Received: 3 September 2018 | Revised: 29 October 2018 | Accepted: 25 November 2018
DOI: 10.1002/cpe.5107

WILEY

SPECIAL ISSUE PAPER

Systematic literature review of sentiment analysis on Twitter using soft computing techniques

Akshi Kumar | Arunima Jaiswal

Department of Computer Science & Engineering, Delhi Technological University, Delhi, India

Correspondence
Akshi Kumar, Department of Computer Science & Engineering, Delhi Technological University, Delhi, India.
Email: akshikumar@dce.ac.in

Summary

Sentiment detection and classification is the latest fad for social analytics on Web. With the array of practical applications in healthcare, finance, media, consumer markets, and government, distilling the voice of public to gain insight to target information and reviews is non-trivial. With a marked increase in the size, subjectivity, and diversity of social web-data, the vagueness, uncertainty and imprecision within the information has increased manifold. Soft computing techniques have been used to handle this fuzziness in practical applications. This work is a study to understand the feasibility, scope and relevance of this alliance of using Soft computing techniques for sentiment analysis on Twitter. We present a systematic literature review to collate, explore, understand and analyze the efforts and trends in a well-structured manner to identify research gaps defining the future prospects of this coupling. The contribution of this paper is significant because firstly the primary focus is to study and evaluate the use of soft computing techniques for sentiment analysis on Twitter and secondly as compared to the previous reviews we adopt a systematic approach to identify, gather empirical evidence, interpret results, critically analyze, and integrate the findings of all relevant high-quality studies to address specific research questions pertaining to the defined research domain.

KEYWORDS
machine learning, review, sentiment analysis, soft computing, Twitter

1 | INTRODUCTION

The incessantly evolving dynamics of the Web in terms of the volume, velocity and variety of opinion-rich information accessible online, has made research in the domain of Sentiment Analysis (SA) a trend for many practical applications which facilitate decision support and deliver targeted information to domain analysts. Interestingly, the buzzing term "*big data*" which is estimated to be 90% unstructured[1] further makes it crucial to tap and analyze information using contemporary tools. Text mining models define the process to transform and substitute this unstructured data into a structured one for knowledge discovery. Use of classification algorithms to intelligently mine text has been studied extensively across literature.[2,3] SA, established as a typical text classification task,[4] is defined as the computational study of people's opinions, attitudes and emotions towards an entity.[5,6] It offers a technology-based solution to understand people's reactions, views and opinion polarities (positive, negative or neutral) in textual content available over social media sources.

Research studies and practical applications in the field of SA have escalated in the past decade with the transformation and expansion of Web from passive provider of content to an active socially-aware distributor of collective intelligence. This new collaborative Web (called Web 2.0),[7] extended by Web-based technologies like comments, blogs and wikis, social media portals like Twitter or Facebook, that allow to build social networks based on professional relationship, interests, etc, encourages a wider range of expressive capability, facilitates more collaborative ways of working, enables community creation, dialogue and knowledge sharing and creates a setting for learners to attract authentic audiences by various tools and technologies. The confluence of **S**ocial media, **M**obile, **A**nalytics and **C**loud has offered the new SMAC[8] technology paradigm, which has transformed the operative environment and user engagement on Web notably. A lateral shift from the conventional e-commerce (electronic commerce) to consequential s-commerce (social commerce) has been observed. S-commerce, as a subset of e-commerce has upheld almost all major innovative practices that assist online commercial activities such as retailing and marketing by incorporating social network(s) in the context of e-commerce transactions. It has expanded the scope of commercial activities by enabling the users to discuss, share, analyze, criticize, compare, appreciate and research about products, brands, services through social platforms like Voonik, Facebook, Twitter, etc. This pool of information can be explored for a mutual benefit of both the customer and the organization. Data Analytics on these social web-based corpora has thus been an ongoing trend where online text/posts/reviews/tweets/comments are transformed into a sentiment-rich knowledge-base that can leverage efficient and effective decision making. So we can say that the mass is relying on such online user generated content for the opinions that extensively marks the increasing significance of SA in our daily lives. For example, if an individual needs to visit a place, instead of asking his friends, relatives, agents, etc, he simply turns on to its online real time reviews of the visitors before taking any decision. Also in terms of business management, if a client needs to buy a product, he first reads all its reviews and then eventually reaches to a decision whether to buy or not to buy that product. Hence we can say that Internet has plethora of data to be analyzed meticulously.

Amongst the Web 2.0 tools, Twitter has evolved as a major revolution in field of social media and has a global reach. It has been the most preferred social channel from which sentiment rich data can be extracted. Originally launched in 2006, Twitter is the currently the most popular and impactful micro-blogging service connecting millions of people worldwide. It is a freely available social networking microblogging[9] media service where a registered user is allowed to broadcast short messages or posts called "tweet" to other registered users in real-time.[10] SA on Twitter has gained popularity due to intrinsic characteristics of the real-time messages shared on it. This is primarily due to the fact that the post size characterizes short text with character-set limit of 280. It has various applications covering wide spectra of domains and has eventually become an indispensable part of an individual's daily digital routine life. Moreover, due to its global connectivity with the diverse user-base and active participation from the users makes it the qualitative and quantitative base for analyzing sentiments. Consequently, we preferred choosing Twitter as the current development multimedia platform for portraying the research aspects on the use of SC techniques in the field of SA. The techniques for SA on Twitter have been also reported across pertinent literature.[5,6,11,12]

The generic SA task includes Data collection; feature selection[13]; sentiment classification and sentiment polarity detection.[2] Effective feature selection is a computationally hard task[14] and has a significant role in determining the sentiment classification accuracy. Moreover, the increased dimensionality, complexity and fuzziness in the user-generated Twitter data further fosters the need to look for improved and optimized sentiment classification techniques. Studies are constantly being conducted to explore new paradigms which handle uncertainty, imprecision, approximation, partial truth, fuzziness and allow replication of human intelligence for personalized and tractable results. Soft Computing (SC) is one such field of study that exploits combination of new computational techniques that mimic consciousness and cognition in several important respects: they can learn from experience; they can universalize into domains where direct experience is absent; and, through parallel computer architectures that simulate biological processes, they can perform mapping from inputs to the outputs faster than inherently serial analytical representations.[15,16] Motivated by this, we present a systematic literature review on "*Sentiment Analysis of Twitter Using Soft Computing Techniques.*" The goal is to gather empirical evidence and analyze results from existing studies to give a critically evaluated discussion on the existing trends in available research, identify gaps in current search and provide future prospects in the area by means of answering the established research questions.

SC techniques are predominantly considered as optimization techniques that help modeling complex real world problems to achieve robust and low cost solutions. These techniques are generally divided into the following five categories[15,16]:

- *Machine Learning (ML)*: Supervised; Unsupervised; or Reinforcement learning. Unsupervised includes Hierarchical, C means, K means clustering, etc). Supervised includes Statistical (Regression, Naïve Bayesian, etc), Structural (Rule Based, Distance Based, etc) and Ensemble methods (Bagging, Boosting, Random Forest, etc,). Deep Learning (DL) is the latest addition to ML and is often regarded as the subset of ML. It is a probable approach used for implementing ML. It includes Deep NN (DNN), Recursive NN, Recurrent NN, Convolutional NN, Long Short Term Memory and Deep Belief Networks.
- *Neural Networks*: (Feed Forward; Multi-Layer Perceptron; Radial Basis; Kohonen Self-Organizing; Modular; Shallow and Deep Neural Networks (DNN)).
- *Evolutionary Computation*: Gene Expression Programming; Differential Evolution; Evolutionary Algorithms (such as Genetic Algorithms); Swarm Intelligence (Nature-Inspired Algorithms such as particle swarm optimization, ant colony optimization, etc)
- *Fuzzy Logic*
- *Probabilistic Reasoning*: Bayesian Networks (Bayesian probability)

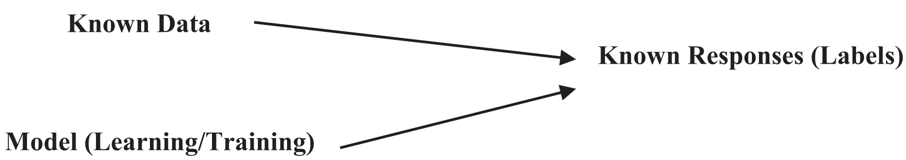
The unique property of all SC techniques is their power of self-tuning, that is, they derive the power of generalization from approximating and learning from experimental data.[15] This generalization for improved precision and certainty is usually done in a high-dimensional space and the big social data space such as Twitter serves as the true source to testify the reasoning and search capabilities of SC techniques when extended to a generic sentiment classification task. Studies that comprehend the use of SC techniques for this supervised learning model of SA on Twitter have continually motivated researchers and practitioners to explore the feasibility and scope of application. The work done in this survey is an insight to this research trend. The use of SC techniques for supervised SA on Twitter was first reported in 2012 by Finn and Mustafaraj.[17] Thus this survey includes papers which demonstrate this effort from the first reported study till June 2018 in a well-structured manner.

The paper is organized as follows: In the next section the basics of SA process and SC are explained. The Section 3 elaborates the review methodology enlisting the Research Questions identified to conduct this study review followed by Section 4, which overviews the pertinent literature of the selected studies concisely. Section 5 provides the results and discussion followed by the conclusion in the final Section 6.

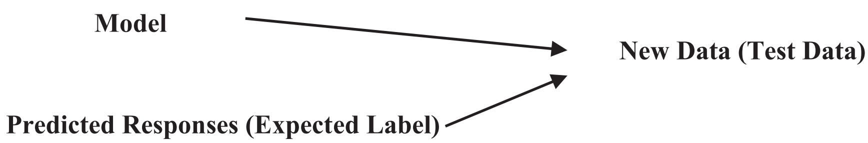
2 | SOFT COMPUTING TECHNIQUES IN SENTIMENT ANALYSIS

SA has been established as a typical text classification task across pertinent literature. A classification task is an instance of supervised learning from examples. In a supervised learning model the data (observations, measurements, etc) are labeled with pre-defined classes and the test data are classified into these classes too. It is a two-step process:

- *Learning (training)*: Learn a model using the training data



- *Testing*: Test the model using unseen test data to assess the model accuracy



The process of SA is depicted in the Figure 1:

SA has emerged as one of the most dynamic area of research in recent times. SA makes it viable to instantly transform unstructured data from social media sources like Twitter into structured data to comprehend intelligence.[18] But the continuously changing dynamics with respect to increasing user-base and user-activity (posts, comments, likes, re-tweets[19]); trending discussions on topics and issues from varied domains, makes Twitter a high-dimensional, complex and fuzzy data space to perform analytics. SC techniques offer a non-trivial solution to the real-world problems which are innately imprecise and uncertain. The guiding principle of SC is to exploit the tolerance for imprecision, uncertainty and partial truth to achieve tractability, robustness, low solution cost, and better rapport with reality.[20] The alliance of these two domains evokes the necessary balancing and necessitates investigation on feasibility, trends and scope of using SC techniques for the *supervised learning model* of SA on Twitter.

The Figure 2 depicts the various SC techniques and their categories. From the observed categorization,[15,16] SC is described as a *"blanket term"* leveraging computational intelligence, comprising of several methodologies which are themselves inter-related to one other in varied forms.

As discussed in Section 1, DL is considered as a sub-category of ML[16] which consists of various techniques such as Deep NN (DNN), Recursive NN, Recurrent NN, Convolutional NN, Long Short Term Memory and Deep Belief Networks. Neural Networks is also established as an independent sub-category of SC techniques[16] and includes Feed Forward; Multi-Layer Perceptron; Radial Basis; Kohonen Self-Organizing; Modular; Shallow and Deep Neural Networks (DNN). There is a broad overlap between these techniques and the inter-connection between SC, ML and DL is shown in Figure 3.

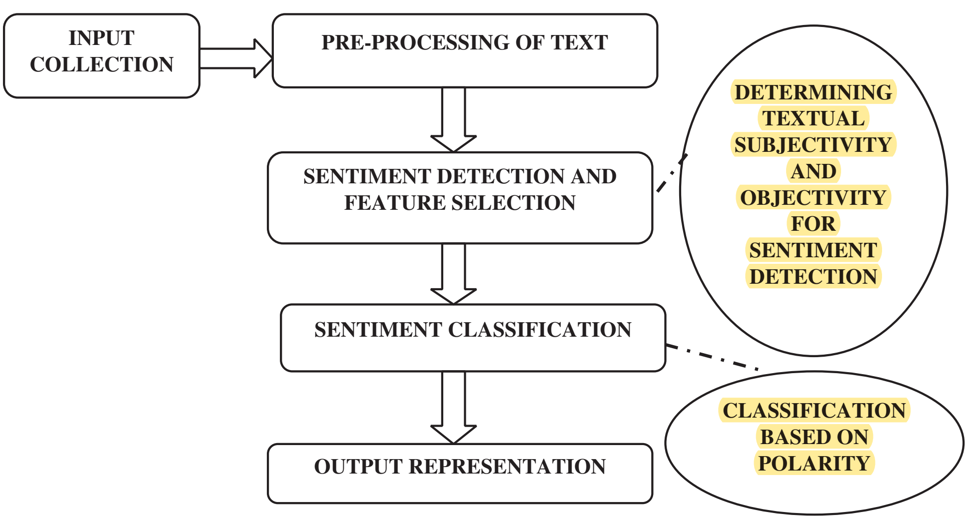


FIGURE 1 Generic sentiment analysis process

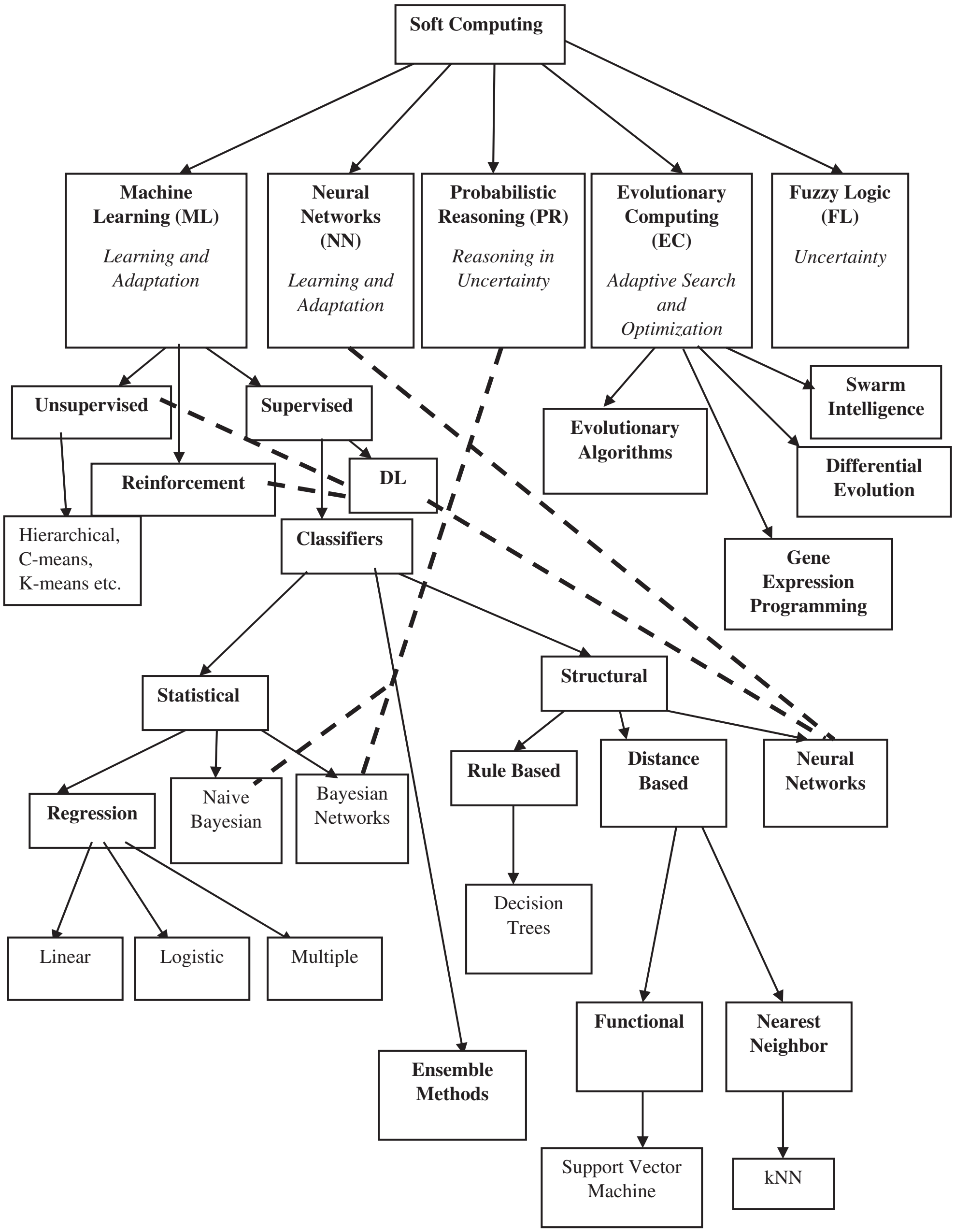


FIGURE 2 Soft Computing Techniques

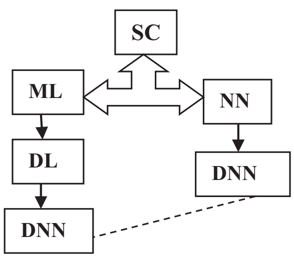


FIGURE 3 Relation between SC, ML and DL

A continuous and an evolving research trend which range from proposing new techniques to new application areas within this unison of using SC techniques for SA on Twitter has been observed in recent times and thus the need to document preliminary work and review the ongoing work is evidently established. Sophisticated and extensive surveys focusing on the techniques, applications, challenges in SA can be found in literature.[5,6,11,21-25] The contribution of this paper is significant because firstly the primary focus is to study and evaluate the use of SC techniques for SA on Twitter and secondly as compared to the previous reviews we adopt a systematic approach to identify, gather empirical evidence, interpret results, critically analyze and integrate the findings of all relevant high-quality studies (primary + secondary) to address specific research questions pertaining to the defined research domain. The next section describes the process of review adopted for this systematic study.

3 | SYSTEMATIC REVIEW PROCESS

This review was planned and conducted based on the format of Systematic Literature Review (SLR) defined by Ketchenham and Charters.[26] The review process was divided into two stages as shown in Figure 4. The first stage was referred as the Research Synthesis and Planning stage which had three phases: Research Questions Formulation, Search Strategy, Selection Criteria and the second stage was the Research Review and Results stage which included: Quality Assessment, Data Extraction and Result reporting as the three phases.

The goal of the first phase was to ascertain and formulate the research questions within the domain recognized for survey. Then in the next phase, a search strategy was designed and adopted to ascertain how the search would be conducted. This was primarily done to find and locate the relevant research studies addressing one or more research questions. The scope of the study was narrowed in the Study Selection phase by using a selection criterion known as inclusion-exclusion criteria. The worthiness of the papers was then calculated using weighted parameters in the Quality assessment phase. The purpose of the Study Selection and Quality Assessment phase was to ensure the quality and similarity of included studies, and clearly define the boundaries of the review. Post this screening and eligibility decisions on the articles, in the next phase, the data was extracted to answer the research questions to finally critically analyze the research domain to output a summarized critique which evaluates, extends, or establishes implications for practice, identify gaps and inconsistencies, if any and provide directions for future research. Thus, the following sub-sections identify the relevant RQ's which this SLR intends to answer followed by the details of selection and examination of the relevant studies to map studies which address one or more RQ.

3.1 | Research questions

The following research questions (RQs) were identified to conduct this SLR:

RQ1. Which are the most distinguished and relevant journals with published studies?
RQ2. On which datasets and domains the studies using SC techniques for SA on Twitter have been conducted?
RQ3. Which are the most frequently used SC techniques for achieving efficient results for the SA on Twitter?
RQ4. What are the widely used performance metrics to evaluate the applied techniques?
RQ5. What is the trend and impact of using SC techniques for SA on Twitter in the past decade?

3.2 | Search strategy

A strategy for exhaustive search of all studies that have been meticulously conducted on the topic was set up to find as many potentially relevant papers as possible that relate to the use of SC techniques in SA since the inception of Twitter. For this, first the research questions were broken into individual concepts to create search terms and then databases/e-portals/digital libraries to be searched were selected. The search terms were identified like Twitter, sentiment analysis, opinion mining, soft computing techniques, machine learning, supervised methods, etc, and were explored in titles, keywords and abstracts of studies to extract all related primary research studies from journals of high repute and highest relevance to the topic of study, available within five prominent digital libraries (publishers), namely, ACM, IEEE, Elsevier, Wiley and Springer. The grammatical variation of these terms such as synonyms, etc, were also used in conjunction with applying wild card for better search or/and Boolean expression were used for expanding or narrowing the sweep of the search in order to collect potentially relevant papers. The reference section of the relevant studies was also examined to extract cross-citations. Some secondary studies were also obtained. Thus, the purpose of this step was to identify, select and extract the desired essential subset of research papers for conducting review. This is often called as study selection criteria and process. These extracted studies were then subject to a selection filter which weeded out the irrelevant and redundant papers based on a criterion.

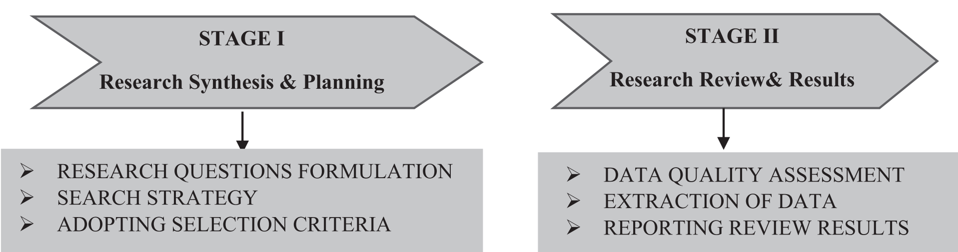


FIGURE 4 Stages of SLR

3.3 | Study selection

In this phase a selection criteria known as "Inclusion-Exclusion criteria" was adopted to limit the scope of search. It was a kind of relevance filter employed to select or reject studies. The intent was to assess all potential studies which facilitate or directly answer at least one research question within the problem domain. We focus on extracting the research articles based on search terms selected, year of publication, journal specified, citation number of the selected article, etc, with the following Inclusion-Exclusion Criteria adopted:
Inclusion criteria:

- Studies published in journals
- Studies representative of SA specifically to the micro blogging portal Twitter
- Studies focusing on the application of supervised ML algorithms like Decision Tree (DT), Support Vector Machine (SVM), Ensemble Methods (EM), k Nearest Neighbor (kNN), Linear Regression (LR), Logistic Regression (LogR), Multiple Regression (MR), etc,
- Ensemble Methods include Random Forests (RF), Bootstrap (BS), Stochastic Gradient (SGD), etc.
- Studies with supervised learning models in SC such as Probabilistic Reasoning which includes Naïve Bayesian (NB), Neural Networks (NN) (comprising Deep NN (DNN), Recursive NN, Recurrent NN, Convolutional NN, Long Short Term Memory and Deep Belief Networks), Fuzzy logic (FL), Evolutionary Computing (EC) (containing models like Genetic Algorithm (GA), etc) for SA on Twitter.
- Studies with hybrids of SC techniques for SA on Twitter.
- Studies involving the comparison of above mentioned techniques
- Studies involving SA of Twitter in English language only

Exclusion criteria:

- Studies published in conferences (Though extended versions published in considered journals were included)
- Studies which are without proper empirical analysis or benchmark comparisons
- Studies using any other social media portal like Quora, Facebook, blogs, etc
- Studies with only textual data are considered, other multimedia (image, video and audio) are not included
- Studies that are purely reviews or surveys on SA without any implementations.
- Studies with non-supervised learning model and techniques for implementing SA.
- Studies involving SA on languages other than English and multilingual SA of Twitter (for example languages like Dutch, Portuguese, Latin, Chinese, Arab, Spanish, etc) are not included.

3.4 | Quality assessment

In order to maintain the quality standard of the selected studies a careful consideration had been affirmed by taking the novelty of technique proposed and the technical content (data set and evaluation methods used). The quality check had already been ensured as we had only considered selective high quality, high impact journals from reputed digital libraries.

3.5 | Data extraction

In this phase finally the key data was extracted from the selected research articles for mapping it to the research questions. The data that was extracted from these studies involved the details about the authors, year of publication, datasets used, techniques applied, domains targeted, type of cross fold validation that was used, key performance indicators that were employed for evaluation of the techniques and the accuracy obtained. All this information was then stored in a table for further data synthesis.

3.6 | Data synthesis

The data synthesis phase summarized and interpreted the data extracted to finally output the result of the SLR as direct answers to the identified RQs using critique analysis, discussions and different representations such as tables, graphs, charts, etc. For enhancing the review process, search and study selection procedures were meticulously carried out twice in order to obtain the most relevant and appropriate studies from the literature resources. Research process started by applying the identified search terms on five selected digital libraries that resulted in 502 papers. After removing redundant studies, we obtained 487 studies. Thereafter, application of inclusion and exclusion criteria yielded 60 potentially relevant studies for further analysis.

Figure 5 depicts the overall search process applied in order to fetch the most relevant studies.

After diligent searching of the required problem statement in all the previously mentioned electronic databases 60 papers were of value to us which had high level of pertinence in response to our chosen domain. Thus we can conclude that total 60 research articles are included under primary studies and 1 research article is included under secondary study.[14]

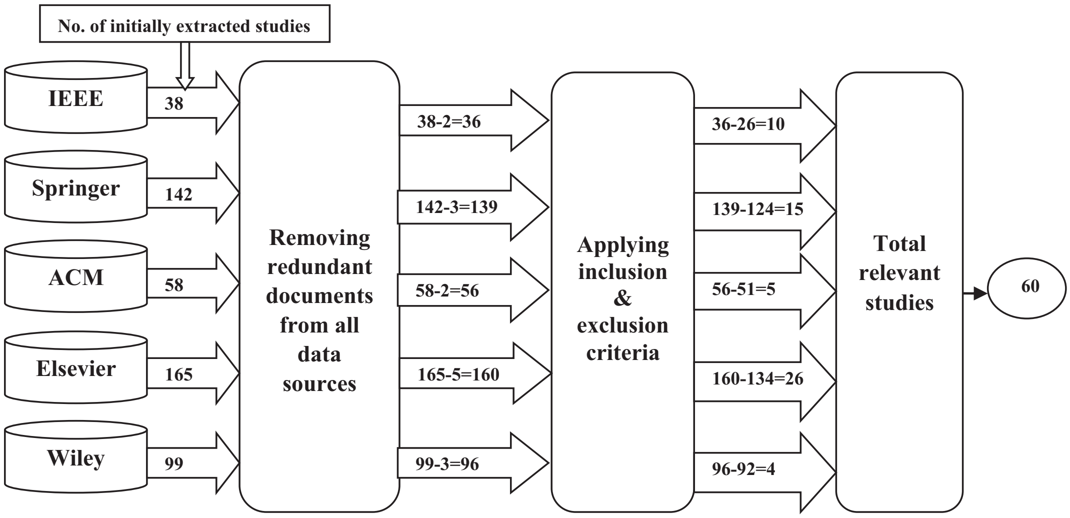


FIGURE 5 SLR Procedure

4 | LITERATURE SURVEY

The review of the final set of studies identified for this SLR on use of SC techniques for SA on Twitter is given in Table 1. As discussed in the data extraction phase, the data extracted from the selected studies included details about the authors, publication, its year of publication, datasets used, techniques applied, domains targeted, type of cross fold validation used [Ten cross (TC), Five cross (FC), Twenty cross (TWC), etc], key performance indicators and accuracy. The details about the types of dataset and domains [WePS3, SemEval, tweets prepared by Stanford University, Sentimentl40, RepLab, STS manual, Sanders Twitter sentiment corpus, Presidential debate corpus (also called as Obama McCain debate, 2008), health care reform (HCR) and generic Twitter posts, etc] and the performance parameters [Confusion matrix (CM), Precision (P), Recall (R), Accuracy (A), Average precision (AP), F score/F1 score/ F measure/ F1 measure/F1 value (F), Macro average F score/Macro F1 score (MAF), AUC (area under the curve), Cohen's Kappa (CK), Error rate (ER), Macro average error rate (MAER), Geometric Mean (GM), Specificity (Sp), Sensitivity (Sn), Normalized Mean Absolute Error (NMAE), Support (Su), Confidence (Cf), Micro average F score (MiAF), False Acceptance Rate (FAR), False Rejection Rate (FRR), Equal Error Rate (EER)] are explained in the Results and Discussion section, Section 5.

In 2012, Finn and Mustafaraj[17] proposed the use of ML techniques like NB, SVM, kNN for labeling the tweets under "political activist" and "general public" categories. NB outperformed the other two with an accuracy of around 92%. Another study was proposed by Yerva et al[27] in 2012 who discussed about the general entity matching issue pertaining to Twitter message classification using hybrid classifiers like NB and SVM for twitter using WePS-2, 3 dataset on the domain like ACL'08 (Association for Computational Linguistics Program committee members), US census data, different companies. They had combined classifiers at the pre-processing level and had observed that the combined classifiers obtained improved A of 96%.

In 2013, Lou et al[28] proposed a novel method to formulate triadic closure social relationships on Twitter and had compared it with classifiers like SVM, etc, they had observed that their approach produces A of 90% for reciprocal relationships on Twitter in comparison to SVM, etc. Another author Arias et al[29] focused on the applicability of LR, NN, SVM, DT techniques for building and evaluating the forecasting models for different time series data sets under different experimental conditions for Twitter. Trilla and Alias.[30] worked majorly on the implementation of the Unigrams so as to adapt to the SA methods with successful classifiers like SVM, NB, LogR for the news headline domain of twitter using Weka and had observed that these methods yield better and improved results in comparison to bigrams, etc, when used with successful classifiers like SVM.

In 2014, Tuarob et al[31] had discussed the usage of techniques like SVM, RF, NB for health related tweets using varied feature sets. SVM is found to perform the best with improved and better F measure of 68.47%. Morchid et al[32] worked towards the analysis and detection of massively retweeted tweets on selected features using techniques like SVM, NB. Improved results were achieved when SVM. A novel approach was presented by Montejo-Ráez et al[33] in 2014, for calculating the scoring of tweets according to their polarity using random walk analysis and comparing it with SVM. Authors approach produces P of approx. 63% in response to SVM which yielded P of approx. 64%. Author Smailović et al[34]proposed an incremental active learning approach for increasing the predictive power of sentiment classifiers for stock market with improved results via applying NB, SVM, kNN. It was observed that NB had lower performance in comparison to SVM. Boella et al[35]proposed an approach for extraction of semantic information automatically from tweets. Then such extracted tweets are fed in to SVM for providing "semantic-aware search queries." Author achieved P of 66.67% and R of 100% for the definitional tweets. In 2014, a methodology was developed by Brynielsson et al[36] for selecting and collecting natural calamity related tweets and tagging them using the annotators like happiness, anger, fear, etc. They had focused on the applicability of the classifiers like SVM, and had observed that SVM outperformed NB with an improved A of about 60%. In 2014, the author Arakawa et al[37] had proposed tweet classification random forests experimentation that analyses number of retweets and the effects of the features based on the "user type." In their work, the results claimed that the "classification by user type" depicted the overall

TABLE 1 Summary of the studies undertaken for review

S. No.	Author	Publication	Year	Techniques	Data Set	Tools	Domain	Cross validation	Performance parameters	Accuracy
1	Finn and Mustafaraj[17]	Springer, KI (Künstliche Intelligenz)	2012	NB, SVM, kNN	Author collected 15,000 political and 50,000 non-political tweets in Nov 2012 and in Sept 2012.	Weka	Politics, News	TC	A	NB achieved highest A of 92%.
2	Yerva et al[27]	Elsevier Journal of Information Systems	2012	NB, SVM	WePS-3	Matlab, open Calais, alchemy API	ACL"08, US census data, different companies.	TC	P, R and F were also used based on the TP, TN, FP and FN (false negatives).	A of 96% of the combined classifiers
3	Lou et al[28]	ACM Transactions on Knowledge Discovery from Data	2013	SVM	35,746,366 tweets from 10/12/2010 to 12/23/2010.	SVM-light.	Elite users (famous personalities, like actors, singers, etc	TC	P,R and F	TriFG method achieves 27% improvement in comparison to SVM
4	Arias et al[29]	ACM Transactions on Intelligent Systems and Technology	2013	LR, NN, SVM, DT	Author fetched all the tweets from 20 Mar to 20 Nov 2011, Jun to Aug 2011	Weka	Stock market, different companies, movies	-	A, P, R, F, CK	SVM achieved A of approx. 68%.
5	Trilla and Alias. [30]	IEEE transactions on audio, speech, and language processing	2013	SVM, NB, LogR	Semeval 2007	Weka, EmoLib	News headlines	TC	F	NB shows improvement by 7% for F as compared to the baseline rate.
6	Tuarob et al[31]	Elsevier Journal of Biomedical Informatics	2014	SVM, RF, NB	700 million tweets from April 2011 to September 2012.	LibSVM, Weka	Health related	TC	P, R, F	Improved P of approx. 63%.
7	Morchid et al [32]	ACM, Pattern Recognition Letters	2014	SVM, NB	6 million tweets from April 14th 2006 to May 13th of 2011.	Weka	General	TC	P, R, F	SVM produced R of 86.9% and P of 59.8%.
8	Montejo-Ráez et al[33]	Elsevier, Computer Speech and Language	2014	SVM	376,296 tweets from September 14th 2010 to March 19th 2011.	SVM-Light	General public board messages	TC	P, R, F.	SVM yielded P approx. 64%.
9	Smailovic et al [34]	Elsevier, Information Sciences	2014	SVM, kNN	1,600,000 tweets prepared by Stanford University. 152,570 tweets discussed about different companies like Apple, Amazon, etc from March 11 to December 9, 2011.	Pegasos SVM	Stock market exchange, companies like Apple, Amazon, Baidu, Cisco, Google, Microsoft, Netflix, people (Bobby Flay, Warren Buffet)	TC	F	The best setting yielded F was approx. 54%.
10	Boella et al[35]	Springer, Journal of Intelligent Information Systems	2014	SVM	Out of dataset of 1 million Twitter posts, only 100 random were extracted.	Weka	General	TC	P, R, F	P of 66.67% and a R of 100%

(Continues)

TABLE 1 (Continued)

S. No.	Author	Publication	Year	Techniques	Data et	Tools	Domain	Cross validation	Performance parameters	Accuracy
11	Brynielsson et al[36]	Springer, open journal of Security Informatics	2014	SVM, NB	2.3 million tweet belonging to Sandy hurricane from Oct 29 to Nov.	Weka	Natural crisis related tweets	TC	CM	Improved A of about 60% using SVM.
12	Arakawa et al[37]	Wiley, Journal of the association for information science and technology.	2014	RF (type of ML that uses DT and creates 1000 bootstrap (BS) by sampling the variables)	Dataset was created using tweets from 40 accounts of the users who were having high number of followers as of September 7, 2011 containing almost 28,756 tweets	MeCab	categories like "entertainers (comedians, "idols," actors, and sportsmen), "names" (politicians, authors, cultural icons, intelligentsia, entrepreneurs, and comic book writers), "characters," and "organizations" (enterprises, local government, etc)	TC	P, R, F	Among all the 15 experiments, experiment 2 showed P of around 84%.
13	Burnap et al[38]	Springer, Social Network Analysis and Mining (SNAM)	2014	DT, Hybrid: (NB+LogR=) BLR	Author collected 427,330 tweets over 14 days for the event (ie, terrorist event "Woolwich London" 2013),	COSMOS 1 tension engine, Senti Strength tool, VGAM package(R software, Weka	Terrorism	TC	P, R, F	The presence of a URL and a hashtag increased the rate of retweets by a factor of 1.78 (78 %).
14	Makazhanov et al[39]	Springer, SNAM	2014	NB, LogR, DT	181,972 tweets from 1 Apr 2013 to 11 May 2013 were collected	Weka	Politics (election)	TC	P, R, F	NB and LR achieved P of 93%.
15	Bogdanov et al[40]	Springer, SNAM	2014	NB	SNAP (dataset contains 467 million posts from Jun to Dec 2009). Author also collected 14.5million tweets from Mar 2006 to May 2012.	-	Business, celebrities, politics, science and sports.	TC	ER, A	Classifier achieves A of 87% and lowest ER of 0.13.
16	Lin and Margolin.[41]	Springer, EPJ Data Science	2014	LR, BS	Tweets were collected from 15 Apr to 19 Apr 2013	-	Terrorism	TWC	ER	Error rate of approx. 10% was obtained.
17	Fu and Shen.[42]	Springer, Neural Computation & Application	2014	FL	Author collected 1,242,522 tweets from 7 Oct 2009 to 13 Nov 2009	-	General	-	Su, Cf	Confidence of approx. 74% was obtained.

(Continues)

TABLE 1 (Continued)

S. No.	Author	Publication	Year	Techniques	Data Set	Tools	Domain	Cross validation	Performance parameters	Accuracy
18	Chen et al[43]	IEEE Transactions on learning Technologies	2014	NB, SVM	Author collected 19799 tweets and 39095 tweets from 01 Nov 2011 to 25 Dec 2012 and 05 Feb 2013 to 17 Apr 2013.	LibSVM, Lemur toolkit	Engineering	TC	CK, A, P, R, F, MiAF, MAF	NB outperformed SVM with A of approx. 96% for negative emotions, etc
19	Liu et al[44]	IEEE Transactions on Knowledge and Data Engineering	2015	SVM, DT, RF	Sanders-twitter sentiment Corpus. 9413 tweets about "Taco Bell" during 24-31 Jan 2011. 3238 tweets of 2008 Presidential Debate corpus.	Weka	Includes public tweet corpuses about Apple, Google, Microsoft, etc	TC	A, P, R, F	Author's proposed techniques showed improved A of about 82%.
20	Kranjc et al[45]	Elsevier, Information Processing and Management	2015	SVM	1,600,000 (800,000 positive and 800,000 negative) tweets prepared by Stanford University	SVMperf, ClowdFlows, RabbitMQ, Django, LATINO *	General *(Link Analysis and Text Mining Toolbox)	TC	A, P, R	SVM produces A of 83.01%
21	Sluban et al[46]	Springer, Computational Social Networks (springer open journal)	2015	SVM	1.6million tweets prepared by the Stanford University. 25,721positive, 23,250negative and 37,951 neutral English tweets. 2,850 positive, 5,569 negative, and 11,439 neutral environmental tweets, from January to December, 2014.	SVMperf, LATINO	General, english, environmental and energy related tweets.	TC	MAF, MAER	For all categories, best model that is the hand-labeled Domain specific model showing highest MAF of 39% and lowest ER of 52.9%
22	Burnap and Williams.[47]	Wiley, Policy & Internet published by Wiley Periodicals	2015	RFDT, SVM, Hybrid of all above.	450,000 tweets of the event "murder of Drummer Lee Rigby in Woolwich, London, UK" on May 22, 2013	CrowdFlower, Stanford Lexical Parser, Weka	online hate speech (cyber hate) regarding murder of Drummer Lee Rigby in Woolwich, London, UK in 2013	TC	P, R, F	RFDT (Random Forest Decision Tree) yields R of 55%, SVM yields 69%. P of 89%.
23	Zubiaga et al[48]	Wiley, Journal of the association for information science and technology.	2015	SVM	567,452 tweets from 348,757 varied users for 1,036 unique trending Topics, where tweets being written in 28 different languages like English, Spanish, Portuguese, Dutch, Indonesian, etc	SVM-light	news, ongoing events, memes, and commemoratives	TC	A, CK	SVM yields A of around 78%.

(Continues)

TABLE 1 (Continued)

S. No.	Author	Publication	Year	Techniques	Data Set	Tools	Domain	Cross Validation	Performance Parameters	Accuracy
24	Magdy et al[49]	Springer, SNAM	2015	NB, SVM, kNN	Author collected 4, 19.5 million tweets, from end of March to the beginning of May 2014	Weka, SVM Light	Politics, Sports, Entertainment, Science, vehicles	TC	P, R, F, A	SVM has emerge as a best performer among all with 58% P.
25	Tsytsarau and Palpanas.[50]	IEEE Transactions on Knowledge and Data Engineering	2016	SVM	7 million tweets were collected from 30 trending topics on twitter from Jun 2009 till Dec 2009.	LK tool, java	General	TC	A, P, R, F	SVM showed A of 78.9% and authors proposed method depicted the A of 82%. P of SVM is around 91%.
26	Andriotis et al[51]	IEEE Transactions on Cybernetics	2016	SVM, NB	Sentiment140 dataset	Weka	Smartphones (Samsung Fame (GT-S6810P)	TC	P, R, F	SVM showed P of approx. 78% for twitter feeds.
27	Tang et al[52]	IEEE Transactions on Knowledge and Data Engineering	2016	kNN, SVM, NN	Tweets collected were from April 1st, 2013 to April 30th 2013 Urban Dictionary, Twitter dataset from SemEval 2013 and 2014.	LibLINEAR	General	TC	A, MAF	F of SVM is 72.1% for 2013 Test and 68.% for 2014 Test. A of hybrid is 86.1% for 2013 Test and 86.% for 2014 Test.
28	Peetz et al[53]	Elsevier, Information Processing & Management	2016	DT	RepLab 2012 and 2013	Weka	Automotive, banking, universities, music	TC	F	RepLab 2013 and 2012 achieved F of 0.55 and 0.49.
29	Sulis et al[54]	Elsevier, Knowledge-Based Systems	2016	DT, RF, SVM, NB, LogR	12,532 Tweets (of Task 11) of SemEval-2015.	Weka	Comedians, etc	TC	F	F of RF for #irony vs #sarcasm is 69.8%, for #irony vs #not, Is 75.2%, for #sarcasm vs #not, is 68.4%.
30	Wu et al[55]	Elsevier, Information Sciences	2016	SVM, NB, LogR	Sanders Twitter sentiment dataset. STS-manual. SemEval 2013.	Matlab R2009b	Companies like Apple, Google, Twitter and Microsoft, etc	TC	A	SVM obtains A of 79%, 82% and 78% for STS, Sanders, and SemEval datasets.

(Continues)

TABLE 1 (Continued)

S. No.	Author	Publication	Year	Techniques	Data Set	Tools	Domain	Cross Validation	Performance Parameters	Accuracy
31	Lo et al[56]	Elsevier, Decision Support Systems	2016	FL, LR, hybrid: SVM + BS, another hybrid: SVM + Bagging.	124,462 Tweets belonged to samsungsg (Twitter account for Samsung Singapore) from 2 Nov 2012 to 3 April 2013. 57,114 tweets of ilovedealssg "(Twitter account for daily deals, promotions and discounts in Singapore) from 26 March 2013 to 15 July 2013. 11,969 tweets of be aqua fitness" (Twitter account of a company focusing on aqua fitness solutions in South East Asia) from 05 Jan 2013 to 11 Nov 2015.	OpenCalais, LibSVM implementation of RapidMiner	Mobiles, fitness, healthy living, daily deals and discounts	TC	P, AP	Hybrid (SVM + EM) obtained F of 98% for samsungsg. 97% for ilovedealssg dataset, 98% for beaquafitness dataset.
32	van Zoonen and Toni.[57]	Elsevier, Computers in Human Behavior	2016	SVM, NB, LogR	578,803 tweets in dutch language being sent by 443 employees who worked in various organizations and work with an average of 39.62 hours per week for an organization having at least thirty employee.	scikit-learn	government/public administration, education/ science, health care, business services, trade/commercial services, industry, financial services, etc	TC	A, R, P, AUC	SVM achieved A of 81%.
33	Wang et al[58]	Springer, Eurasip Journal on Wireless Communications and Networking	2016	SVM, kNN, LogR, NB	Author collected from Twitter tweets between Mar. 1st and May 1st of 2015	scikit-learn	News related subjects	FC	P, R	SVM has achieved A and R of 89.8% and 89%.
34	Celli et al[59]	Elsevier, Information Processing and Management	2016	LogR, RF	Gold Standard,	-	News	TC	P, R, F	61.7% F was obtained.
35	Igawa et al[60]	Elsevier, Information Sciences	2016	RF, NN	Author collected all the tweets of the FIFA World Cup 2014	-	Sports	TC	A, P	RF achieved 88.7% A.
36	Korkmaz et al[61]	Springer, SNAM	2016	LogR	500 million tweets were collected from Nov 2012 to Aug 2014.	-	Social and political	-	P, R, F	Average F scores are in range 0.68–0.95.

(Continues)

TABLE 1 (Continued)

S. No.	Author	Publication	Year	Techniques	Data Set	Tools	Domain	Cross validation	Performance parameters	Accuracy
37	Burnap and Williams.[62]	Springer, EPJ Data Science	2016	SVM, RF	1803 tweets of sexual orientation, 1876 tweets of racism, 1914 tweets of disability were collected on 30 Apr 2013, 6 Nov 2012, 29 Aug 2012.		Cyber hate	TC	P, R, F	The mean precision of the individual classifiers for cyber hate was 0.85.
38	Oliveira et al[63]	Elsevier, Expert Systems With Applications	2016	MR, NN, SVM, RF	Author collected 31 million tweets from 22 Dec 2012 to 29 Oct 2015.	R-Tool, Mongodb, Stanford CoreNLP	Stock Market	TC	NMAE	SVM majorly produced the most accurate results.
39	Perikos and Hatzilygeroudis.[64]	Elsevier, Engineering Applications of Artificial Intelligence	2016	NB	Author collected 250 random tweets	Python's NLTK toolkit	News	-	P, A, Sp, Sn	NB achieved A of approx. 85%
40	Brocardo et al[65]	Wiley, International Journal for Communication System	2016	SVM	Author collected 3194 tweet as on before 6 Nov 2013	Weka	General	TC	FAR, FRR, EER	The best setting yielded error rate as 10.08%
41	Bouazizi and Ohtsuki.[66]	IEEE Access	2016	SVM, kNN, RF	Author collected 7628 tweets from Dec 2014 to Mar 2015.	Apache OpenNLP, Weka, LibSVM	General	-	A, P, R, F	SVM obtained the highest P of 98%.
42	Farias et al[67]	ACM Transactions on Internet Technology	2016	NB, DT, SVM	Author collected more than 30,000 sarcastic or ironic tweets	Weka	Education, humor, politics, news	TC	F	SVM achieved highest F of 0.90
43	Sintsova and Pu.[68]	ACM Transactions on Intelligent Systems and Technology	2016	NB, LogR	Author extracted 52218 tweets	Weka, LibLINEAR	Sports	-	A, P, R, F, MAF	Highest macro R is obtained by LogR
44	Nair et al[69]	Elsevier, Computers and Electrical Engineering	2017	DT	processed.cleveland.data from Heart Disease Data Set of UCI machine learning repository where a user tweets the necessary data like age, heart rate, etc	Apache Spark's machine learning library, MLlib written with Scala	Health (heart diseases)	70:30	A (Author took 70% data for training and 30% for testing.)	It yields higher accuracy in less time and free of cost.
45	Cui et al[70]	Springer, Expert Systems With Applications	2017	SVM	132.6 million tweets in April, May and June 2014 by 23.2 million accounts were considered	LibSVM	NGOs, charities, events, journalists/ bloggers (freelance media professionals or news agencies), celebrities, politicians, sportsmen, etc	80:20	P, R, A, F (Author took 80% data for training and 20% for testing.)	Improved results were obtained using enhance distant based supervised algorithm along with SVM.

(Continues)

TABLE 1 (Continued)

S. No.	Author	Publication	Year	Techniques	Data Set	Tools	Domain	Cross validation	Performance parameters	Accuracy
46	Pérez-Gállego et al[71]	Elsevier, Information Fusion	2017	NB, LogR, SVM	1, 60 0, 0 0 0 tweets with emoticons were collected from April 6, 2009 to June 16, 2009.	LibLINEAR and LibSVM	General	FC	GM	Improved results were obtained using NB.
47	Alsine et al[72]	Elsevier International Journal of Approximate Reasoning	2017	SVM	Author collected tweets belonging to the month of Mar 2016 and Apr 2017	natural language Toolkit (NLTK)	Taxation, politics, public campaign	-	A	Improved A approx. 60% obtained for SVM.
48	Jianqiang and Xiaolin.[73]	IEEE Access	2017	SVM, NB, LogR,RF	Stanford Twitter Sentiment, SemEval 2014, Stanford Twitter Sentiment Gold (STS-Gold), Sentiment Strength Twitter(SS-Twitter), Sentiment Evaluation (SE-Twitter)	scikit-learn	-	TC	A, F	NB achieved highest F of 0.37 for SemEval 2014.
49	Jain and Kumar[74]	Journal of Computational Science, Elsevier	2017	SVM, NB, LogR	Author collected tweets from Sept 2016 to Nov 2016 using Twitter API.	LibLinear	Health	TC	F, P, R, A	Performance of SVM was observed to be better as compared to NB.
50	Keshavarz and Abadeh[75]	Knowledge-Based Systems Elsevier	2017	GA	Author had utilized benchmark datasets, namely Sanders, Presidential debate corpus, Healthcare Reform (HCR), SemEval 2013 and Stanford.	-	Companies like Apple, Google, Twitter and Microsoft, etc, Politics (Elections), Health.	TC	F, A, P, R	Accuracy of more than 85% was obtained.
51	Xiong et al[76]	Neurocomputing Elsevier	2017	SGD SVM, NB-SVM (NB enhanced SVM), NN, CNN	SemEval 2013			TC	MAF	CNN yielded improved results with MAF score of around 85%.
52	Neppalli et al[77]	International Journal of Disaster Risk Reduction, Elsevier	2017	NB, SVM	Author collected geo-tagged tweets from Hurricane Sandy collection comprising of 74,708 tweets with geo-location.	SentiStrength	Environmental crisis	TC	A	SVM produced enhanced results A of with around 76%.
53	Singh et al[78]	Transportation Research, Elsevier	2017	SVM, NB	Author collected random 10,664 tweets using Twitter handlers.	-	Food (health)	FC	A	It was observed that performance of SVM was better than NB.
54	Xiaomei et al[79]	Knowledge-Based Systems, Elsevier	2017	SVM, NB,	Author had utilized benchmark datasets, namely Sanders, Presidential debate corpus and Healthcare Reform (HCR).	-	Health, Obama, Republicans, Democrats, conservatives, liberals, elections, politics, Tea Party	FC	A	SVM yielded improved accuracy.

(Continues)

TABLE 1 (Continued)

S. No.	Author	Publication	Year	Techniques	Data Set	Tools	Domain	Cross Validation	Performance Parameters	Accuracy
55	Khan et al[80]	International Journal of Information Technology, Springer	2017	NB	Author collected 20,000 political and non-political.	Ipython notebook, Apache Spark, Python nltk library	Politics	-	P	Author proposed algorithm achieved highest A of 85%.
56	Bouazizi and Ohtsuki[81]	IEEE Access	2017	RF	Author collected 21,000 tweets for training and 19740 tweets for testing purposes.	SENTA	Random	FC	A,P,R, F	RF achieved A of 60.2% for multi-class SA.
57	Li et al[82]	Information Systems, Elsevier	2017	NB, DT	Author collected 196,370 tweets and classified them into 20 classes.	MongoDB, Java	Stock Market	TC	A	NB yielded A of more than 72%.
58	Jianqiang et al[83]	IEEE Access	2018	SVM, CNN	Author used benchmark datasets namely STS-Test, STS-Gold, SS-Twitter, SE-twitter.	GloVe	Random	TC	A, P, R, F	GloVe-CNN highest A achieved of around 87.62% using STS dataset.
59	Ghiassi and Lee[84]	Expert Systems with Applications, Elsevier	2018	NN, SVM	Author collected around 40,000 tweets from 8th Jan 2013 to 11th April 2013 related to Starbucks, Governor Christie, Southwest airlines and Verizon.	WEKA, Java, MS SQL Server	Consumer products and services, Politics, Entertainment.	DAN2	P, R, F	Author achieved domain transferability for different datasets. SVM yielded enhanced results.
60	Symeonidis et al[85]	Expert Systems with Applications, Elsevier	2018	NB, SVM, LogR, CNN	Author used benchmark datasets namely SS-Twitter and SemEval 2013-2017.	NLTK, Sklearn	Random	-	A	Author achieved best results with CNN.

best performance. It was also observed that information roles and function words are important aspects in the retweeted classes and analyzing number of features is important in determining user types. Among all the 15 experiments, experiment 2 showed precision of around 84%.

A model was built by Burnap et al[38] that predicted information flow size and survival on Twitter following a terrorist event via the action of retweeting using DT, BLR techniques. The novel findings were the time lags between retweets, the co-occurrence of URLS and hashtags, and the sentiment expressed in the tweet. Makazhanov et al[39]predicted user political preference from their Twitter behavior towards political parties for 2012 Albertan and 2013 Pakistani elections. They build prediction models based on a variety of behavioral and contextual features using NB, LR, DT techniques. A genetically inspired framework was proposed by Bogdanov et al[40]for modeling individual social media users which they termed a genotype. They extracted topic-specific influence backbone structures based on content adoption and further showed that genotype model with combination of NB enable more than 20% improvement. Lin and Margolin[41] studied the expression of fear and social support in Twitter communication during and after a terrorist attack using methods like LR, BS. Using nearly all geo-tagged tweets they had examined the temporal correlation in these expressions. Their findings suggests that not all fear is necessarily bad and could be considered interesting for general prospects of terrorism as a strategy for political change in the era of social media. A novel method was proposed by Fu and Shen[42] for extracting useful behavioral trends of users on Twitter using the mass assignment theory based fuzzy association rules. The paper uses FL in developing the new scheme and gave improved results. Another work focusing on the usage of techniques like NB, SVM, etc, was proposed by Chen et al[43] for depicting the student understandings, issues, problems and challenges faced by them during their studies using social media like Twitter. In 2015, a model was developed by the author Liu et al[44] for fetching unlabeled tweets from a mixed group of labeled and un-labeled tweets for maintaining the dynamism of the selected tweets and classifying them depending on the trends of the topics selected with improved A using SVM, DT, RF classifiers. Kranjc et al[45] focused on the implementation of the active learning scenario for twitter using cloud based data mining platform with improved performance by classifying tweets as positive and negative only, ie, two way classification. SVM produces A of 83.01%. Sluban et al[46] had proposed a work that divided sentiment model into three categories as Smiley-labeled general, Hand-labeled general, and Hand-labeled domain-specific sentiment model utilizing the already processed negative and positive tweets. They had further proved that the "high-quality domain specific tweets" provides a much better sentiment model despite the number of available tweets for it. They had implemented SVM. Author observed that among all the three categories stated, best model that outraged others is the hand-labeled domain specific model that showed lowest ER of 52.9% and the highest MAF of 39% on the test set. A classifier was developed by Burnap and Williams[47] for monitoring public reactions to emotional hateful events like death of Rigby using SVM, RF, DT and hybrid of these. Prime aspect of the author was to evaluate the hybrid classifier which could be further used by policymakers for effective and efficient decision making process for such cyber hate on social media. Based on the implementation results, author lead to the conclusion that ensemble classification process is most effective and efficient for classification of such cyber hate events, provided the current feature sets. Zubiaga et al[48] proposed a method so as to efficiently categorize trending topics irrespective of the need of any external data using SVM and it was observed that SVM yields A of around 78%. Magdy et al[49] experimentally demonstrated the effectiveness of a "distant supervision" approach to tweet classification, consisting in automatically obtaining labeled data from one social media platform (YouTube) and using this data for training a classifier for another such platform (Twitter) using kNN, SVM, DT techniques.

Tsytsarau and Palpanas[50] had focused on the aggregation of the large real time datasets having diversified sentiments and thereafter had developed a model that performs the sentiment contradiction diversification at different time scales. Author had used a novel data structure which is incrementally maintained and helps in scaling large amount of datasets, often called as contradiction tree. The results claim that the SVM had shown improved results in terms of measuring the contradiction level and ranking of the dataset. In 2016, another study depicting the investigation of the effect of textual data on the short message services (SMS) so as to perform SA on the smart-phone users for revealing the mood trends in them and comparing them with the twitter feeds was given by Andriotis et al[51] They had majorly focused on the data being stored in the internal storage of our smart phones and illustrating inter-connections within the entities at all the levels of the ecosystem and their approach primarily targets the data that is found in the smart phones, which is linking the users to the universal digital community. The results claimed that the P and F score of SVM was just more or less comparable to other classifiers being implemented by the author like NB. Tang et al[52] had implemented a recursive neural network and convolution neural network with dedicated loss functions so as to record sentiments of sentences or words as well as contexts of those words for learning word embedding, ie, author had developed NN based ranking model for learning sentiment embedding by utilizing sentence level sentiment information as "task-specific evidences." In their work, author had used urban dictionary in order to make clusters of all related words together and then applied kNN classification to these clusters so as to classify them into positive, negative and neutral clusters of high quality similar words sentiments. Peetz et al[53] had focused on the estimation of the polarity of the tweets in terms of reputation, for which DT was used for combining, learning and finding the optimal number of features in a set. In the preliminary experimentation, author had observed that the SVM and RF showed poor performances in comparison to DT for varied selected features. The results exclaimed that the DT was more successful in making decisions when applied to the tweets belonging to domains like automotive, banking, universities, music, etc.

Sulis et al[54] had briefly discussed about the figurative content of the tweets such as hashtag's "for not, sarcasm and irony" by applying techniques like DT, RF, SVM, NB, LogR. Author's aim was to explore all the differentiating traits among these figurative tweet contents. In their work, it was observed that the RF gained the highest F score and DT got the lowest F score for all the #tag combinations for figurative intent tweet messages. The best result had been observed by RF for the case of #irony vs #not classification, being approx. 75.2% which profoundly provides better insight into the use of these types of hashtags for labeling the tweets (whether they are ironical, or sarcastic, etc) being twittered on social media. In 2016, author Wu et al[55] had worked towards extraction of the useful sentiment oriented knowledge from the unlabeled tweets in order to improve and enhance the microblog sentiment classification using SVM, NB, LogR. The experimental results proved that the author's approach improved the process of sentiment classification effectively by reducing the dependency on the labeled data. Lo et al[56] had worked towards establishment of the top-n followers and ranking them which could eventually help the companies (belonging to mobiles, fitness, healthy living, daily deals and discounts) and promoters to publicize their businesses on Twitter. Author had implemented the concept by applying methodologies like LR, Fuzzy logic, hybrid: BS ensemble using SVM model, another hybrid: bagging ensemble using SVM models. van Zoonen and Toni[57] proposed an approach that enabled to perform the analysis of the entire tweet texts and thereby helped in reducing the risk involved with the sampling errors. Author established that it could be applied to other social media content as well for varied topics in demand using multiple method approaches. SVM outperformed the methods like NB, LogR and produced acceptable and higher level values for all the performance parameters when applied to it thus yielding highest reliability statistics. In 2016, another author sisWang et al[58] had focused on the determination of "multivariate emotional model classification." Author had also applied deep learning for the "entity recognition" via using "SENNA" deep learning toolkit. Author applied classifiers like SVM, kNN, LogR, NB. SVM has emerged to give most promising results in comparison to other methods applied. It produced a lower error rate and higher accuracy. SVM has achieved accuracy and recall of 89.8% and 89% respectively which is higher in regard to other classifiers.

Celli et al[59] had analyzed the role of personality and communication styles in the diffusion of news articles using LR, RF. They had automatically annotated personality types and communication styles of Twitter users and analyzed the correlations between personality, communication style, Twitter metadata (such as following and followers) and the type of mood associated to the articles they shared. Another study demonstrating a wavelet-based approach for account classification was given by Igawa et al[60] that detects textual dissemination by bots on an Online Social Network. Their main objective was to match account patterns with humans, cyborgs or robots, improving the existing algorithms that automatically detect frauds. Experiments were performed using a set of posts crawled during the 2014 FIFA World Cup, obtaining accuracies within the range from 94 to 100% via RF and NN (multilayer perceptron). Korkmaz et al[61] had presented a model for predicting civil unrest through the combination of heterogeneous online data sources and provide a critical evaluation of the approach via implementing LogR. They had evaluated the predictive power of disparate datasets and methods, and provide interpretable insights into unrest events. Another author Burnap and Williams[62] had developed novel machine classification models to identify different types of cyber hate individually. The resulting cyber hate classification models have been shown to be applicable to a range of protected characteristics including race, disability and sexual orientation, and provide new ability to automatically identify content perceived by a group of human annotators as hateful or antagonistic. They had implemented SVM and RF. A model was proposed by Oliveira et al[63] for assessing the impact of tweets on stock market variables such as returns, volatility, etc. The author had applied ML techniques such as MR, NN, SVM and RF to detect whether the predictions based on sentiments are influential on the stock market or not. Another study for deriving dependencies and the emotional state of the sentences was put forward by the author Perikos and Hatzilygeroudis.[64] This ensemble classifier was implemented on a dataset of news headlines and Twitter posts to reveal the best performer with higher A and P using NB. A stylometric analysis technique in continuous authentication was explored by Brocardo et al.[65] It proceeds by breaking an online document into a sequence of short texts on which the CA decisions happen. The method yielded promising results with an equal ER varying from 8.21% to 16.73%. Bouazizi and Ohtsuki[66] proposed the implementation of techniques like SVM, kNN, RF for detection of sarcastic comments using pattern based features. Farías et al[67] proposed the usage of techniques like NB, DT, SVM for differentiating between the ironic and the non-ironic content using affective information. Sintsova and Pu[68] had focused on the usage of NB and LogR for building emotion classifiers.

Nair et al[69] had developed a health monitoring application for prediction of heart diseases based on spark cluster ML model using DT methodology for prediction of health status of an individual by harnessing real time data from Twitter. The application has been deployed using Cloud in "Amazon Elastic Compute Cloud (EC2)." Cui et al[70] had formulated the use of SVM together with other distant supervised classification algorithms for classifying Twitter accounts as Branding and Personal account types without the involvement of any manual labeling. Pérez-Gállego et al[71] had focused on the usage of the EM together with NB and SVM for incorporating binary quantification. The results showed that the better performances were obtained by using the ensemble versions. Alsine et al[72] had presented a model based on valued abstract argumentation for automatically labeling the relationship between the sentiments via implementing SVM. It also reasons about the accepted and the rejected sentiment tweets for the controversial discussions on Twitter. Jianqiang and Xiaolin[73] discussed about the performance of the classifiers like NB, SVM, LogR, RF on five different benchmark Twitter datasets. The results indicate that NB and RF are more sensitive in comparison to the other classifiers for varied pre-processing methods. Jain and Kumar[74] in 2017, had applied SVM, NB and LogR to health domain for SA. The results were evaluated using precision, accuracy, recall and F measure parameters. The highest accuracy was obtained by SVM. Keshavarz and Abadeh[75] demonstrated the applicability of genetic algorithm (GA) model for SA to benchmark datasets namely, Sanders, Presidential debate corpus, Healthcare Reform (HCR), SemEval 2013 and Stanford. Improved results were obtained using GA. Xiong et al[76] applied soft computing techniques such as Stochastic Gradient Descent (SGD), SVM, NB-SVM (NB enhanced SVM), MLP and CNN to SemEval 2013 benchmark corpus using ten-fold cross validation. CNN yielded improved results with MAF score of around 85%. Neppalli et al[77] collected geo-tagged tweets from Hurricane Sandy Collection for analyzing sentiments of the tweets belonging to the environmental crisis. Author gathered around 74, 708 tweets with geo-location using SentiStrength and applied SVM and NB. SVM produced enhanced results with A of around 76%. Another similar work was given by Singh et al[78] in 2017 where the author applied SVM and NB to analyze sentiment polarity of the tweets belonging to health domain. It was again observed that performance of SVM was better than NB. Xiaomei et al[79] utilized benchmark datasets, namely Sanders, Presidential debate corpus and Healthcare Reform (HCR) for SA using five-fold cross validation. The tweets belonged to varied domains such as Health, Obama, Republicans, Democrats, conservatives, liberals, elections, politics, etc. SVM yielded improved accuracy. Khan et al[80] demonstrated the applicability of NB for SA by collecting around 20,000 political and non-political tweets as dataset. The results were evaluated using precision efficacy criterion. Bouazizi and Ohtsuki[81] applied ensemble method namely RF for SA. Author collected 21,000 tweets for training and 19740 tweets for testing purposes using FC validation technique. RF achieved A of 60.2% for multi-class SA. Li et al[82] focused on the use of NB and DT for SA for the tweets belonging to stock market exchange domain. NB yielded A of more than 72%.

Jianqiang et al[83] in 2018 applied SVM and presented a deep neural network model namely CNN for SA using GloVe. Author utilized benchmark datasets namely STS-Test, STS-Gold, SS-Twitter, SE-Twitter for SA. GloVe-CNN achieved highest A of around 87.62% using STS data set. Ghiassi and Lee[84] analyzed sentiment polarity using techniques such as NN and SVM. Author collected around 40,000 tweets from 8th Jan 2013 to 11th April 2013 related to Starbucks, Governor Christie, Southwest airlines and Verizon. Author achieved domain transferability for different datasets. SVM yielded enhanced results. Symeonidis et al[85] applied soft computing techniques namely NB, SVM, LogR and CNN for analyzing sentiments. Author used benchmark datasets namely SS-Twitter and SemEval 2013-2017. Amongst all, CNN yielded best accuracy.

Thus, we can infer that very few studies exists that demonstrates the application of DL methods for SA on Twitter, making it completely a potential area for research and development in the field of SA. Researchers and academicians are open to substantiate the influence of DL for SA on social media such as Twitter.

5 | RESULTS AND DISCUSSION

In this section, we discuss the results obtained as answers to the RQs defined in the SLR. Table 2 depicts the mapping of the research articles to the respective RQs they address.

- *Most distinguished and relevant journals with published studies in the identified research domain (RQ1)*

Reviewing the literature exhaustively helped us identify the journals within the selected digital libraries and is prominently publishing research in our domain. Table 3 demonstrates the distribution of the research articles in the identified journals, their proportions and corresponding cumulative proportions of the studies included in this SLR.

The pie chart in Figure 6 shows the distribution of the published papers in the selected journals from the five decided digital libraries. In the past decade, the percentage of published studies on use of SC for SA on Twitter is the highest in journals of Elsevier (43%), followed by Springer (25%), IEEE (17%), ACM (8%) and Wiley (7%).

Table 4 suggests the most relevant and distinguished journals in this domain of study. Amongst the final selected articles for this SLR, majority of them belonged to Elsevier (Knowledge-Based Systems) and Springer (Social Network Analysis and Mining (SNAM)), followed by IEEE Access. Next comes, IEEE Transactions on Knowledge and Data Engineering, Elsevier (Information Sciences and Information Processing and Management), followed by Springer (EPJ Data Science), ACM Transactions on Intelligent Systems and Technology and Wiley Online Library (Journal of the Association for Information Science and Technology Journal).

- *Widely used datasets and domains in which the studies for SC techniques in SA on Twitter have been conducted (RQ2)*

Reviewing the pertinent literature it was observed that the research studies have considered a gamut of data sources including benchmark datasets. The probed datasets was either benchmarked or a set of random tweets collected in real-time within a selective topic/subject/domain. The various benchmark datasets used in the past decade were WePS-3,[27] SemEval,[30,52,54,55,73,75,76,85] tweets prepared by Stanford University,[34,45,46,75] SNAP,[40] Sanders Twitter Sentiment Corpus (denoted as Sanders),[44,55,75,79] 2008 Presidential Debate Corpus,[44,75,79] Sentiment140,[51] RepLab 2012,[53] RepLab 2013,[53] STS-manual,[55] Gold Standard personality labeled Twitter dataset,[59] Cleveland Heart Disease data,[69] STS-Gold,[73,83] SS-Twitter,[73,83,85] SE-Twitter,[73,83] STS-Test,[73,83] HCR.[75,79]

TABLE 2 Mapping of RQ's with the relevant research articles

S.No.	RQ number addressed	Research Reference
1	RQ1	34,38-40,44,45,49,50,52,53,55,59-61,75,79,81,83-85
2	RQ2	17,27-29,31,33,34,36,38-41,43-49,51-64,67-70,72-80,82-85
3	RQ3	17,27-36,38-40,43-52,54-58,62-68,70-74,76-85
4	RQ4	27-35,37-39,43,44,47,49-51,53,54,56-62,64,66-68,70,73-75,77-85
5	RQ5	17,27-85

TABLE 3 Distribution of the research articles in regard to the respective journals

Rank	Journal name	Description	#papers	Proportion (%)	Cumulative Proportion (%)
1	Elsevier	• Elsevier, Journal of Information Systems	1	43	43
		• Elsevier, Journal of Biomedical Informatics	1		
		• Elsevier, Computer Speech and Language	1		
		• Elsevier, Information Sciences	3		
		• Elsevier, Information Processing and Management	3		
		• Elsevier, Knowledge-Based Systems	5		
		• Elsevier, Decision Support Systems	1		
		• Elsevier, Computers in Human Behavior	1		
		• Elsevier, Expert Systems With Applications	1		
		• Elsevier, Engineering Applications of Artificial Intelligence	1		
		• Elsevier, Computers and Electrical Engineering	1		
		• Elsevier, Information Fusion	1		
		• Elsevier, Journal of Approximate Reasoning	1		
		• Elsevier, Journal of Computational Science	1		
		• Elsevier, Neurocomputing	1		
		• Elsevier, Journal of Disaster Risk Reduction	1		
		• Elsevier, Transportation Research	1		
		• Elsevier, Information Systems	1		
2	Springer	• Springer, KI - Künstliche Intelligenz	1	25	68
		• Springer, Journal of Intelligent Information Systems	1		
		• Springer, open journal of Security Informatics	1		
		• Springer, Social Network Analysis and Mining (SNAM)	5		
		• Springer, EPJ Data Science	2		
		• Springer, Neural Computation & Application	1		
		• Springer, Computational Social Networks	1		
		• Springer, Eurasip Journal on Wireless Communications and Networking	1		
		• Springer, Expert Systems With Applications	1		
		• Springer, Journal of Information Technology	1		
3	IEEE	• IEEE Transactions on Audio, Speech, and language processing	1	17	85
		• IEEE Transactions on Learning Technologies	1		
		• IEEE Transactions on Knowledge and Data Engineering.	3		
		• IEEE Transactions on Cybernetics.	1		
		• IEEE Access	4		
4	ACM	• ACM Transactions on Knowledge Discovery from Data	1	8	93
		• ACM Transactions on Intelligent Systems and Technology	2		
		• ACM, Pattern Recognition Letters	1		
		• ACM Transactions on Internet Technology	1		
5	Wiley	• Wiley, Journal of the Association for Information Science and Technology	2	7	100
		• Wiley, Policy & Internet published by Wiley Periodicals.	1		
		• Wiley, International Journal for communication system	1		

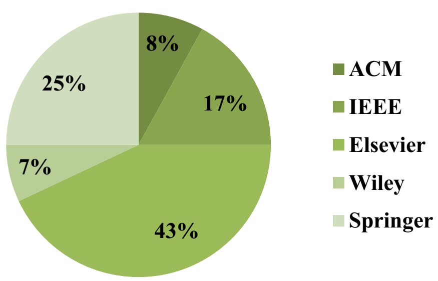


FIGURE 6 Distribution of papers in accordance to the digital libraries (expressed in percentages)

TABLE 4 Relevant and distinguished journals

Rank	Journal Name	# Papers
1	• Elsevier, Knowledge-Based Systems	5
	• Springer, Social Network Analysis and Mining (SNAM)	5
2	• IEEE Access	4
3	• IEEE Transactions on Knowledge and Data Engineering	3
	• Elsevier, Information Sciences	3
	• Elsevier, Information Processing and Management	3
4	• Springer, EPJ Data Science	2
	• ACM Transactions on Intelligent Systems and Technology	2
	• Wiley, Journal of the Association for Information Science and Technology	2

WePS-3[27] is available at http://nlp.uned.es/weps/weps-3/data and consists of tweets belonging to fifty companies. SemEval is a series of "computational semantic analysis" systems comprising of various tasks. The SemEval 2007[30] dataset comprises of tweets belonging to the news headlines of the major newspapers consisting of 1250 headlines. The SemEval 2013 is available at http://www.cs.york.ac.uk/semeval-2013/task2/ and consists of 6,237, 1654 and 3813 manually annotated tweets in the training, development and test sets.[52,55,75,76,83] The SemEval 2014 dataset includes data related to ten different tasks evaluating varied "computational semantic systems," provided by the SemEval 2014 conference.[52,73,83] The SemEval 2015[54,83] is available at http://alt.qcri.org/semeval2015/task11 and consists of tweets which are enriched with metaphors and ironical content. Similarly, SemEval series provides SemEval 2016[85] and SemEval 2017[85] set of tweets as well belonging to varied topics such as consumer products, environmental issues, etc. Another benchmark dataset[34,45,46,75] SNAP dataset[40] is a collection of 1,600,000 positive and negative tweets prepared by the "Stanford University" which focus on the "general purpose smiley tweets.". consists of 467 million tweets including a network wide view for a period of around six months from June to Dec 2009. The tweets belonged to the domains like business, celebrities, politics, science or technology and sports. "Sanders Twitter Sentiment Corpus" referred to as Sanders[44,55,75,79] comprises of 5513 manually annotated tweets available at http://www.sananalytics.com/lab/twitter-sentiment/. The tweets belonged to four different companies which are Apple, Google, Twitter and Microsoft. The 2008 Presidential Debate Corpus[44,75,79] comprises of the 3238 tweets by Amazon Mechanical Turk related to the presidential elections with sentiment judgments, also called as Obama-McCain debate. It had been the most widely viewed presidential debate. Sentiment140[51] is a corpus comprising of the sentiments associated with any topic, brand or product on Twitter, available at http://help.sentiment140.com/home. RepLab 2012[53] and RepLab 2013[54] standard benchmark datasets comprises of the tweets depicting the sentiments associated with the reputation of the companies or any organizations or celebrities, etc, on the Twitter. The STS-manual dataset[55] comprises of 498 tweets belonging to people, product, companies, etc. Gold Standard personality labeled Twitter dataset consisting of the personality types of the 118 males and 92 females aged between 14 to 65 years by means of short BFI-10 personality test.[59] Cleveland Heart Disease data[69] is a data corpus found in the UCI machine learning repository comprising of fourteen variables measured on 303 heart disease patients. Stanford Twitter Sentiment Test[73,83] (denoted as STS-Test) dataset consists of in totality 498 tweets. Out of which, 177 are negative, 182 are positive and 139 are neutral tweets belonging to products, companies and people. Stanford Twitter Sentiment Gold[73,83] (referred as STS-Gold) corpus consists of 2034 tweets. Among those 1402 are negative and 632 are positive labeled tweets. It also contains 27 positive, 13 negative and 18 neutral entities. Sentiment Strength Twitter[73,83,85] (denoted as SS-Twitter) is a dataset comprising of 4242 tweets. Out of which, 1037 are negatively labeled, 1953 are neutral and 1252 are positively labeled tweets. The original dataset is available at http://sentistrength.wlv.ac.uk/documentation/. Sentiment Evaluation Twitter[73,83] (referred as SE-Twitter) comprises of 6745 tweets being manually annotated by 3 Mechanical Turk workers. Out of which, 990 were labeled as negative tweets, 4097 as neutral tweets and 1658 as positive tweets. HCR[75,79] is a dataset comprising of 621 tweets for training and 665 tweets for testing purposes. The tweets were gathered in March 2010 and were two way labeled, as either positive or negative.

Many reported researches were carried on the tweets fetched directly from Twitter using its API. The tweets were from a variety of domains, topics and time period (referred as topic specific/topic oriented tweets). These prominently included tweets from or about elite personalities like actors; singers; sportsperson; comedians; politicians, authors, idols; entertainers,[28,34,37,40,54,55,67,70,73,75,79] etc, news and commemoratives,[17,30,48,58,59,64,67,79] health and fitness,[31,56,57,69,74,75,78,79] stock market exchanges,[29,34,63,82] companies like AT&T; Amazon; Apple; Google; Microsoft,[27,29,34,44,55,73,75] consumer products like kindle, smart-phones, etc,[51,55,56,73,84] natural calamities, energy and environmental related,[36,46,77] cyber hatred,[47,62] entertainment[29,49,53,84] which includes tweets about music and movies, automotive or vehicles,[49,53] banking,[53] government or public campaign or public administration,[33,37,39,57,72,79] education or universities,[43,53,57,67] science or technology,[40,49,57] politics,[17,39,40,49,61,67,72,79,80,84] sports,[40,49,60] daily deals and discount,[56] trade or commercial services or business or financial services,[40,57] NGO's (charities),[70] blogger's or journalists (freelance media professionals (FMP),[70] taxation,[72] terrorism,[38,41] etc.

Thus, a number of datasets have been used across selected studies to conduct empirical evaluations of SC techniques for SA on Twitter. Amongst the benchmarked datasets, it was observed that SemEval datasets, especially SemEval 2007, 2013, 2014, and 2015 respectively were the most widely used ones. Next to follow was the data collected and prepared by the Stanford University and then comes the Sanders dataset. Apart from these most frequently used datasets, the other data sources that were also considerably explored are WePS-3, 2008 Presidential Debate Corpus, Sentiment140, RepLab 2012, RepLab 2013, STS-manual, Gold Standard personality labeled Twitter dataset, Cleveland Heart Disease data, STS-Gold, SS-Twitter, SE-Twitter, STS-Test. Table 5 exhibits the most widely used datasets for SA of Twitter using SC techniques.

Figure 7 depicts the distribution of the variety of benchmark datasets used by the researchers in the past decade for SA of Twitter using SC techniques.

Further, with the increased trend in the usage of Twitter it was observed that random tweets on general topics from various domains have been considered for research evaluations, especially since the year 2010. Figure 8 depicts the year-wise trend of published work with random tweets from various domains taken as dataset for empirical evaluation.

TABLE 5 Widely used datasets for Twitter SA

Rank	Dataset Used	# Papers	Referenced Study
1	SemEval	8	30,52,54-55,73,75-76,85
2	Data collected and prepared by Stanford University	4	34,45-46,75
2	Sanders	4	44,55,75,79
3	Presidential Debate Corpus	3	44,75,79
3	SS-Twitter	3	73,83,85
4	STS-Gold, STS-Test, SE-Twitter	2	73,83
4	HCR	2	75,79

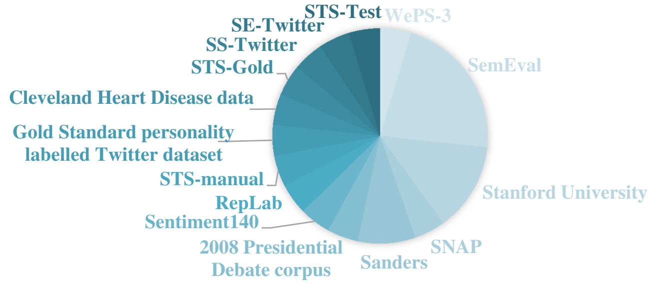


FIGURE 7 Distribution of different benchmark datasets

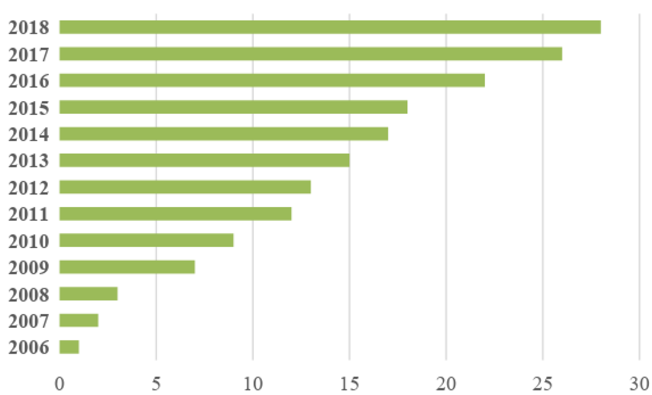


FIGURE 8 Year-wise cumulative assessment of random (general) tweets used as datasets

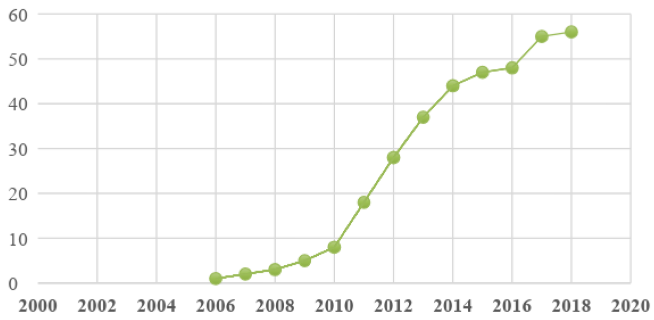


FIGURE 9 Year-wise cumulative assessment of topic oriented tweets used as datasets

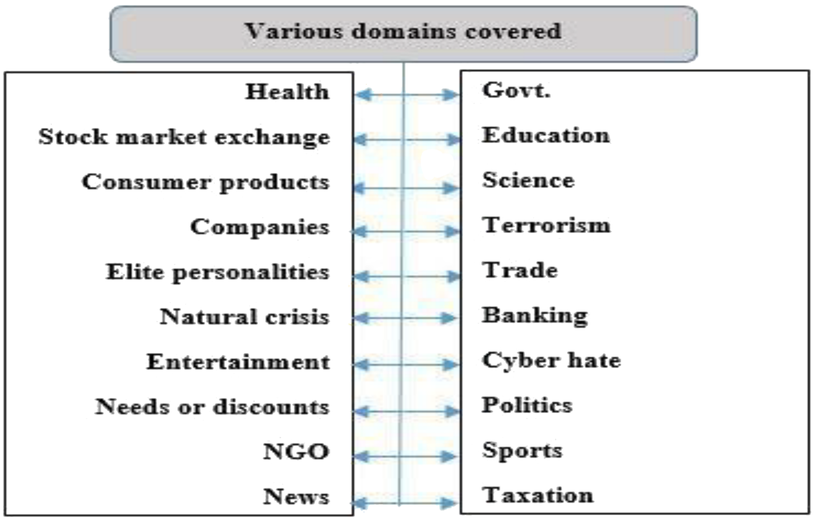


FIGURE 10 Various domains covered for research in the past decade for Twitter

Figure 9 shows the year-wise trend of published work with "topic specified/topic oriented" tweets from various domains taken as dataset for empirical evaluation.

Figure 10 specifies the wide array of topics from which the randomized tweets have been taken to finally evaluate studies applying SC techniques for SA.

Table 6 shows the mapping of the various SC techniques used for SA across domains in the last decade.

- *Most frequently used SC techniques for achieving efficient results for the SA on Twitter (RQ3)*

Table 7 illustrates the year-wise usage of various SC techniques in the past decade over a wide range of domain applying SA on Twitter.

Thus, the most frequently used SC techniques for achieving efficient results for the SA on Twitter is Machine Learning-Support Vector Machine (SVM), followed by Naïve Bayesian (NB) and ensemble methods (EM). The following Figure 11 shows the graph depicting the quantitative extent of use of various SC techniques over the past decade for SA on Twitter.

Figure 12 depicts the distribution of various categories of SC techniques over the past decade.

Table 8 depicts the implementation of hybrid techniques over the past decade.

- *Widely used performance measures (RQ4)*

Performance Measures

The following performance measures were observed in all selected studies to measure the performances of an applied SC technique. They are referred to as the "key performance indicators" (KPIs) or "performance parameters" or "efficacy measures." Table 9 illustrates the KPI's used over past decade.

TABLE 6 Mapping of the techniques with the varied domains covered in the last decade

TECHNIQUES APPLIED	Health	Stock Market Exchange	ConsumerProd	Companies	Elite Users	Environment/ Natural Crisis	Entertainment	Deals/ Discount	News	Govt.	Education	Science	Terrorism.	Trade/ Finance	Banking	Cyber Hate	Politics	Sports	Taxation	NGO	FMP	Automotive
EM	✓	✓		✓	✓				✓				✓		✓	✓		✓				
DT	✓	✓		✓	✓		✓		✓		✓		✓			✓	✓					✓
kNN		✓		✓	✓		✓		✓			✓					✓	✓				✓
SVM	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓		✓		✓	✓	✓	✓	✓	✓	✓
LR	✓	✓	✓	✓			✓	✓					✓									
LogR	✓			✓	✓				✓	✓	✓	✓		✓			✓	✓				
SGD			✓		✓	✓			✓													
MR		✓																				
NB	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓		✓			✓	✓				✓
FL	✓		✓					✓														
NN		✓	✓	✓	✓	✓	✓		✓								✓	✓				
GA	✓			✓													✓					
CNN			✓		✓	✓			✓													
HYBRID (NB-SVM)			✓		✓	✓			✓													
HYBRID (RFDT+BS)					✓					✓												
HYBRID (BLR)													✓									
HYBRID (SVM + RFDT)																✓						
HYBRID (SVM + BS)	✓		✓					✓														
HYBRID (SVM+ Bagging)	✓		✓					✓														

TABLE 7 Year-wise distribution of applicability of the techniques

Rank	Technique	Papers	Year	Referenced study
1	SVM	44	2012-2018	17,27-36,43-52,54,55, 57,58,62, 63,65-67,70,71,74,76-79,83-85
2	NB	29	2012-2018	17,27,30-32,36,38-40,43,47,49,51, 54-58,64,67,68,71,73,74,76-79, 80,82,85
3	EM	15	2014-2017	31,37,41,44,45,54,56,59,60,62,63, 66,73,76,81
4	LogR	14	2013-2014, 2016-2018	30,38,39,54,55,57-59,61,68,71, 73,74,85
5	DT	9	2013-2017	29,39,44,47,53,54,67,68,82
6	kNN	6	2012, 2014-2016	17,34,49,52,58,66
6	NN	6	2013, 2017-2018	28,52,60,63,76,84
7	CNN	4	2017-2018	76,83-85
8	LR	3	2013, 2014, 2016	29,41,56
9	FL	2	2014, 2016	42,56
10	MR	1	2016	63
11	GA	1	2017	75

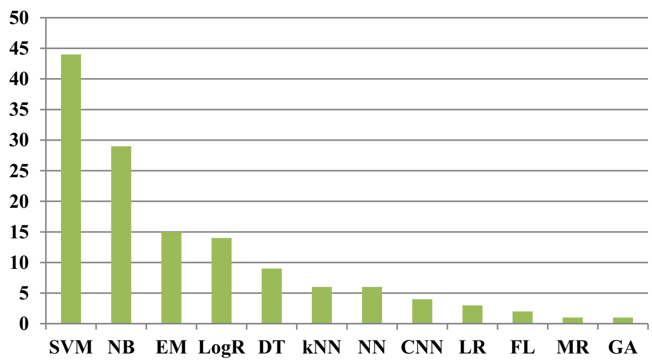


FIGURE 11 Quantitative extent of use of various SC techniques over past decade

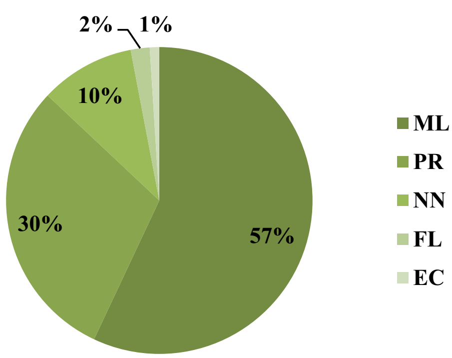


FIGURE 12 Distribution of various categories of SC techniques over the past decade (expressed in percentages)

TABLE 8 Year-wise distribution of applicability of the hybrid techniques

S:No.	Technique	Papers	Year	Referenced Study
1	RFDT+BS	1	2014	37
2	BLR	1	2014	38
3	SVM + RFDT	1	2015	47
4	SVM + BS, SVM + Bagging	1	2016	56
5	NBSVM	1	2017	76

TABLE 9 Key performance indicators

S. No.	KPI	Description	Referred study
1.	Precision (P)	It defines the exactness of any classifier. A higher precision value indicates fewer "false positives" (FP) and vice versa. It is given as the ratio of true positives (TP) to all the predicted positives.	27-29,31-33,35,37-39,43-45,47,49-51,56-62, 64,66,68,70,74,75,80,81,83,84
2.	Recall (R)/ Sensitivity (Sn)	It defines the sensitivity or the completeness of any classifier. A higher recall value indicates less "false positives" and vice versa. Recall and precision are bounded by inverse relation with each other. It is given as the ratio of TP to all the actual positives (TP + FN).	27-29,31-33,35,37-39,43-45,47,49-51,57-59, 61,62,64,66,68,70,74,75,81,83,84
3.	Accuracy (A)	It is defined as proximity of a measurement to its true value. It is calculated as a proportion of TP and true negatives (TN) among total inspected cases.	17,29,40,43-45,48-50,52,55,57,60,64, 66,68-70,72-75,77-79,85,81-83
4.	Confusion matrix (CM)	It is also called as error matrix and is used for evaluating performances of the supervised learning classifiers based on the actual and the predicted classifications on a set of test data for which the true values are known. Each cell has fields like false negatives (FN), FP, TN, TP.	36
5.	Average precision (AP)	It is defined as the mean of precision values for a system taken on different data sets for overall evaluation of the system performances.	56
6.	F score/F1 score/ F measure/F1 measure/F1 value(F)	It is defined as "weighted" harmonic mean of Recall and Precision. It is the combination of the precision and recall measures.	27-35,37-39,43,44,47,49-51,53,54,59,61,62, 66-68,70,74,75,81,83,84
7.	Macro average F score/ Macro F1 score (MAF)	This metric is used in order to understand the overall performance of the system across varied data sets. It is defined as the harmonic mean of the macro average precision and macro average recall.	43,46,52,68,76
8.	AUC (area under the curve)	It is interpreted as AUROC (area under the receiver operating characteristics curve) and is also used for measuring the efficacy of the classifiers based on two metrics, ie, true positive rate and false positive rate.	57
9.	Cohen's Kappa (CK)	This metric performs a comparison between the observed and the expected accuracy. It is used for evaluating performance of a single classifier and also among different classifiers.	29,43,48
10.	Error rate (ER)	It is majorly measured on testing data sets. Error rate determines that the predicted output values are inaccurate, ie, it is defined as a proportion of total number of bad predictions among all total inspected cases.	40,41
11.	Macro average error rate (MAER)	It is used for measuring the efficacy of the overall system by calculating the mean of per label classification error rate.	46
12.	Specificity (Sp)	Sp is referred as true negative rate and is denoted as ratio of TN to all the actual negatives (TN + FP).	64
13.	GM	It measures the ability of a classifier to balance Sp (accuracy on the negative examples) and Sn (accuracy on positive examples).	71
14.	NMAE	Mean absolute error (MAE) computes the average of the absolute difference between the predicted and the true ratings. NMAE normalizes MAE by dividing it with difference between the maximum and the minimum rating items.	63
15.	Su & Cf	Su is defined as a probability that a randomly chosen item set will be present in all the items in a rule. Cf is defined as conditional probability that all the items will be present, given the presence of one of the items.	42
16.	MiAF	Micro average measures are used in case of varied size datasets. It is defined as the harmonic mean of micro average precision and micro average recall.	43
17.	FAR, FRR, EER	FAR is defined as the ratio of the number of the falsely accepted items/client patterns to the number of all imposter patterns. FRR is denoted as ratio of the rejected items/client patterns to the total number of the items/client patterns. When FAR and FRR are equal, that common value is called as EER.	65

- *Trend and impact of using SC techniques for SA on Twitter in the past decade (RQ5)*

There has been a notable upward trend in the past decade, evident from studies within the domain with an increase in interest of both researchers and practitioners. Figure 13 is a line graph which plots the year of publication on X-axis and the number of papers published in that corresponding year on the Y-axis. All the papers referred in this SLR were journal publications, and from the Figure, we can make out that the maximum numbers of papers were published in the year 2016. That is, 19 papers were published in 2016. Since the inception of Twitter, the popularity and interest of using SA on it is discernable with the increased and focus research implementing variety of SC techniques and within diverse domains like movies, music, sports, news, health, stock exchange, etc. Figures 13 and 14 show the year-wise distribution of number of papers and a marked increase in distribution of papers post 2014.

Figure 13 is a clear indication that the popularity of SA for Twitter using SC techniques has increased enormously after year 2014, hence the future researchers and practitioners should inspect the papers being published after year 2014 so to reach to the most recent and significant research articles for the latest developments in this field. Table 2 is showing the insights of application of SC techniques for SA on Twitter for the past decade. SC is a broad term consisting of various techniques. One such sub-category of SC is called as ML that further comprises of DL methodologies. Very few studies exists that demonstrates the usage of DL for SA on Twitter in the past decade. These techniques have the power of self-learning and enables multi-level automatic feature representation learning. They outperformed the state of art algorithms due to the capability to harness large amount of data in a time and space efficient manner, yielding improved performance. Thus, we can say that the application of DL to SA is an upcoming research area that could be used for discovering the hidden emotions, sentiments and opinions for uncovering the appropriateness of the feelings behind these textual tweet posts. Furthermore, other parallel areas of application of SA using DL on social media such as Twitter includes assessing context-aware SA using DL, sarcasm detection using DL (as sarcasm is one of the type of negative polarity based sentiment), humor detection using DL (where humor again is considered to be a positive based sentiment), etc. In a nut shell, we can say that the SA associated with the micro-blogging social media like Twitter can be a magnificent gamut of information for efficient decision making. It provides relevant insights that can aid in determining effective social well-being for varied marketing strategy, improving product messaging, customer servicing, brand monitoring, etc.

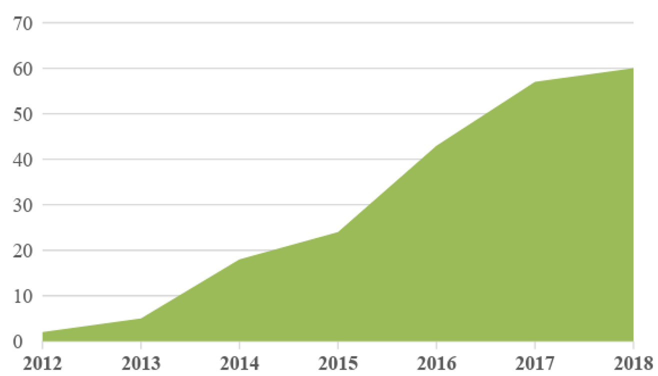


FIGURE 13 Year-wise distribution of number of papers

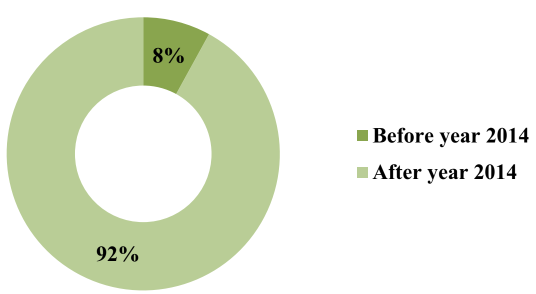


FIGURE 14 Distribution of papers after year 2014 (number of papers/percentage of total papers)

6 | CONCLUSION

From the systematic literature review conducted, it was observed that there has been a recent trend in research studies as well as reports of novel implementation models which improve the sentiment analysis process on social media. The primary goals have been to assist decision making, augment intelligent analytics and enhance personalized web experience. SC has emerged as a paradigm which combines new computational techniques that mimic consciousness and cognition in several important respects. Application of SC techniques for sentiment classification on social media is a promising direction of research with practical domains for finding, exploring and understanding the extensibility of human expressions via both textual and non-textual unstructured web-data available. The unique property of all SC techniques is their power of self-tuning, that is, they derive the power of generalization from approximating and learning from experimental data that is usually performed in a high-dimensional space, such as Twitter, that has emerge as the true source in order to testify the reasoning and searching capabilities of these techniques for improved precision and certainty. Consequently studies to understand the theory, research and practice of using SC techniques for this supervised learning model of SA on Twitter to explore its feasibility and scope have been conducted and the following research gaps were identified across pertinent literature:

- Mining, analyzing and classifying sentiments from web data is challenging as effective feature selection is computationally hard.
- Incessant need to enhance the performance of the sentiment classification tools which are now in practical use within various business and social domains.
- The tools and software are useable and affordable only by organizations (both private and government) but currently unavailable to generic users for assisting intelligent and personalized data analysis.
- Though researchers are keen in applying soft computing techniques, only few approaches have been explored. Techniques such as Deep learning, Evolutionary computing, optimization algorithms and hybrid approaches including Neuro-Fuzzy models have been least explored or implemented to substantiate their influence on sentiment analysis.
- Analyzing sentiment analysis using Fuzzy Logic approaches like Type 2 Fuzzy models is yet another novel dimension that is open for further exploration.
- Also, Nature Inspired Algorithms (NIA) including Swarm Intelligence algorithms and Bio-Inspired algorithms like flower pollination, grey wolf, moth flame, etc, have not been implemented yet in order to signify their impact on sentiment analysis when examined for varied social media.
- The existing models using soft computing techniques for sentiment analysis on social media have majorly considered Twitter as the database (Benchmark SemEval, random tweets, etc), making other social media technologies such as Flickr, Tumblr open to application and testing.
- Most of the work in the field of SA majorly ponders on polarity detection and classification. The reported studies focused on the use of sentiment words such as nouns, adverbs, verbs, adjectives that often decrease the precision rate of SA. Thus, an optimum model is needed that could yield adequate precision rate for effective SA.
- Most of the work done to analyze the sentiments is on textual un-structured web data, whereas multimedia (non-textual: audio, video, image, gif's, emoticons) sentiment analysis has not been explored much.
- Fine-grain sentiment analysis which includes, emotion analysis, sarcasm detection, rumor detection, irony detection have been identified as potential directions of research.
- Social media has become an informal way of communication with increased use of slangs and emoticons, mal-formed words, colloquial expressions, thus increasing the dimensionality, fuzziness and complexity of the content.
- Analyzing such man-made words and expressions for such informal and mashed-up web content has become altogether a very challenging task in the area of sentiment analysis.
- Hence, a multi-prolonged approach utilizing the self-tuning capabilities of SC techniques is required that could aid in meticulous prediction of sentiment polarity.

Thus, the need to exploit new computational techniques for improving the sentiment classification accuracy on social Web is identified, making this domain of study a potentially active and dynamic for both researchers and practitioners.

ORCID

Akshi Kumar https://orcid.org/0000-0003-4263-7168
Arunima Jaiswal https://orcid.org/0000-0002-2376-9416

REFERENCES

1. International Data Corporation. 2017. https://www.idc.com/about/about.jsp. Accessed April 22, 2017.
2. Aggarwal CC, Zhai C. A survey of text classification algorithms. In: *Mining Text Data*. Boston, MA: Springer Science+Business Media; 2012.
3. Rajman M, Besançon R. Text mining-knowledge extraction from unstructured textual data. *Advances in Data Science and Classification: Proceedings of the 6th Conference of the International Federation of Classification Societies (IFCS-98) Università "La Sapienza", Rome, 21-24 July, 1998*. Berlin, Germany: Springer-Verlag Berlin Heidelberg; 1998:473-480.
4. Kumar A, Sebastian TM. Sentiment analysis on Twitter. *Int J Comput Sci Issues*. 2012;9(4):372-378.

5. Pang B, Lee L. Opinion mining and sentiment analysis. *Found Trends® Inf Retr*. 2008;2(1-2):1-135.
6. Kumar A, Sebastian TM. Sentiment analysis: a perspective on its past, present and future. *Int J Intell Syst Appl*. 2012;4(10):1-14.
7. Reilly T. Web 2.0 Compact Definition: Trying Again. Sebastopol, CA: O'Reilly Media; 2017. http://radar.oreilly.com/2006/12/web-20-compact-definition-tryi.html. Accessed April 24, 2017.
8. Alfouzan HI. Introduction to SMAC-social mobile analytics and cloud. *Int J Sci Eng Res*. 2015;6:128-130.
9. Li D, Luo Z, Ding Y, et al. User-level microblogging recommendation incorporating social influence. *J Assoc Inf Sci Technol*. 2017;68(3):553-568.
10. Pak A, Paroubek P. Twitter as a corpus for sentiment analysis and opinion mining. In: Proceedings of the Seventh Conference on International Language Resources and Evaluation; 2010; Valletta, Malta.
11. Liu B. *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. Cambridge, UK: Cambridge University Press; 2015.
12. Kumar A, Joshi A. Ontology driven sentiment analysis on social web for government intelligence. In: Proceedings of the Special Collection on eGovernment Innovations in India; 2017; New Delhi, India.
13. Dave K, Lawrence S, Pennock DM. Mining the peanut gallery: opinion extraction and semantic classification of product reviews. In: Proceedings of the 12th international conference on World Wide Web; 2003; Budapest, Hungary.
14. Kumar A, Khorwal R, Chaudhary S. A survey on sentiment analysis using swarm intelligence. *Indian J Sci Technol*. 2016;9(39):1-7.
15. Aggarwal CC. *Neural Networks and Deep Learning: A Textbook*. Cham, Switzerland: Springer International Publishing; 2018.
16. Sivanandam SN, Deepa SN. *Principles of Soft Computing (With CD)*. Hoboken, NJ: *Wiley*. 2007.
17. Finn S, Mustafaraj E. Learning to discover political activism in the Twitterverse. *KI-Künstliche Intelligenz*. 2013;27(1):17-24.
18. Saif H, He Y, Fernandez M, Alani H. Contextual semantics for sentiment analysis of Twitter. *Inf Process Manag*. 2015;52(1):5-19. https://doi.org/10.1016/j.ipm.2015.01.005
19. Wu B, Shen H. Analyzing and predicting news popularity on Twitter. *Int J Inf Manag*. 2015;35(6):702-711.
20. Balas VE, Fodor J, Várkonyi-Kóczy AR, eds. *New Concepts and Applications in Soft Computing*. Berlin, Germany: Springer-Verlag Berlin Heidelberg; 2013.
21. Tsytsarau M, Palpanas T. Survey on mining subjective data on the web. *Data Min Knowl Discov*. 2012;24(3):478-514.
22. Cambria E, Schuller B, Xia Y, Havasi C. New avenues in opinion mining and sentiment analysis. *IEEE Intell Syst*. 2013;28(2):15-21.
23. Feldman R. Techniques and applications for sentiment analysis. *Commun ACM*. 2013;56(4):82-89.
24. Montoyo A, Martínez-Barco P, Balahur A. Subjectivity and sentiment analysis: an overview of the current state of the area and envisaged developments. *Decis Support Syst*. 2012;53(4):675-679.
25. Medhat W, Hassan A, Korashy H. Sentiment analysis algorithms and applications: a survey. *Ain Shams Eng J*. 2014;5(4):1093-1113.
26. Kitchenham B, Charters S. *Guidelines for Performing Systematic Literature Reviews in Software Engineering*. Technical Report EBSE 2007-001. Keele, UK and Durham, UK: Keele University and Durham University; 2007.
27. Yerva SR, Miklós Z, Aberer K. Quality-aware similarity assessment for entity matching in Web data. *Inf Syst*. 2012;37(4):336-351.26(28.)
28. Lou T, Tang J, Hopcroft J, Fang Z, Ding X. Learning to predict reciprocity and triadic closure in social networks. *ACM Trans Knowl Discov Data*. 2013;7(2):1-25.
29. Arias M, Arratia A, Xuriguera R. Forecasting with Twitter data. *ACM Trans Intell Syst Technol*. 2013;5(1).
30. Trilla A, Alias F. Sentence-based sentiment analysis for expressive text-to-speech. *IEEE Trans Audio Speech Lang Process*. 2013;21(2):223-233.
31. Tuarob S, Tucker CS, Salathe M, Ram N. An ensemble heterogeneous classification methodology for discovering health-related knowledge in social media messages. *J Biomed Inform*. 2014;49:255-268.
32. Morchid M, Dufour R, Bousquet P-M, Linarès G, Torres-Moreno J-M. Feature selection using principal component analysis for massive retweet detection. *Pattern Recognit Lett*. 2014;49:33-39.
33. Montejo-Ráez A, Martínez-Cámara E, Martín-Valdivia MT, Ureña-López LA. Ranked WordNet graph for sentiment polarity classification in Twitter. *Comput Speech Lang*. 2014;28(1):93-107.
34. Smailović J, Grčar M, Lavrač N, Žnidaršič M. Stream-based active learning for sentiment analysis in the financial domain. *Inf Sci*. 2014;285:181-203.
35. Boella G, Di Caro L, Ruggeri A, Robaldo L. Learning from syntax generalizations for automatic semantic annotation. *J Intell Inf Syst*. 2014;43(2):231-246.
36. Brynielsson J, Johansson F, Jonsson C, Westling A. Emotion classification of social media posts for estimating people's reactions to communicated alert messages during crises. *Secur Inform*. 2014;3:1-11.
37. Arakawa Y, Kameda A, Aizawa A, Suzuki T. Adding Twitter-specific features to stylistic features for classifying tweets by user type and number of retweets. *J Assoc Inf Sci Technol*. 2014;65(7):1416-1423.
38. Burnap P, Williams ML, Sloan L, et al. Tweeting the terror: modelling the social media reaction to the Woolwich terrorist attack. *Soc Netw Anal Min*. 2014;4:1-14.
39. Makazhanov A, Rafiei D, Waqar M. Predicting political preference of Twitter users. *Soc Netw Anal Min*. 2014;4:1-15.
40. Bogdanov P, Busch M, Moehlis J, Singh AK, Szymanski BK. Modeling individual topic-specific behavior and influence backbone networks in social media. *Soc Netw Anal Min*. 2014;4:1-16.
41. Lin YR, Margolin D. The ripple of fear, sympathy and solidarity during the Boston bombings. *EPJ Data Sci*. 2014;3.
42. Fu X, Shen Y. Study of collective user behaviour in Twitter: a fuzzy approach. *Neural Comput Appl*. 2014;25(7-8):1603-1614.
43. Chen X, Vorvoreanu M, Madhavan K. Mining social media data for understanding students' learning experiences. *IEEE Trans Learn Technol*. 2014;7(3):246-259.
44. Liu S, Cheng X, Li F, Li F. TASC: topic-adaptive sentiment classification on dynamic tweets. *IEEE Trans Knowl Data Eng*. 2015;27(6):1696-1709.
45. Kranjc J, Smailović J, Podpečan V, Grčar M, Žnidaršič M, Lavrač N. Active learning for sentiment analysis on data streams: methodology and workflow implementation in the ClowdFlows platform. *Inf Process Manag*. 2015;51(2):187-203.
46. Sluban B, Smailović J, Battiston S, Mozetič I. Sentiment leaning of influential communities in social networks. *Comput Soc Netw*. 2015;2:1-21.
47. Burnap P, Williams ML. Cyber hate speech on Twitter: an application of machine classification and statistical modeling for policy and decision making. *Policy Internet*. 2015;7(2):223-242.

48. Zubiaga A, Spina D, Martinez R, Fresno V. Real-time classification of Twitter trends. *J Assoc Inf Sci Technol*. 2015;66(3):462-473.
49. Magdy W, Sajjad H, El-Ganainy T, Sebastiani F. Bridging social media via distant supervision. *Soc Netw Anal Min*. 2015;5:1-12.
50. Tsytsarau M, Palpanas T. Managing diverse sentiments at large scale. *IEEE Trans Knowl Data Eng*. 2016;28(11):3028-3040.
51. Andriotis P, Oikonomou G, Tryfonas T, Li S. Highlighting relationships of a smartphone's social ecosystem in potentially large investigations. *IEEE Trans Cybern*. 2016;46(9):1974-1985.
52. Tang D, Wei F, Qin B, Yang N, Liu T, Zhou M. Sentiment embeddings with applications to sentiment analysis. *IEEE Trans Knowl Data Eng*. 2016;28(2):496-509.
53. Peetz MH, de Rijke M, Kaptein R. Estimating reputation polarity on microblog posts. *Inf Process Manag*. 2016;52(2):193-216.
54. Sulis E, Farías DIH, Rosso P, Patti V, Ruffo G. Figurative messages and affect in Twitter: differences between #irony, #sarcasm and #not. *Knowl Based Syst*. 2016;108:132-143.
55. Wu F, Song Y, Huang Y. Microblog sentiment classification with heterogeneous sentiment knowledge. *Inf Sci*. 2016;373:149-164.
56. Lo SL, Chiong R, Cornforth D. Ranking of high-value social audiences on Twitter. *Decis Support Syst*. 2016;85:34-48.
57. van Zoonen W, Toni GLA. Social media research: the application of supervised machine learning in organizational communication research. *Comput Hum Behav*. 2016;63:132-141.
58. Wang Z, Cui X, Gao L, Yin Q, Ke L, Zhang S. A hybrid model of sentimental entity recognition on mobile social media. *EURASIP J Wirel Commun Netw*. 2016;1:1-12.
59. Celli F, Ghosh A, Alam F, Riccardi G. In the mood for sharing contents: emotions, personality and interaction styles in the diffusion of news. *Inf Process Manag*. 2016;52(1):93-98.
60. Igawa RA, Barbon Jr S, Paulo KCS, et al. Account classification in online social networks with LBCA and wavelets. *Inf Sci*. 2016;332:72-83.
61. Korkmaz G, Cadena J, Kuhlman CJ, Marathe A, Vullikanti A, Ramakrishnan N. Multi-source models for civil unrest forecasting. *Soc Netw Anal Min*. 2016;6:1-25.
62. Burnap P, Williams ML. Us and them: identifying cyber hate on Twitter across multiple protected characteristics. *EPJ Data Sci*. 2016;5:1-15.
63. Oliveira N, Cortez P, Areal N. The impact of microblogging data for stock market prediction: using Twitter to predict returns, volatility, trading volume and survey sentiment indices. *Expert Syst Appl*. 2017;73:125-144.
64. Perikos I, Hatzilygeroudis I. Recognizing emotions in text using ensemble of classifiers. *Eng Appl Artif Intell*. 2016;51:191-201.
65. Brocardo ML, Traore I, Woungang I, Obaidat MS. Authorship verification using deep belief network systems. *Int J Commun Syst*. 2017;30(12):1-10.
66. Bouazizi M, Ohtsuki TO. A pattern-based approach for sarcasm detection on Twitter. *IEEE Access*. 2016;4:5477-5488.
67. Farías DIH, Patti V, Rosso P. Irony detection in Twitter: the role of affective content. *ACM Trans Internet Technol*. 2016;16(3):1-24.
68. Sintsova V, Pu P. Dystemo: distant supervision method for multi-category emotion recognition in tweets. *ACM Trans Intell Syst Technol*. 2016;8(1):1-22.
69. Nair LR, Shetty SD, Shetty SD. Applying spark based machine learning model on streaming big data for health status prediction. *Comput Electr Eng*. 2018;65:393-399.
70. Cui L, Zhang X, Qin AK, Sellis T, Wu L. CDS: collaborative distant supervision for Twitter account classification. *Expert Syst Appl*. 2017;83:94-103.
71. Pérez-Gállego P, Quevedo JR, del Coz JJ. Using ensembles for problems with characterizable changes in data distribution: a case study on quantification. *Inf Fusion*. 2017;34:87-100.
72. Alsinet T, Argelich J, Béjar R, Fernández C, Mateu C, Planes J. Weighted argumentation for analysis of discussions in Twitter. *Int J Approx Reason*. 2017;85:21-35.
73. Jianqiang Z, Xiaolin G. Comparison research on text pre-processing methods on Twitter sentiment analysis. *IEEE Access*. 2017;5:2870-2879.
74. Jain VK, Kumar S. Effective surveillance and predictive mapping of mosquito-borne diseases using social media. *J Comput Sci*. 2018;25:406-415.
75. Keshavarz H, Abadeh MS. ALGA: adaptive lexicon learning using genetic algorithm for sentiment analysis of microblogs. *Knowl Based Syst*. 2017;122:1-16.
76. Xiong S, Lv H, Zhao W, Ji D. Towards Twitter sentiment classification by multi-level sentiment-enriched word embeddings. *Neurocomputing*. 2018;275:2459-2466.
77. Neppalli VK, Caragea C, Squicciarini A, Tapia A, Stehle S. Sentiment analysis during Hurricane Sandy in emergency response. *Int J Disaster Risk Reduct*. 2018;21:213-222.
78. Singh A, Shukla N, Mishra N. Social media data analytics to improve supply chain management in food industries. *Transp Res Part E Logist Transp Rev*. 2018;114:398-415.
79. Xiaomei Z, Jing Y, Jianpei Z, Hongyu H. Microblog sentiment analysis with weak dependency connections. *Knowl Based Syst*. 2018;142:170-180.
80. Khan I, Naqvi SK, Alam M, Rizvi SNA. An efficient framework for real-time tweet classification. *Int J Inf Technol*. 2017;9(2):215-221.
81. Bouazizi M, Ohtsuki T. A pattern-based approach for multi-class sentiment analysis in Twitter. *IEEE Access*. 2017;5:20617-20639.
82. Li B, Chan KC, Ou C, Ruifeng S. Discovering public sentiment in social media for predicting stock movement of publicly listed companies. *Inf Syst*. 2017;69:81-92.
83. Jianqiang Z, Xiaolin G, Xuejun Z. Deep convolution neural networks for Twitter sentiment analysis. *IEEE Access*. 2018;6:23253-23260.
84. Ghiassi M, Lee S. A domain transferable lexicon set for Twitter sentiment analysis using a supervised machine learning approach. *Expert Syst Appl*. 2018;106:197-216.
85. Symeonidis S, Effrosynidis D, Arampatzis A. A comparative evaluation of pre-processing techniques and their interactions for Twitter sentiment analysis. *Expert Syst Appl*. 2018;110:298-310.

How to cite this article: Kumar A, Jaiswal A. Systematic literature review of sentiment analysis on Twitter using soft computing techniques. *Concurrency Computat Pract Exper*. 2020;32:e5107. https://doi.org/10.1002/cpe.5107