further elaborates the study report on the social data analysis. Section IV describes the proposed ML integrated social media marketing (ML-SMM) approach and analyses its implementation and performance. Finally, Section V concludes the paper by describing the potential advantages and applications of the proposed ML-SMM approach.

## 2. Literature review

Many marketers prefer to use artificial intelligence (AI) to transform data into valuable customer insights. Information gathering is an art [4] that involves identifying the benefits of online marketing for improving information gathering and

☆ This paper is for regular issues of CAEE. Reviews processed and recommended for publication to the Editor-in-Chief by Associate Editor Dr. S. Smys.

* Corresponding author.
*E-mail addresses:* arasu@nitt.edu (B.S. Arasu), jonathbaskaran@gmail.com (B.J.B. Seelan), selvan@nitt.edu (N. Thamaraiselvan).

https://doi.org/10.1016/j.compeleceng.2020.106723
0045-7906/

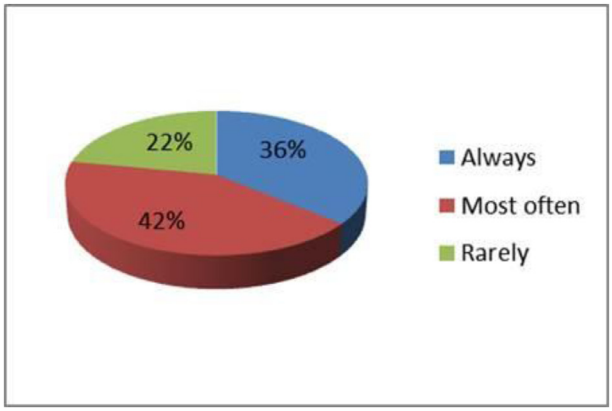


**Fig. 1.** SM considered when making buying decisions.

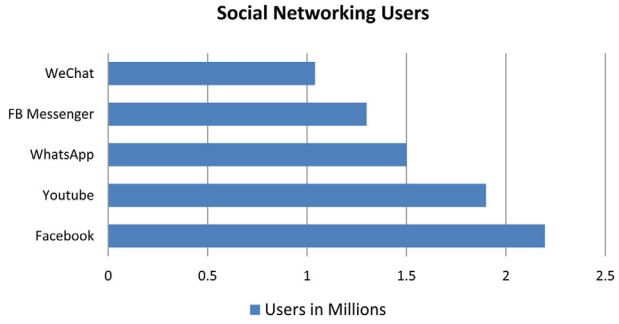


**Fig. 2.** The most popular social networking platforms based on users.

feedback. Most users use social media platforms, such as those shown in Fig. 2, to share their views on products of interest and their expected requirements. Other benefits of social media include the following:

1. High potential to share purchase experience or product knowledge/experience,
2. To increase the promotion of firms,
3. Magnify internal and external relationships, and
4. Know a user's consumption habits and preferences.

This research explores the possibility of identifying efficient classification, clustering and prediction algorithms with the use of a machine learning tool called WEKA [5]. The features and limitations of other machine learning tools and limitations are discussed in Tables 3 and 4 in the following sections. WEKA uses different algorithms for different scenarios; it has a sufficient number of algorithms to predict different market conditions. Machine learning (ML) is one of the sub-disciplines of AI. Various AI concepts help in addressing different market challenges that involve recovering from and rectifying business crises. A large portion of the human workload can be reduced The human workload is reduced by artificial intelligence systems. Business models are generated using a training set of samples. The unknown test set samples are analysed using the models.

Recurring challenges include the following:

(i) The ability to generate and leverage shopper perceptions,
(ii) Using social media data analytics (structured and unstructured data), and
(iii) Maximizing the effectiveness of analytical capabilities.

Social analytics include sentiment analysis, natural language processing (NLP), text analysis, predictive analysis, content analysis, and statistical and behavioural analysis [6,7]. The applications that support marketing for business promoters include a few social media types/platforms, as shown in Table 1.

Mining techniques are used for text analytics. To implement these concepts, a machine learning tool called WEKA has been used; WEKA has built-in classifiers, clustering tools, prediction algorithms, and a data visualization tool for visualizing and comparing results.

**Table 1**
Social media platforms.

| Type | Platforms |
|---|---|
| SN sites | Facebook, LinkedIn, Myspace |
| Blogs | BlogSpot, dig word press |
| Microblogs | Twitter, tweet peek |
| Content communities | YouTube, Flickr, Slideshare |
| Video sharing communities, forums | Skype, Messenger, Google Talk |
| Photo sharing | Instagram, Pinterest, Snapchat |

## 3. Social data analysis system study

NLP allows AI systems to analyse human language to derive the meanings of blogs and product reviews and billions of tweets, Facebook posts and status updates [8]. A report sponsored by the EMC predicts that the volume of data will grow to 40 zettabytes by the end of 2020, resulting in 50-fold growth from the beginning of 2010. The large volumes of data [9], which are mostly unstructured and generated by social networks, drive text mining to play a vital role in social data analysis.

Retail companies use ML technologies as catalysts and tools to support solving market problems [3]. The sample use cases include assessing freshness and the markdown detection problem and classification techniques. Prediction techniques have been used for predicting sales data and shelf-out scenarios. Clustering algorithms are useful for customer segmentation, advertisements, and personalized communication [10]. ML has been used for product listings and for ranking advertising concepts.

Many companies currently perform various kinds of analytics, such as sentiment analysis, to achieve a better understanding of and response to what customers observe about them and their products in online marketing [11]. In the future, business organizations will rely more on various mining techniques and ML tools for dynamic data analysis.

Applications such as fraud detection, stock market prediction, customer relationship management and summarization involve a variety of text mining techniques that include keyword extraction, named entity recognition, gender prediction and sentiment analysis [12]. YouTube is the second largest search engine after Google, the third most frequently visited website and the largest video content sharing website in the world [13]. The massive use of YouTube generates billions of dollars in marketing communication. This shows how social media platforms play a dominant role in online marketing. Mechanisms that integrate machine learning tools with marketing strategies can be employed to achieve improved performance.

ML combines statistics and AI. ML involves learning from input data and generating knowledge as a model for making smart decisions on unfamiliar test data [14–18]. Social data analysis has reached new heights in the microblogging era, and unstructured data analyses have been used to find frequent patterns in recurrent activities [19]. AI and ML are the main analytical tools. Automatic (machine) recognition, description, classification and pattern grouping are important areas of interest problems in a variety of engineering and scientific disciplines such as biology, psychology, medicine, marketing, computer vision, artificial intelligence, and remote sensing [20].

## 4. The proposed ML integrated SMM with implementation and performance analysis

ML integrated social media marketing (ML-SMM) is our proposed approach. The steps of the process involved in the proposed ML-SMM approach are as follows: (i) Text mining, (ii) Machine learning integrated with social media marketing, and (iii) ML-SMM analysis using WEKA.

### *4.1. Text mining*

Text mining occupies a prominent position in a few research fields. On the web, just 20% of the information is organized, and the remaining 80% is unstructured information. Information access through content mining has generally been of high quality, and the goal of content mining is to create more business value. Text mining involves examining unstructured information and identifying important data patterns quickly. Individuals compose words or sentences that may contain errors via web-based networking media such as Facebook, twitter, blogs, and WhatsApp. Content mining is utilized to allow the investigation of the data using legitimate language and organized sentences.

For example, any social media site that is used to advertise products or services also requests customer feedback. However, customers do not give/discuss their feedback in a structured form; customers' ways of writing words differ. The analyst collects the feedback, analyses it (referred to as feedback analysis) and arranges it into meaningful patterns that can be further used in marketing. This work is done with the help of text mining algorithms [21,22] and different kinds of text mining techniques, which are listed in Table 2.

For example, Facebook is a popular social media site used by a large number of individuals to share their thoughts, ideas, and happy and unhappy experiences. Most importantly, those data are unstructured.

**Table 2**
Text mining techniques.

| Techniques | Concept |
|---|---|
| Retrieval | Retrieve valuable facts from unstructured text |
| Extraction | Extract knowledge from a structured database |
| Summarization | Keep the main points that have the same meaning |
| Categorization | Document-based classification (supervised learning), Pattern mining |
| Clustering | Grouping documents into clusters (unsupervised learning), Term-based ontology, k-means |
| Filtering | Support vector machines |

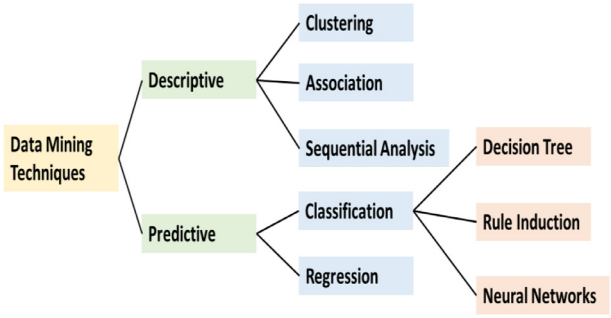


**Fig. 3.** Data mining techniques.

### *4.2. ML integrated SMM (ML-SMM) approach*

ML is particularly relevant for creating decision support systems and for analysis. One of the major innovations in digital marketing strategy has been the introduction of AI tools to streamline marketing processes and make businesses more effective. Organizations use ML outcomes to gain an in-depth understanding of consumer perceptions and to optimize their marketing strategies. ML tools can be a beneficial tool for digital marketers, allowing them to reveal and understand data better. By tracking consumer trends and producing actionable insights, one can predict online consumer behaviour. With more than 2 billion monthly users, Facebook is using AI to flag posts automatically.

LinkedIn uses ML and AI for almost all its products. LinkedIn uses algorithms with the capability to predict users who may be the best fit for a role. Using ML algorithms, it highlights candidates who are the most likely to respond or those who are seeking new opportunities [23]. Twitter recently launched an update to its service in which it uses AI to crop an image using face detection or create a thumbnail from an entire image [24,25].

### *4.3. ML-SMM analysis using WEKA*

WEKA is a data mining tool that performs data analysis and produces the results needed to achieve efficient marketing. Consequently, businesses can obtain more revenue and greater competitive advantage. This form of data analysis offers an understanding of customer behaviour with regard to purchasing. Data mining techniques can be classified into two broad categories, namely, descriptive and predictive, and the subdomains are given in Fig. 3.

WEKA is a combination of ML algorithms for executing data mining tasks. It has four interfaces that start from the main GUI (graphical user interface) window, as shown in Fig. 4 below. The Explorer, Knowledge Flow, and Simple CLI (Command Line Interface) all handle data preprocessing, classification, regression, clustering, and association. The Experimenter handles only classification and regression problems. Each interface has different utilities and different benefits. This tool also supports a variety of file formats such as ARFF, CSV, LibSVM, and C4.5.

***Building the data set for WEKA***

WEKA accepts data sets in attribute-relation file format (ARFF). In an ARFF data file, users can define each column. The proposed ML-SMM approach uses an ARFF file in WEKA, which is represented in Fig. 5 below.

***Features of WEKA***

- WEKA is a Java-based open-source data mining tool that is a collection of many data mining and machine learning algorithms.
- It is platform-independent software.
- It is stronger than other machine learning techniques and suitable for developing new machine learning schemes.
- Because it is open-source and extensible, it can be integrated with other Java packages. It provides three graphical user interfaces.

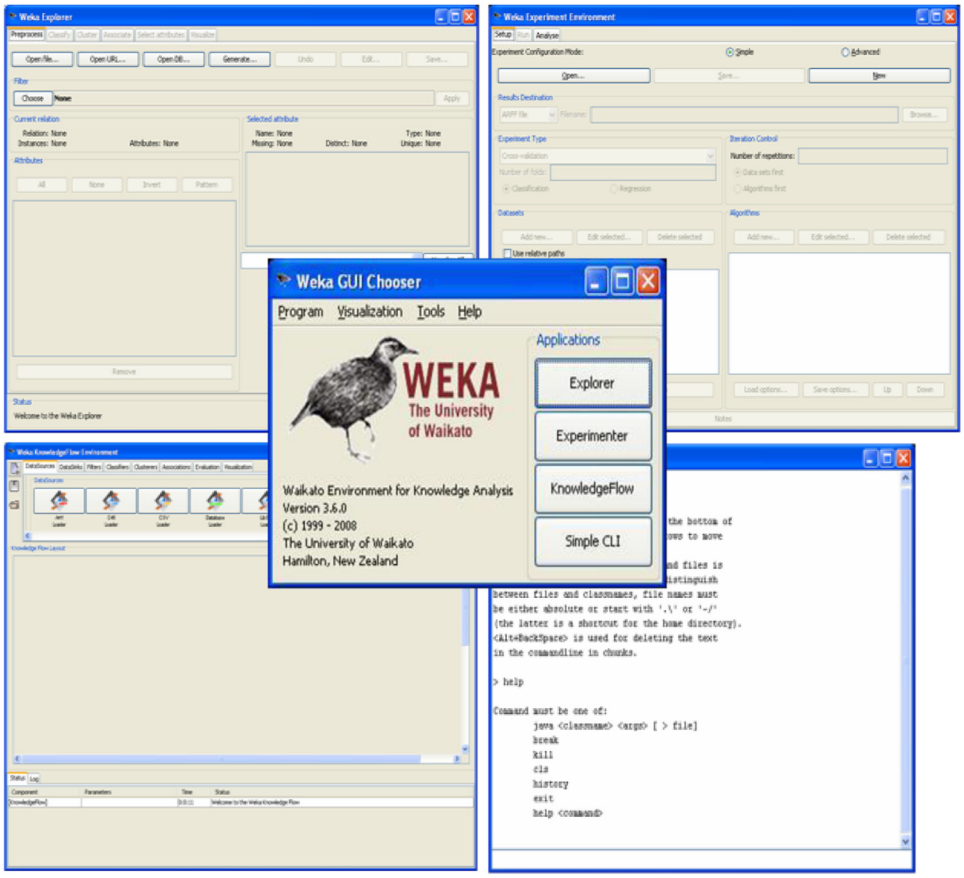


**Fig. 4.** Interfaces of WEKA.

```
@RELATION house

@ATTRIBUTE houseSize NUMERIC
@ATTRIBUTE lotSize NUMERIC
@ATTRIBUTE bedrooms NUMERIC
@ATTRIBUTE granite NUMERIC
@ATTRIBUTE bathroom NUMERIC
@ATTRIBUTE sellingPrice NUMERIC

@DATA
3529,9191,6,0,0,205000
3247,10061,5,1,1,224900
4032,10150,5,0,1,197900
2397,14156,4,1,0,189900
2200,9600,4,0,1,195000
3536,19994,6,1,1,325000
2983,9365,5,0,1,230000
```

**Fig. 5.** WEKA File format in ARFF form.

## Comparison of WEKA with other tools

Currently available ML tools were studied and analysed based on their usage and supporting features (shown in Table 3) and limitations (shown in Table 4) to select the most appropriate tool to apply to SM and ensure a breakthrough in the evolution of SMM.

**Table 3**
Comparison of ML tools.

| Tool(s) | Type | Features |
|---|---|---|
| RAPID MINER | Predictive and mathematical analysis | Many functions for data analysis and handling intuitive GUI. |
| ORANGE | Machine learning, data mining | Visual programming, data analytics, extensive toolset and programming interface. |
| KNIME | Business intelligence and mining | Sophisticated data handling and well-defined API. |
| R | Statistical tool | Time series analysis association discovery, data exploration, cluster discovery, parallel computing, text mining, social network analysis, and outlier detection. |
| WEKA | Machine learning | Data preprocessing, search algorithms for feature selection classification analysis, regression algorithms, clustering algorithms, attribute/subset evaluators, association rules, and efficient GUI assistance. |

## Limitations of ML tools

**Table 4**
Limitations of ML tools.

| Tool | Limitation |
|---|---|
| RAPID MINER | Requires prominent knowledge of database handling. |
| ORANGE | Large installation size, limited reporting capabilities. |
| KNIME | Limited error measurement. |
| R | Less specialized for data mining. |
| WEKA | Performs better by overcoming the above-stated limitations. |

**Table 5**
Data set description.

| Data set characteristic | Multivariate, text |
|---|---|
| Type of data | Categorical |
| Number of instances | 330 |
| Number of attributes | 4 |
| Associated task | Classification |
| Missing values | – |

**Table 6**
Simulation results.

| Algorithm (Total Instances, 61) | Correctly Classified Instances% (value) | Incorrectly Classified Instances% (value) |
|---|---|---|
| J48 | 96.7213 | 3.2787 |
| Navies Bayes | 93.44261 | 6.5574 |
| SMO | 85.2459 | 14.7541 |
| ZeroR | 63.93 | 36.06 |
| Part | 96.7213 | 3.2787 |

## Sample Instances (of implementation) using WEKA

The data set to be analysed is first collected from the web. The collected social media data may be structured or unstructured. Before analysing the social media data, the proposed mechanism preprocesses and converts the data to the acceptable format. Then, feature vectors are extracted, and machine learning techniques such as classification, prediction, and clustering are applied to the extracted data. The diagram in Fig. 6 shows the general implementation steps.

***An example of using data set1:***

The web data used for the data analysis are downloaded from the University of California, Irvine (UCI) ML repository. The data set includes 1650 data elements, which are formatted as 330 rows and four columns. Seventy-five percent of the data are used for training, and the remaining data are used for the testing process. Here, WEKA classification techniques are used for sample implementation, and performance is analysed. The classifiers are randomly chosen for this example because WEKA has more built-in classifiers. The proposed approach is implemented, and the classifiers' performance is compared; we identify which classifier works best for which kind of dataset. The data set description is shown in Table 5.

The simulation results based on the algorithm are shown in Table 6 and Fig. 7.

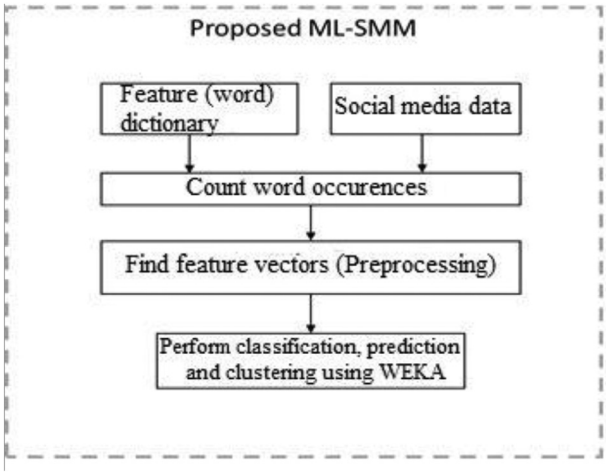


**Fig. 6.** Basic feature extraction architecture.

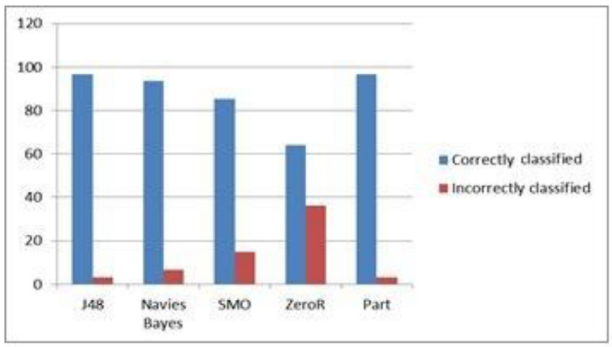


**Fig. 7.** Visualization of the simulation results.

| Metric | Formula |
|---|---|
| True positive rate, recall | $\frac{TP}{TP+FN}$ |
| False positive rate | $\frac{FP}{FP+TN}$ |
| Precision | $\frac{TP}{TP+FP}$ |
| Accuracy | $\frac{TP+TN}{TP+TN+FP+FN}$ |
| F-measure | $\frac{2 \cdot precision \cdot recall}{precision + recall}$ |

**Fig. 8.** Metrics and Formulas
TP - Positive tuples that are correctly classified as positive
TN- Negative tuples that are correctly classified as negative.
FP- Negative tuples that are incorrectly classified as positive.
FN- Positive tuples that are incorrectly classified as negative.

**Table 7**
Performance Values of the Precision, Recall, and F-Measure.

| TP (instances) | FP (instances) | Precision | Recall | F-Measure |
|---|---|---|---|---|
| 318 | 12 | 0.96 | 0.9 | 0.93 |
| 308 | 22 | 0.93 | 0.84 | 0.88 |
| 282 | 48 | 0.85 | 0.72 | 0.78 |
| 209 | 121 | 0.63 | 0.58 | 0.6 |
| 318 | 12 | 0.96 | 0.9 | 0.93 |

### WEKA algorithm's performance measures

To measure the performance of algorithms such as classifiers, clustering methods, and other mining techniques, different metrics and formulas are employed. By using these measures, one can analyse and compare the performance of different techniques and identify which technique is suitable for which kind of application. Those metrics are given in Fig. 8, and the performance assessments are listed in Table 7.

## 5. Conclusion

The proposed work on ML-SMM mechanisms explains the concepts of social media marketing and machine learning and integrates the WEKA machine learning tool to predict online consumer behaviour to ensure effective marketing. Sample datasets are collected and analysed using the WEKA tool, and the results show that it achieves better performance than other tools. Even though many tools are available for this task, the combination of ML-SMM with WEKA performs better than other tools in terms of applying different kinds of mining techniques, business applications and data analysis methods. This combination also overcomes the limitations of other tools by providing better reporting capabilities. This tool can also be extended to SMM campaigns pertaining to different business domains such as online education, health care, and music to explore their intentions, requirements and preferences.

### Ethics declarations

### Human and animal participants

This article does not contain any studies with human participants performed by any of the authors.

### Informed consent

All the authors are well informed about this submission.

### Declaration of Competing Interest

The authors declare that they have no conflicts of interest.

### CRediT authorship contribution statement

**B. Senthil Arasu:** Supervision, Conceptualization, Methodology, Software, Visualization, Writing - review & editing, Investigation. **B. Jonath Backia Seelan:** Conceptualization, Methodology, Writing - original draft, Validation, Data curation, Validation. **N. Thamaraiselvan:** Data curation, Formal analysis, Writing - review & editing.

### References

[1] Weber L. Marketing to the Social Web: How digital Customer Communities Build your Business. John Wiley & Sons; 2009.
[2] Tapscott D, Williams AD. Wikinomics: How Mass Collaboration Changes Everything. Penguin; 2008.
[3] Drummond WJ, French SP. The future of GIS in planning: converging technologies and diverging interests. J. Am. Plann. Assoc. 2008;74(2):161–74.
[4] Kietzmann J, Paschen J, Treen E. Artificial intelligence in advertising: how marketers can leverage artificial intelligence along the consumer journey. J. Advert. Res. 2018;58(3):263–7.
[5] Misirlis N, Vlachopoulou M. Social media metrics and analytics in marketing–S3M: a mapping literature review. Int. J. Inf. Manage. 2018;38(1):270–6.
[6] Duffett R, Petroşanu DM, Negricea IC, Edu T. Effect of YouTube marketing communication on converting brand liking into preference among millennials regarding brands in general and sustainable offers in particular. Evidence from South Africa and Romania. Sustainability 2019;11(3):604.
[7] Graves, A., 2013. Generating sequences with recurrent neural networks. arXiv preprint arXiv:1308.0850.
[8] Alves H, Fernandes C, Raposo M. Social media marketing: a literature review and implications. Psychol. Market. 2016;33(12):1029–38.
[9] Kietzmann J, Paschen J, Treen E. Artificial intelligence in advertising: how marketers can leverage artificial intelligence along the consumer journey. J. Advert. Res. 2018;58(3):263–7.
[10] Herz, F.S., Pinpoint Inc, 2002. *System and method for providing customized electronic newspapers and target advertisements*. U.S. Patent 6,460,036.
[11] Fan W, Gordon MD. The power of social media analytics. Commun. ACM 2014;57(6):74–81.
[12] Pejić Bach M, Krstić Ž, Seljan S, Turulja L. Text mining for big data analysis in financial sector: a literature review. Sustainability 2019;11(5):1277.
[13] Duffett R, Petroşanu DM, Negricea IC, Edu T. Effect of YouTube marketing communication on converting brand liking into preference among millennials regarding brands in general and sustainable offers in particular. Evidence from South Africa and Romania. Sustainability 2019;11(3):604.

[14] Bashar A, Ahmad I, Wasiq M. Effectiveness of social media as a marketing tool: an empirical study. Int. J. Market. Financ. Serv. Manag. Res. 2012;1(11):88–99.
[15] Hurtado JL, Agarwal A, Zhu X. Topic discovery and future trend forecasting for texts. J. Big Data 2016;3(1):7.
[16] Miklosik A, Kuchta M, Evans N, Zak S. Towards the adoption of machine learning-based analytical tools in digital marketing. IEEE Access 2019;7:85705–18.
[17] Papoutsoglou M, Ampatzoglou A, Mittas N, Angelis L. Extracting Knowledge from on-line Sources for Software Engineering Labor Market: a Mapping Study. IEEE Access 2019;7:157595–613.
[18] Spekman RE, Kamauff JW, Myhr N. An empirical investigation into supply chain management: a perspective on partnerships. Supply Chain Manag. Int. J. 1998.
[19] Kottursamy K, Raja G, Saranya K. A data activity-based server-side cache replacement for mobile devices. In: Artificial Intelligence and Evolutionary Computations in Engineering Systems. New Delhi: Springer; 2016. p. 579–89.
[20] Jain AK, Duin RPW, Mao J. Statistical pattern recognition: a review. IEEE Trans Pattern Anal Mach Intell 2000;22(1):4–37.
[21] Zhong N, Li Y, Wu ST. Effective pattern discovery for text mining. IEEE Trans. Knowl. Data Eng. 2010;24(1):30–44.
[22] Berry MW. Survey of text mining. Comput. Rev. 2004;45(9):548.
[23] Schwartz HA, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones SM, Agrawal M, Shah A, Kosinski M, Stillwell D, Seligman ME, Ungar LH. Personality, gender, and age in the language of social media: the open-vocabulary approach. PLoS ONE 2013;8(9):e73791.
[24] Pang B, Lee L. Opinion mining and sentiment analysis. Found. Trends® Inf. Retr. 2008;2(1–2):1–135.
[25] Liu B. Sentiment analysis and subjectivity. Handb. Natural Lang. Process. 2010;2(2010):627–66.

**Dr. B. Senthil Arasu** is an electronics and communication engineer with a master's in business administration and a doctorate in stock price behaviour. He was an equity research analyst and is now an associate professor with NIT Trichy. He also serves as a board of studies member, an external examiner, an academic auditor and a reviewer for international journals. He has conducted management development programmes for TNPL, L&T, BHEL, NLC, etc.

**B. Jonath Backia Seelan** is currently a research scholar in the Department of Management Studies at the National Institute of Technology, Tiruchirappalli. His-research interests include machine learning tools for social media marketing, social media advertising and integrated marketing.

**Dr. N. Thamaraiselvan** is an associate professor in the Department of Management Studies at the National Institute of Technology, Tiruchirappalli. His-research interests include brand leveraging strategies, customer satisfaction analytics, services branding, and social media services and digital marketing. He also serves as a board of studies member, an external examiner, an academic auditor and a reviewer for international journals.