Computer Science Review 40 (2021) 100395

Contents lists available at ScienceDirect

Computer Science Review

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

Review article

Machine learning algorithms for social media analysis: A survey

Balaji T.K. [a], Chandra Sekhara Rao Annavarapu [b,*], Annushree Bablani [a]

[a] Department of Computer Science and Engineering, Indian Institute of Information Technology, Sri City, Andhra Pradesh, 517646, India
[b] Department of Computer Science and Engineering, Indian Institute of Technology (Indian School of Mines), Dhanbad, Jharkhand, 826004, India

ARTICLE INFO

Article history:
Received 13 July 2020
Received in revised form 2 March 2021
Accepted 4 March 2021
Available online 20 March 2021

Keywords:
Social Media
Machine learning
Social network analysis
Applications of social media analysis

ABSTRACT

Social Media (SM) are the most widespread and rapid data generation applications on the Internet increase the study of these data. However, the efficient processing of such massive data is challenging, so we require a system that learns from these data, like machine learning. Machine learning methods make the systems to learn itself. Many papers are published on SM using machine learning approaches over the past few decades. In this paper, we provide a comprehensive survey of multiple applications of SM analysis using robust machine learning algorithms. Initially, we discuss a summary of machine learning algorithms, which are used in SM analysis. After that, we provide a detailed survey of machine learning approaches to SM analysis. Furthermore, we summarize the challenges and benefits of Machine Learning usages in SM analysis. Finally, we presented open issues and consequences in SM analysis for further research.

Contents

* Corresponding author.
E-mail addresses: balajitk7@gmail.com (Balaji T.K.), acsrao@iitism.ac.in (C.S.R. Annavarapu), annushree.bablani@iiits.in (A. Bablani).

https://doi.org/10.1016/j.cosrev.2021.100395
1574-0137/

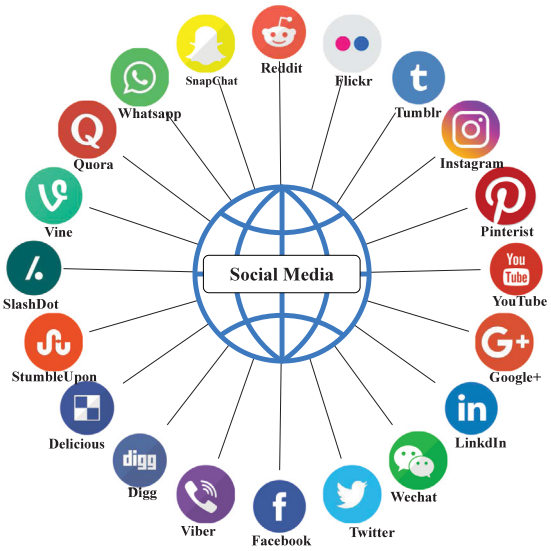

Fig. 1. Common Social Media platforms.

1. Introduction

Social Media (SM) are often observed to be quite large, as it is embedded in many sources of data and at different scales. SM plays a vital role in human communication and improves business applications [1]. In Fig. 1, we presented popular social media applications, which are getting more attention from users in the past few decades. The data generating among these SM applications are mammoth, from tweets to swipes, likes to shares, the online world is exploding as Terabytes to Petabytes [2,3]. The number of online users has increased by over a billion in the last few years. More than half of the globe's web traffic increased because of smartphones as they are pretty flexible to access SMs. As per the statistics, the data is on overdrive. It is being generated at a break-neck pace, flooding out of the dozens of associated devices that we use regularly, and it gives no signs of slowing down.

In this dynamic world, text is the biggest way of communication and more than 18.2 million text messages are transmitting in a minute. Over the text, multimedia messages are also grasping the attention in recent days, among the jpg, png, tiff, and videos, the gif images posting are increased 22% over the year. YouTube, Amazon, Netflix, and Facebook are making users use high network bandwidth. These high rates of usage of streaming services makes unstoppable climbing of data [4]. The youth in the world demands internet data. More than 18 million requests got to The Weather Channel forecasting requests online every minute to determining the climate, shine, or rain. It is a 22% hike as compared with the previous year per minute requests. Netflix users around the world are watching 69,444 h of videos per minute.[1]

These comprehensive and streaming data is not a gentle task to store, process, and extract knowledge. It requires higher storage devices and efficient processing mechanisms. A possible way to gain this knowledge is by merging the intelligence concepts with smart learning approaches through Machine Learning (ML) techniques [5–7]. ML gives, the systems able to learn itself without being programmed from outside. ML has lots of intra-disciplinary and interdisciplinary domain scope such as medical sciences, weather forecasting [8,9], image analysis [10–12], opinion mining, and textual forensics, fake news identification [13], wireless sensor networks [14], Internet of Things (IoT), and many [15]. For all these applications, we use various ML algorithms, summary of most common machine learning algorithms are discussed in [14]. There are several advantages and disadvantages of ML in these domains. However, ML disadvantages can be mastered by their outstanding advantages. In this context, SM also adopts the ML to get the benefits from it. The general processing steps of SM through machine learning are shown in Fig. 2.

Currently, researchers are working on SM analysis and presented their surveys; however, all their papers limited to a specific application of SM analysis. For example, in [16], surveyed only on disaster management through SM. In [17], authors studied sentiment analysis approaches, methods, and applications only. In [18], the authors provided a survey report on the recommender systems. In [19], surveyed social media analytics in hospitality and tourism. In [20], authors surveyed on fake profile detection techniques in SM. In [21], authors surveyed how to utilize SM data for pharmacovigilance. In [22], the authors presented a survey on the then software tools available for analyzing social media. Their article includes social media analytics techniques such as sentiment classification, supervised learning methods, social media analytics tools, and platforms. In [23], authors presented a review on recent trends and challenges to store, and process the social media big data.

Hence, Most of the existing surveys limited to specific SM application. Best of our knowledge, it is the first ever feat identifying and giving a comprehensive survey on multiple applications to Machine learning algorithms for social media analysis. We survey recent articles on various applications of SM analysis published up to 2020. This survey mainly focus on various applications such as anomaly detection, behavioral analysis, bioinformatics, business intelligence, crime detection, epidemics, event detection, image analysis, recommendations, relationships and reputations, and sentiment analysis. Hence this paper addresses multi-application knowledge of SM analysis that adopting ML techniques. The main contribution of our work can be summarized as follows:

[1] https://www.domo.com/learn/data-never-sleeps-7.

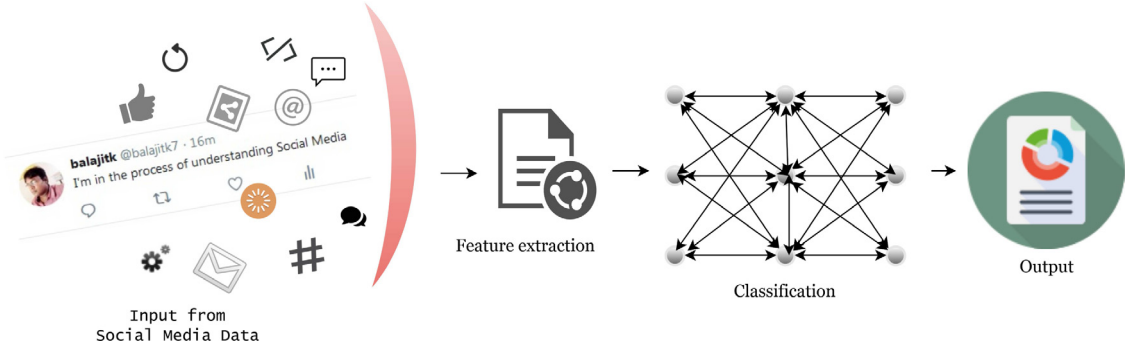

Fig. 2. General data processing of social media through Machine Learning.

1. We provide the summaries of various machine learning approaches which are used in SM.
2. We present the survey of machine learning in SM analysis.
3. The statistical analysis of SM over the ML are presented along with their limitations.
4. We Provide open issues identified in SM.

The rest of this paper is structured as follows. Abbreviations used in this paper are listed in Table 1. Section 2 discusses the approaches, and regularly used algorithms of ML. Section 3 explains the applications of SM analysis through ML. Section 4 explains the consequence and statistical analysis. Section 5 explains the open issues for further research. Section 6 concludes the paper.

2. Approaches, and algorithms of ML

In this section, we explore various machine learning approaches, commonly used algorithms and classifiers for SM analysis, which helps for a better understanding of the further sections.

2.1. Common machine learning approaches and algorithms

In general, we have three root learning approaches available in ML, such as supervised learning (SL), unsupervised learning (UL), and reinforcement learning (RL). The detailed process of SM raw data with various learning techniques are shown in Fig. 3. *Supervised learning* is helpful in cases where a label property is available for an individual dataset. The Decision Tree, Regression, Random Forest, Logistic Regression, KNN, etc are the example SL algorithms [24]. The *UL* is beneficial in cases where the challenge is to identify implicit relationships in a given unlabeled dataset. The examples of UL are clustering algorithms such as K-means and Apriori algorithms. The *RL* falls between supervised and unsupervised machine learning extremes, in this a form of feedback open for each predictive step or action, but no specific label or error message. In RL, actual input or output combinations are nevermore awarded, nor sub-optimal operations explicitly altered. Markov Decision Process is an algorithm that respects the Reinforcement algorithm. Additional approaches belong to the ML are Association rule learning, Inductive logic programming (ILP), Sparse dictionary learning, Representation learning, genetic algorithms, and Similarity and metric learning.

The *Association rule learning* is a rule-based ML method for identifying awesome relations between variables in huge databases. It is specified to recognize strong rules identified in databases using some measure of interestingness. For example, the rule {milk,flour} $\geq$ {bread} found in the supermarket sales data. This rule delivers as if a customer buys milk and flour together, and they also likely to buy cheese — this kind of information used as the basis for decisions about marketing

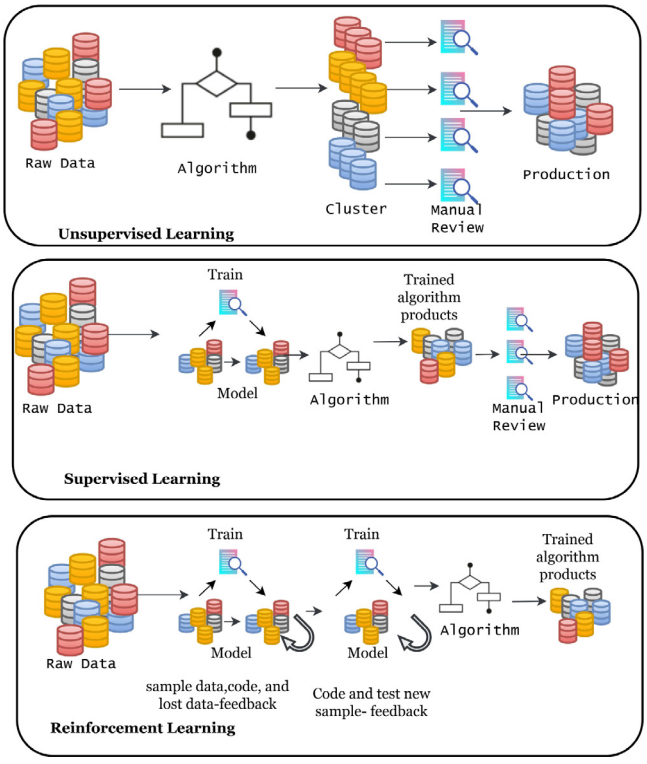

Fig. 3. Traditional data processing in various Learning Techniques of ML.

activities. There are multiple association rule-based algorithms — Apriori algorithm, Eclat algorithm, and FP-growth are notable in it [25]. The *Inductive Logic Programming (ILP)* rule learning is an approach using logic programming as a natural description for input examples, training information, and opinions [26]. *Sparse dictionary learning* (SDL) is very famous for recognition of patterns, in machine learning, vision analysis in computers, medical image, and so on. The sparse dictionary allows state of the art in pattern recognition [27] and computer visions.

The *Representation learning* is a collection of techniques that permit the model to automatically identify the methods required for feature selection or classification from traditional data. It substitutes manual feature engineering and enables a machine to both learn the features and utilize them to run a specific task [28]. The *Genetic algorithms* is a technique for solving both constrained and unconstrained optimization problems that are based on real selection, the method that encourages biological evolution [5]. The *Similarity and metric learning* used in Recommendation systems for recommending new similar objects that are learned from the similarity prediction system [29].

Table 1
Abbreviations used in this article.

Abbreviation	full form
AA	Adamic–Adar Predictor
AI	Artificial Intelligence
ANN	Artificial Neural Networks
ARX	Auto-Regression with Exogenous Inputs
AUC	Area Under the Curve
BCI	Brain–machine interfaces
BCR	Boosted Logistic Regression
CART	Classification and Regression Tree
CDC	Centers for Disease Control and Prevention
CESNA	Community detection from Node attributes and Links
CNN	Convolutional Neural Networks
COSMOS	Cardiff Online Social Media Observatory
CRF	Conditional Random Field
CRW	Cumulative Random Walk
DAG	Direct Acyclic Graph
DASS	Dengue Active Surveillance System
DCNN	Deep Convolutional Neural Networks
DL	Deep Learning
DT	Decision Tree
EMR	Electronic Medical Records
FN	False Negatives
FP	False Positives
GIS	Geographical Information System
GMG	Generalized Markov Graph
IDSC	Infection Disease Surveillance Center
IFM	Information Flow Model
ILP	Inductive Logic Programming
JST	Joint sentiment-topic
K-NN / KNN	K-Nearest Neighbors
LAIM	Latent Aspect Influence Model
LBM	Latent Bias Model
LDA	Latent Dirichlet Allocation
Lir	Linear Regression
LR	Logistic Regression
LRCV	Logistic Regression with built-in Cross-Validation
LT	Linear Threshold
MAP	Mean Average Precision
MDS	Multidimensional Scaling
MEMM	Maximum Entropy Markov Model
MGT	Mixed graph of terms
ML	Machine Learning
MLDS	Money Laundering Detection System
MRF	Markov Random Field
NDGC	Normalized Discounted Cumulative Gain
NEL	Named Entity Link
NeLP	Negative Link Prediction
NER	Named Entity Recognition
NLP	Natural Language Processing
NN	Neural Network
OAIM	Observed Aspect Influence Mode
OCCRF	One-Class Conditional Random Fields
RA	Resource Allocation Predictor
RF	Random Forest
RGDBN	Relational Generative Deep Belief Nets
RL	Reinforcement Learning
SBU	Stony Brook University
SCADA	Supervisory Control and Data Acquisition
SHM	Structural Health Monitoring
SL	Supervised Learning
SM	Social Media
SRL	Statistical Relational Learning
SRW	Supervised Random Walk
SVM	Support Vector Machine
TP	True Positives
UGC	User Generated Content
UL	Unsupervised Learning

The taxonomy of machine learning algorithms are shown in Fig. 4. Here we discuss each of the ML approaches used in the further section for better understanding.

2.1.1. Artificial neural networks and deep learning

An artificial neural network (ANN) learning algorithm is motivated by the composition and working aspects of organic neural

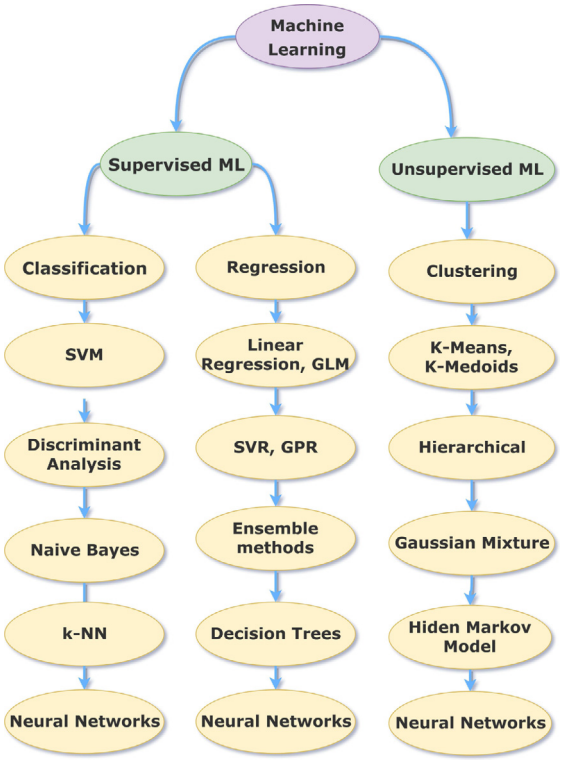

Fig. 4. Taxonomy of ML algorithms.

circuitry [30]. Computations are figured concerning a correlated assortment of unnatural neurons, processing knowledge using a connectionist approach to computation. Similarly, deep learning (DL) consists of multiple hidden layers when compared with an artificial neural network [31]. Deep learning approach strives to model the way the human brain processes light and sound into vision and hearing. Remarkable applications of deep learning are computer vision, image recognition, visual art processing, Natural Language Processing, drug discovery, bioinformatics, cyber security [32], emotion detection and speech recognition. The benefits of the DL and ANN are also adopting by the SM for solving various data analytical issues discussed in further sections.

2.1.2. Naïve Bayes

Naïve Bayes is the simplest classifier, which does fast classification on the label data. For example, phishing is a favorite application of Naïve Bayes algorithm [33]. The phishing filter classifier classifies a label "phishing" or "Not a Phishing" on emails we received. To build machine learning designs, Naïve Bayes Classifier works utilizing grouped by identical elements which follow Bayes Probability Theorem. It is the classification of a natural word for subjective analysis of content based on the Bayes Probability Theorem [34]. A Bayesian network is a probabilistic graphical representation that represents a set of arbitrary variables and their qualified independencies via a directed acyclic graph (DAG). Following equations shows the mathematical representation of the Naïve Bayes classification.

$$P(k|a) = \frac{p(k/a)p(k)}{p(a)} \tag{1}$$

$$P(k|A) = p(a_1/k) \times p(a_2/k) \times \cdots .. \times p(a_n/k) \times p(k) \tag{2}$$

Where $P(k|a)$ in Eq. (1) means posterior probability, $P(k)$ means class prior probability, $P(a|k)$ means likelihood, and $P(a)$ indicates predictor. Here Eq. (2) represents the posterior probability series up to nth number [35].

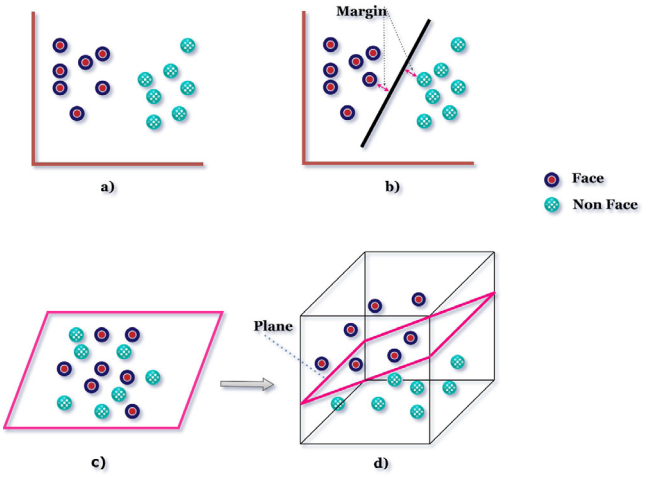

Fig. 5. An Example of support vector machine. (a) Raw data, (b) Grouping the data using plane, (c) Ungroupable data, and (d) data split using hyper plane.

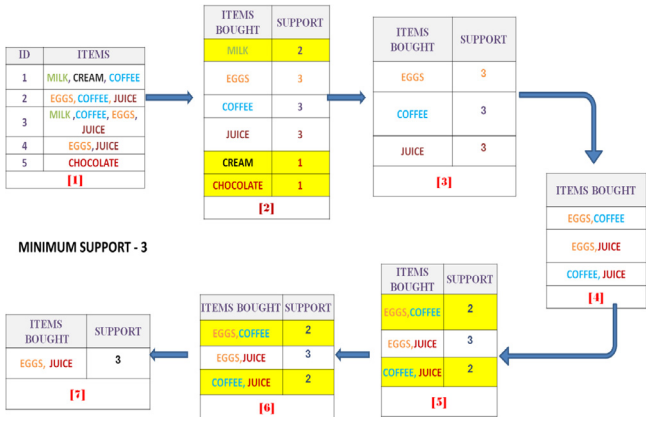

Fig. 6. Illustration of Apriori algorithm.

2.1.3. K-means

K-means is an unsupervised iterative algorithm, and it is used for clustering the input unlabeled data [36]. In this initially, we determine the number of clusters that need to be represented with a constant K. We then select a random K point as centroids. These centroids are not necessarily from our data. Then assign each data point to the nearest centroid, which forms K clusters. Calculate and place the new centroid for each of the newly formed clusters. Finally, we assign each data point again to the new nearest centroid. If any reassignment happens, we need to compute and place each cluster's new centroid again until achieving the best. For instance, let us consider this for Google Search results. When we search for SONY in Google, which returns all links that represent SONY as mobile, camera, and TV. If we apply the K-means algorithm to cluster these web pages based on device type, then it clusters similar devices into the same group. SONY mobiles pages as one group, SONY TV's pages as another group, and SONY camera related pages in the further group.

2.1.4. Support vector machine

Support vector machines (SVM) is a most popular SL method, and it is used for regression and classification [37]. The SVM classifies the given data by drawing a plane on them, and this shows the classification between the data inputted. The best classification means a better gap on the sides of a plane called the maximization of margin. So, this data can be visualized in multi-dimensional too [38].

In SVM, we use hyperplane to divide the data into groups. Fig. 5 illustrates the SVM for a better understanding of this method. In this example, we take two groups, such as a face group and a non-face group. In Fig. 5(a) has a set of raw data, which is divided into two groups using hyperplane, as shown in Fig. 5(b). While pointing the hyperplane, the margin between the groups must be the same. It may not be possible always to use hyperplane to group the data. In such situations, we require using multi-dimensional planes. In Fig. 5(c), we represent two types of data that cannot be divided using hyperplane. We solve this situation using the multi-dimensional plane, as explained in Fig. 5(d).

2.1.5. Apriori algorithm

For the given data, we can construct association rules using the Apriori algorithm, an unsupervised ML algorithm [39]. If X happens, then Y also happens, this rule is called association rule with a particular probability. Association rules produced *IF-THEN* arrangement. For instance, 'IF' somebody buys milk 'THEN'; they also purchase sugar to use it. For the algorithm to determine before-mentioned outcomes, it first watches how many members bought sugar along with milk. Out of 102 members who purchased milk, 86 members also obtained sugar in this way we can calculate. This Apriori algorithms works on few principles:

i. If a specific set frequently happens, then all the subsets of the specific set, also happen repeatedly.
ii. If a specific set occurs occasionally, then all the super sets of the specific set have occasional occurred.

In Fig. 6, it illustrates an example of the Apriori algorithm with minimum support three. First, one should take the data and count the support of all items. Remove the items with the minimum backing less than three. Combine two items and calculate the support of items then discard the items with the minimum support less than three. Finally, combine all three elements, calculate their support, and drop the items with a minimum of less than three. There is no frequent item because all items have minimum support of less than three.

2.1.6. Linear regression

Linear regression (LiR) algorithm describes the change in one attribute affects the other attribute, and this algorithm specifically explains the understanding of these two attributes. It describes how the dependent attribute communicates with the independent attribute. The dependent attribute is the reason for prediction, and the independent attribute gives meaning to the prediction [40]. We illustrate LiR with an example in Fig. 7. In this example, we plotted the linear regression graph from the given data. It also includes calculations of linear regression.

Linear regression generally used for predictive analysis. Linear regression is good when the relationship between a response variable and covariances are identified to be linear. This approach is good because it transfers focus to preprocessing and data analysis from statistical modeling. Linear regression ensures us without disturbing about the model's intricate details for learning to play with data. In summary, it is suitable for the learning process of analyzing the data. However, it will not be suggested for various practical applications because it generalizes the real-world problems.

2.1.7. Logistic regression

Logistic Regression ML algorithm is for classification tasks. We use Logistic Regression (LR) when the immature attribute is mentioned, and there are one or more autonomous attributes [41,42].

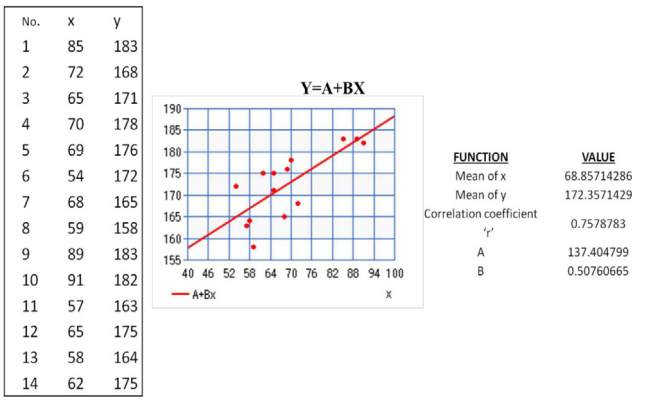

Fig. 7. A linear regression example with a sample data.

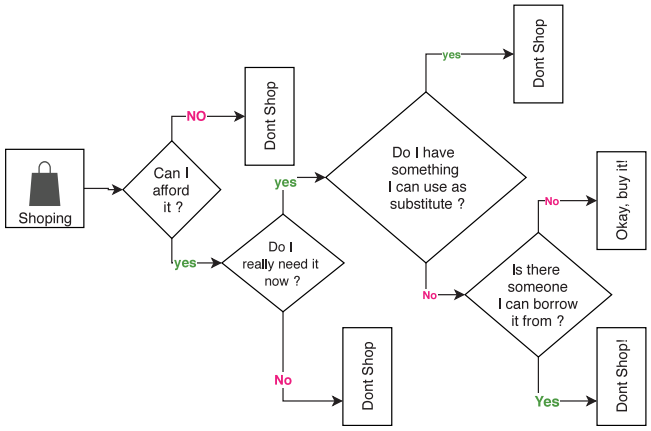

Fig. 8. Shopping decision guides using DT.

It is analogous to various linear regressions, and all of the same warnings employ. The goal of logistic regression is to find the best appropriate model to describe the relationships within the dichotomous characteristic of interest and a set of autonomous attributes. There are various types of Logistic regressions are available such as binary, multi, ordered, mixed, and conditional logistic regression. In binary or binomial logistic regression has two possible outcomes as *zero* or *one*. If an outcome of the depended variable is a remarkable outcome, then this outcome generally codes as *one* and the other outcome denoted as *zero*. Binomial logistic regression will use the independent variables to predict the odds. Multinomial logistic regression or multinomial logit can handle more than two dependent variables to represent as multiple categorical dependency. Ordered logistic regression handles ordinal dependent variables.

2.1.8. Decision trees

A decision tree (DT) is a graphical illustration that makes use of branching methodology to represent all possible outcomes of a decision, based on specific conditions [43]. There multiple regression tree algorithms available; those are very popular — Classification and Regression Tree (CART) [44], random forest, boosted trees. Decision trees (DT) cover both classification and regressing of machine learning [45]. For every branching, there is an outcome decision using a specific constraint for the given input. Giving graphical representation to all branching, we use Decision Trees. In a Decision Tree, we can get the classification using path from root to leaf node. A test on the attributes described in the Decision Tree is called as an internal node. The outcome of these tests is grabbed from the branch of the Decision Tree represented as leaf nodes. Decision trees are useful in data exploration and seek less data cleaning. But over-fitting issues are one of the challenging problems in DT. This issue will be solved by arranging constraints on model parameters and pruning. Decision trees are not fit to the continuous variables.

We explain the Decision Tree with an example using shopping guides in Fig. 8. In this example, we describe a shopping decision guides of an individual to buy a thing. A series of decision guides occurred while shopping to buy that product or not. These shopping decision guides include, whether that particular item is affordable to you and also checks the availability of substitute to it. Decision trees are also considered as generalized systems to induction rules derived from logistic data. While gathering decision, it reduces the number of questions in generating the decision tree is optimized, then that tree said to an optimal decision tree. We have several optimal decision tree algorithms such as CART, ID3, ASSISTANT, and Corner Last Slot(CLS).

2.1.9. Random forests

The Random Forest (RF) is a robust machine learning algorithm used for classification and regression [46]. The RF uses a bagging approach that describes with the group of DT [41,47]. The RF trains the system multiple times with an arbitrary dataset sample to provide an extraordinary prediction model. It gives definitive prediction by using the outcomes DTs as in the ensemble learning method. If a prediction that occurs multiple times in the decision tree, then that is the high-grade prediction of random forest. The central power of the random forest algorithm lies in its convenience in solving both classification and regression that afford good calculations of these. It handles the large datasets very gently without losing the dimensionality. Hundreds and thousands of input variables can easily handle and identify the signification variables that leads to noticing this as another dimensionality reduction method. It provides efficient methods for analyzing the missing data and maintains accuracy when the significant dimension of data considered absent. Bootstrap sampling and out of bag samples added advantages to Random forest algorithms.

2.1.10. Nearest neighbors

Nearest neighbors also have the capability of decision trees in applying to both classification and regression problems [48]. Nearest Neighbors classification problems used extensively in the enterprise, and it is a simple machine learning algorithm. This classification works as if there is a neighbor who votes for a case, nearest ones save that case of a neighbor. In the same way, it stores all appropriate circumstances and classifies new cases by considering all neighbor's votes. We can measure the case of each class using a distance function of k nearest neighbors. We use Euclidean, Hamming, Minkowski, and Manhattan distance for measuring distance among k nearest neighbors. The hamming distance used for categorical and Euclidean, Manhattan Minkowski are continuous functions [49]. If $k = 1$ means only one class of its nearest neighbors to a case.

2.2. Machine learning and big data

The data with a huge volume, variety and veracity structure called the big data. The SM data also belongs to the big data because of its variety, volume and veracity in nature [50–53]. The Fig. 9 shows comparisons among big-data with machine learning, artificial intelligence, and neural networks for the reference. We are aware of data and how it is generating in the world of internet and intranet. These data are generating from various sources like generating because of the extensive use of IoT's, from

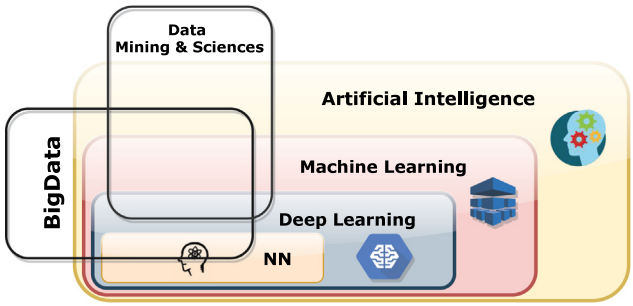

Fig. 9. Relationship between AI, ML, Deep Learning, Big Data and Data Sciences.

Table 2
Confusion-matrix for binary classification.

		Actual	
		Positives (1)	Negatives (0)
Predicted	Positives (1)	TP	FP
	Negatives (0)	FN	TN

Wireless sensor nodes, Uploading video's to the SM, uploading our photographs to the Instagram, posting our status or sharing our posts in the Facebook, Twitter tweets, and from many SM. The study says that 90% of the data generating on the internet are in the form of videos and photos [54].

Processing colossal amount of data leads to learn about Big Data. Applying machine learning techniques to big data [55] give much-sophisticated information. The data in big data are constricted using 5v's - Volume, Velocity, Variety, Veracity, and Value. Volume refers to the immense amounts of data generated every second. Velocity means the rate of newly generated data and the rate at which data moves around. Variety refers to the assorted types of data we can now use. Veracity refers to the complexity or dependability of the data. Value: is all well and good having access to big data, but unless we can turn it into value, it is useless.

2.3. ML classifier performance evaluations

Machine learning algorithms mainly applies four classification performance measures [56] available namely binary classification, multi-class classification, multi-topic classification, and measures for Hierarchical classification. In order avoid the ambiguity in the learning of performance measures, we start with a confusion matrix to find the best outcomes.

We can estimate the accuracy of classification by calculating the number of accurately identified class samples called true positives (TP). If the sample is identified, but that is not related to the particular desired class, then it is true negative (TN). If the sample is not identified, and it is not associated with any class, then it is false positive (FP). If they are not recognized as a class sample, then it is a false negative (FN). These four calculations organize a confusion matrix. We see this confusion matrix mostly in a tabular form that describes the performance of a classifier on a collection of test data to identify the real values. We presented a confusion matrix, which is utilized in performance evaluations of ML in Table 2.

2.3.1. Binary classification measures

Performance measures presents the most often used measures for binary classification based on the values of the confusion matrix. The Accuracy(AUC) captures a single point on the Reception Operating Characteristic curve; We measure the point at which Precision, Recall, combined AUC/Accuracy, and Fscore's properties because of its extensive use in text classification.

Accuracy- classifier total effectiveness

$$Accuracy = \frac{(tp + tn)}{(tp + fn + fp + tn)} \tag{3}$$

Precision – the portion of the identified as positive that necessarily positives [57]

$$Precision = \frac{tp}{(tp + fp)} \tag{4}$$

Recall (sensitivity)– The part of predicted concrete (positive) samples that have identified as concretes (positives) [58].

$$Recall = \frac{tp}{(tp + fn)} \tag{5}$$

F-score — It gives relation between a classifier and its chosen labeled [59].

$$F - score = \frac{((\beta^2 + 1)tp)}{((\beta^2 + 1)tp + \beta^2 fn + fp)} \tag{6}$$

Specificity – How efficiently a classifier recognizes negative labels

$$Specificity = \frac{tn}{(fp + tn)} \tag{7}$$

AUC – Classifier can withdraw the false classification.

$$AUC = \frac{1}{2}\left(\frac{tp}{(tp + fn)} + \frac{tn}{(tn + fp)}\right) \tag{8}$$

These equations are used to measure the performance of the binary classification. The Eq. (3) specifies the overall effectiveness of a classifier. Eq. (4) presents class agreement of the data samples with the positive samples given by the classifier. The Eq. (5) identifies the positive samples to enhance the effectiveness of a classifier. Eq. (6) specifies the relation among the classifier given positive data samples. Specificity (7) gives a possible efficient way to identify the negative labels of a classifier. Using Eq. (3) we can avoid false classification [60].

2.3.2. Multi-class classification measures

Multi-class classification Measure using overview of the measures for many classes C_i: tp_i are True Positive for C_i, and fp_i is False Positive, fn_i is False Negative, and tn_i is True Negative respectively. M and μ and indices represent macro- and micro-averaging with constant i.

$$Average\ Accuracy = \frac{\sum_{i=1}^{l} \frac{(tp_i + tn_i)}{(tp_i + fn_i + fp_i + tn_i)}}{l} \tag{9}$$

$$Error\ Rate = \frac{\sum_{i=1}^{l} \frac{(fp_i + fn_i)}{(tp_i + fn_i + fp_i + tn_i)}}{l} \tag{10}$$

$$Precision_{\mu} = \frac{\sum_{i=1}^{l} tp_i}{\sum_{i=1}^{l}(tp_i + fp_i)} \tag{11}$$

$$Recall_{\mu} = \frac{\sum_{i=1}^{l} tp_i}{\sum_{i=1}^{l}(tp_i + fn_i)} \tag{12}$$

$$Fscore_{\mu} = \frac{((\beta^2 + 1)Precision_{\mu}Recall_{\mu})}{(\beta^2 Precision_{\mu} + Recall_{\mu})} \tag{13}$$

$$Precision_M = \frac{\frac{\sum_{i=1}^{l} tp_i}{\sum_{i=1}^{l}(tp_i + fp_i)}}{l} \tag{14}$$

$$Recall_M = \frac{\frac{\sum_{i=1}^{l} tp_i}{\sum_{i=1}^{l}(tp_i + fn_i)}}{l} \tag{15}$$

$$Fscore_M = \frac{((\beta^2 + 1)Precision_M Recall_M)}{(\beta^2 Precision_M + Recall_M)} \tag{16}$$

From Eq. (9) to Eq. (16) are used to measure multi-class classification. Eq. (9) identifies the classifiers average per class effectiveness. The Eq. (10) specifies the classification error by average per class. In Eq. (11) specifies the agreement of the data class samples of a classifier. The sums of per text decisions gives effectiveness of classifier to identify the class samples by Eq. (12). Eq. (13) identifies the relations between positive samples data and classifier give sums of per text decisions. This Eq. (14) gives an average per class agreement of the class labels data of a classifier. The Eq. (15) used to identify the class labels by an average per class effectiveness of a classifier. Using Eq. (16) we can specify the relations between positive labels data given based on a per class average, which is given by classifier [61,62].

2.3.3. Multi-topic classification measures

The I is the marker function; $L_i = L_i[1], \ldots, L_i[l]$ denotes class labels to x_i, $L_i[j] = 1$ if C_j is present between the labels and *zero*, if not; L_i^c are labels of a classifier, L_i^d are the labels of data with constants i and j.

$$Exact\ Match\ Ratio = \frac{(\sum_{i=1}^{n} I(L_i^d = L_i^d))}{n} \tag{17}$$

$$Labeling\ F-score = \frac{\sum_{i=1}^{n} \frac{2\sum_{j=1}^{l} L_i^c[j] L_i^d[j]}{\sum_{j=1}^{l}(L_i^c[j]+L_i^d[j])}}{n} \tag{18}$$

$$Retrieval\ F-score = \frac{\sum_{j=1}^{l} \frac{2\sum_{j=1}^{n} L_i^c[j] L_i^d[j]}{\sum_{j=1}^{n}(L_i^c[j]+L_i^d[j])}}{l} \tag{19}$$

$$Hamming\ Loss = \frac{\sum_{i=1}^{n}\sum_{j=1}^{l} I(L_i^c[j] \neq L_i^d[j])}{nl} \tag{20}$$

Multi-topic classification can be measured using the formulas specified from (Eq. (17) to Eq. (20). The exact match ratio Eq. (17) used to specify the average per-text exact classification. Eq. (18) specifies the average per text classification with partial matches for labeling F-score. The retrieval F-score Eq. (19) specifies the average per class classification. The hamming Loss formula (Eq. (20)) identifies the total error of the average per example per class [63,64].

2.3.4. Measures for hierarchical classification

$C_{\downarrow}$ is class C's subclasses; $C_{\downarrow}^c$ is classifier assigned subclasses ; $C_{\downarrow}^d$ – labels of data class; same notations use to super classes, that are represented by $C_{\uparrow}$.

$$precision_{\downarrow} = \frac{(|C_{\downarrow}^c \bigcap C_{\downarrow}^d|)}{(|C_{\downarrow}^c|)} \tag{21}$$

$$Recall{\downarrow} = \frac{(|C_{\downarrow}^c \bigcap C_{\downarrow}^d|)}{(|C_{\downarrow}^d|)} \tag{22}$$

$$Fscore_{\downarrow} = \frac{((\beta^2+1)Precision_{\downarrow} Recall_{\downarrow})}{(\beta^2 Precision_{\downarrow} + Recall_{\downarrow})} \tag{23}$$

$$precision_{\uparrow} = \frac{(|C_{\uparrow}^c \bigcap C_{\uparrow}^d|)}{(|C_{\uparrow}^c|)} \tag{24}$$

$$Recall{\uparrow} = \frac{(|C_{\uparrow}^c \bigcap C_{\uparrow}^d|)}{(|C_{\uparrow}^d|)} \tag{25}$$

$$Fscore_{\uparrow} = \frac{((\beta^2+1)Precision_{\uparrow} Recall_{\uparrow})}{(\beta^2 Precision_{\uparrow} + Recall_{\uparrow})} \tag{26}$$

Hierarchical classification can be measured using the formulas from (Eq. (21) to Eq. (26). The Eq. (21) used to specify the classifier's positive agreement on subclass samples. Eq. (22) specify data's positive agreement on subclass samples. The Eq. (23) specify the classifier through relations between positive sample data.

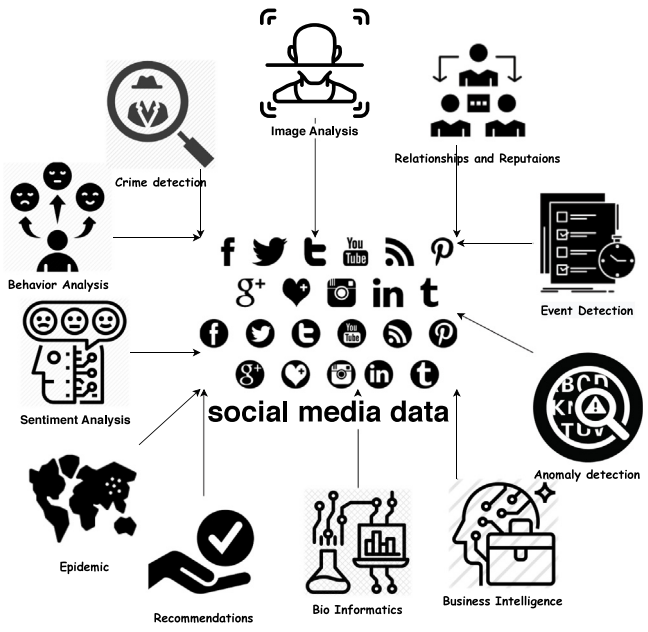

Fig. 10. Applications of SM analysis through ML.

The Eq. (24) identifies the classifier's positive agreement on super class samples. Eq. (25) specifies the data's positive agreement on super class samples. Eq. (26) represents the relations among classifiers positive superclass samples [65,66].

3. Machine learning in social media analysis

The SM analysis is the procedure of studying social compositions toward the utilization of networks and graph theory. Its analysis characterizes networked compositions as far as nodes (people, things inside the network, or individual actors) and the ties, joins, or edges (interactions or relationships) that correlate them. There are many SM applications available in the world of networks. Some of the regularly using SM in our routine life are Facebook, Snapchat, QQ, WhatsApp, WeChat, QZone, Tumblr, Instagram, Twitter, etc. To optimize our work, we are narrowing the SM Analysis [5] to multiple application areas based on their scope to apply Machine Learning concepts. In this paper, we systematically survey the articles on SM analysis published up to 2017 by exploiting the methodological decisions on these application areas of SM. The applications considered for surveying are anomaly detection, behavioral analysis [67], bioinformatics, business intelligence [68], crime detection, epidemics, event detection, image analysis, recommendations, relationships [69], and reputations, and sentiment, opinion and emotions analysis. Fig. 10 shows the applications of social media analysis through machine learning.

3.1. Anomaly detection

Anomaly detection is a procedure used to identify unusual patterns that do not agree to expected behavior, called outliers. It has multiple applications in business, from intrusion detection to system health monitoring, Network failure anomaly, information diffusion anomaly, and fraud detection in credit card transactions to fault discovery in operating environments. Future research depends on extreme learning techniques to identify a defect in the area of SM. Network failure anomaly causes a reduction in bandwidth to the users of SM. In [70], authors presented their article on root cause analysis to detect network failures by applying machine learning techniques. They discussed statistical methods

for root cause analysis, such as applications of random forests for diagnosis and training and study of system properties. They applied different methodologies to find errors that lead to network failures. In [71], the author's primary intention was to focus on the problem of distributed detection in multi-agent networks. Authors developed a Kullback–Leibler cost to compare algorithms to its centralized counterpart, which provides an optimization-based framework. By considering network size, a spectral gap, they implemented asymptotic learning with finite-time analysis. Their data analyzation limits only a few areas of datasets; if we use the wide range of datasets, we get even more meaning full results.

SM has many studies available to explore the topic of information diffusion anomaly. For example, in [72], the authors discussed on the visualization of information diffusion anomaly discovery in SM. They applied Multidimensional scaling (MDS), hierarchical topic clustering, and One-class conditional random fields (OCCRF) machine learning algorithms. Fluxflow is the working title of their paper for providing the visual representation of information diffusion across social media. They used to work on Twitter tweets to convey their work vicinity. They processed the Twitter data using Apache Hadoop and stored it for queried analysis. They applied the hierarchical topic clustering algorithm and feature-space distributions (tweets view, states view, and features view) via MDS to clustering the data. Finally, they discover anomaly using OCCRF. FluxFlow foundationally joins an arrangement of algorithms to portray re-tweeting threads, did anomaly discovery in that, and create a successful visual interface, comprising of a few new representations and numerous organized perspectives, to introduce the model yields.

In [73], the authors presented the philosophical representations of statistical learning and extreme machine learning. The statistical learning theory will assess the performance of a predictive model by applying essential and adequate conditions for non-parametric inference. Their statistical learning technique implemented using the worst-case approach, unlike others. Authors described an example of parallel extreme machine learning implementation by a regression-based Map-Reduce framework using SPARK framework for emotion detection and polarity recognition applying on social media generated data. They employed a statistical learning technique's predictive model to achieve extreme machine learning for scaling large data sets. We can extend this work not just by applying on supervised data but also on semi-supervised data too.

If an anomaly exists in a network of social media, which not only disturbs the networked people but also changes their behavior. In [74], authors presented a way to deal with anomaly detection among unauthorized users and their behaviors. Authors specifically consider the fact that social media will make a huge impact on society. The content and messages passed, and communicating will leads to negative effects on society. To avoid such a situation, it is always better to prevent the spreading of rumors. For this, the authors developed an algorithm TargetVue, which presents the communication of user activities to represent the behaviors visually using ego-centric glyphs.

Anomaly detection in SM is a challenging task. We addressed the variety of suitable features that describe the typical behavior of a user in SM for anomaly detection. It is essential to take the best feature, which represents the normal user behavior in the SM environment. Another major challenge for detecting anomalies in SM is the evaluation and comparison of different methods. Several existing strategies are disclosed with distinct problem domains and data formats in memory. Besides, there is a need for openly accessible datasets with identified facts. In the future, anomaly detection may become the most dominant study area that plays a crucial role in privacy-preserving [75,76], detecting cloning, malware, phishing, spamming, and so on. Sophisticated algorithms are available to analyze the anomalies in SM for future work in this area.

3.2. Behavioral analysis

Behavior analysis is a natural science that seeks to understand the behavior of individuals. In this digital era, analysts search for versatile ways to analyze the behaviors of a person. One such a possible way is identifying the behavior of a user using ML on SM analysis for opinion mining [77], text forensics, and so many [78–80]. Behavioral analysis is also used to develop behavioral prediction [81] patterns too. Authors in [82], developed a method to identify the behavior of normal users from spammers in social media. They worked mostly for identifying pornographic spammers from the twitter data using NodeXL. They apply few Machine Learning algorithms (Logistic Regression, Bayes Net, AdaboostM1, J48, and Random Forest) to classify the data and discovering the behavior of illegitimate users. Authors can use opinion mining, text mining, Natural Language Processing for more accuracy. As per our study, authors can use image analysis techniques for identifying pornographs to get even more sophisticated results. In [83], the authors presented a survey on user behavior in SM. They studied different perspectives, such as using mobile platform social behavior, connection and communication among users, traffic movement, and malicious behavior of SM users.

A community in the real world has a specific job to do. For instance, the biker community is a community that plans voyages, or senior citizens club has senior people to share their feelings in the community. Like this every online Networked group has specific intentions, identifying such groups requires community detection techniques. Authors in [84], developed a new concept, "sentiment communities" for improving the online marketing and business strategies. Sentiment communities are users group who have an exclusive mindset on particular products or services on social networks. For detecting sentiment communities, they apply semi-definite programming and random rounds to combine the community detection. In [85], authors proposed the model of rumor dissemination in SM community. They developed a rumor dissemination model to adopt a specific strategy based on a combined logit model. Using the mean-field theory approximation method, they derived a rumor dissemination character and developed a framework for it.

The Health of an SM group depends on the behavior of users in it. We cannot always expect that all users are legitimate and good users. Some users are illegitimate, and their behavior could hurt other user's sentiments [86]. To identify such users, analysts of social media utilize the approaches of affective computing [87]. Authors in [88], recorded general tasks of affective computing and sentiment analysis and presented a broad categorization for illegitimate users. We observed that authors only focused on identifying the user's credibility based on the bullying applied by the user. Their article placement and content delivery are good though they miss exhibiting the relation between the other illegitimate users. For their study, the authors used the Markov chain model to develop the user's credibility. Authors in [89] developed a novel concept to discover the group of peoples who will get thrilled and elegant to share their promoted content. The authors developed the model that identifying the influential node was influential activity maximization model. They said their model finds the pre-eminent and trusted users to the online users. They said they used that trusted user to spread their information. This activity maximization model uses the NP-hard's diffusion models, such as the Linear threshold (LT) model and Independent cascade (IC) model.

In [90], the authors analyzed human behavior using social media networks and projected how these could be quantifiable. They developed a method egoSlider to visualize the nature of online users. They make them into a cluster according to their

egocentric view to identifying the behavior. In [91], authors address the text-based cyberbullying discovery. They developed a semantic dropout noise and enforcing sparsity, and they produced a semantic-enhanced marginalized denoising auto-encoder learning model for cyberbullying discovery. Besides, word embeddings used to build and improve bullying word lists that are initialized by domain expertise. The performances of their approaches verified through two datasets from social media sites such as Twitter and MySpace. As per our study, we can improve the recognition of bullying using word order in messages.

The behavioral analysis of SM not just defines community interest but also gives fluidic business solutions to critical problems. In [92], they reviewed how to prevent customer churning in the telecommunication industry. It is a fact to the industry that the customer retention costs more than the customer acquisition. Authors developed an approach to stop customers from churning using machine learning techniques. They worked on Monte Carlo simulations to support their work. They compare this simulated data on five machine learning techniques, such as Artificial Neural Networks, Decision Trees, Support Vector Machines, Naïve Bayes classifiers, and Logistic Regression classifiers to identify the best classifier for this work. They found that under the boosted network settings, SVM gave the best score to use. Hence machine learning in customer churning prediction advances the business of this domain.

In [93], authors developed an approach that matches many objects with many other linguistic objects using Rematch. A rematch is a short form of the relational matching model. Their work gives different language, and locale people of same interests can be identified using this method. They cluster similar words of multiple languages and use an infinite relational model. They said their model is a nonparametric Bayesian extension of the stochastic block model. The researchers in the world of social media developed many tools to find the behavior of users. This behavioral analytics not only extracts the information from SM but also improves the performance of the systems. In [94], the authors mentioned the importance of tools used to present better information on animal movement and behavior. The primary factor that needs to consider is to monitor the individual as well as environment interaction. By taking the help of network theory, we implement the network analysis, which plays a significant role in refining, identification of animal movement ecology, and behavior.

In [95], the authors described a behavior-aware user response system. They use a hierarchical ensemble learning framework for heterogeneous SM data. Their framework automatically extracted the feature and transmitted the data. They also developed the extended version of the multiple Kernel SVM and Hierarchical multiple kernel SVM (H-MK-SVM. They said they applied four types of data for improving the prediction performances, such as keywords, tags, longitudinal engagement, and individual behavioral data. They classified the user response model using SM's heterogeneous data and ensemble learning; the framework they developed intelligently blends the e-commerce and SM data available in their databases. To improve the performance of a prediction system, they combined the multi-channel network, video, text, audio, and unlabeled data toward straight marketing models and customer relationship management.

The journey of understanding human behavior at scale has quite recently started. Most present work in SM considers examining behavior from an information mining, or machine learning point of view [97]. Hence, behavior analysis applying to social media could lead to developing good communities in society. It leads to understanding the people in a meaningful way, so as it reduces the people participating in the crime. Based on the community's identified interest, one can able to enhance or develop their business strategies. Sometimes the behavior analysis drives to practice imperfections too. The understood behavior of users derived from the behavior analysis can be used as a trusted weapon for fraud persons to imitate real users. We presented this behavior analysis summary table shown in Table 3.

3.3. Bioinformatics

In 1979, Paulien Hogeweg invented the bioinformatics terminology that studies the biological systems using information technology [98]. Bioinformatics is the area of molecular biology that combines both information technology applications and computational science. This bioinformatics used to develop biological software and construct algorithms to analyze the data that belongs to biology. For example, bioinformatics is very useful in improving drug instances [99], proteins, metabolic pathways, and genes data. We knew that biological data was being in raw form, we cannot store this as we do to store traditional data, it requires specific storage house for organizing, storing, and manipulating such extraordinary data. Using biological databases and software, data scientists extracts and uses this information for improving the existed data.

Bioinformatics applying in social media data leads to a new category of information extraction. It adds another power to social media and confirming it has a broad spectrum of data available for researchers to explore its applications [100]. Bioinformatics data could lead to developing biological devices to enhance our daily life more meaningful [101]. For example, in [102], the authors developed a device which identifies breathing pattern and recognizes the snoring time during sleeping and corrected by applying several sleeping order techniques. They find breathing disorders by applying machine learning algorithms to the wireless sensory device that has participation in SM. Bioinformatics and social media combination helps to create healthy community for bio informaticians [103]. Such communities help the society to discover more meaningful drugs [104].

Bioinformatics is very useful in improving drug instances, proteins, metabolic pathways, and genes data. This leads to developing biological event detection [105] and monitoring systems too. In [21], authors reviewed pharmacovigilance using social media data. Diversified data available on social media can give users concern about the deceases and some medicinal recommendations to other users. These data are much helpful to the pharmacy companies to develop new drugs and also improve the existing drugs. In most of the cases, users face adverse effects while using inadvertent drugs. So mainly, adverse drug reactions monitor gives the best hope to the pharmacy companies to improve their medications by analyzing these data. The authors explained unwanted drug monitor using ML techniques on social media data.

In [106], authors applied machine learning in real-time applications in the field of medical, biological parts. They developed an approach that will identify the pre-functional behavior of organs with inbuilt behavior in a sinister way for the neurodevelopmental outcome. By applying such techniques helps to study the growth of the brain in infants. Authors explained multivariate pattern analysis, machine learning, and pattern classification approaches are well-suited for high-dimensional neuroimaging data. In [107], authors developed a smartphone application, which works in real-time for patient monitoring. Continuous monitoring of the patient and observing the symptoms of the patient will be intimated as a message or call to the nearest clinical person or doctor. By comparing smartphone collected data with predefined data, necessary actions will take place.

In [108], authors presented a system on Location-based Services in an indoor environment by using an Indoor Positioning System. They developed an indoor positioning system using

Table 3
Summary of Machine learning in Behavior analysis.

Study	ML Used /Model/Framework	Description	Evaluation	Datasets/ Data utilized
[82]	RF	Behavioral analysis and classification of spammers distributing pornographic content in social media	Accuracy - 91.96%	Twitter
[92]	RF, Naive Bayes, DT, LR, SVM	Predicting the churn rate of a customer.	accuracy - 97% F1-over 84%.	UCI ML Repository Churn dataset
[93]	ReMatch model	Unsupervised Many-to-Many Object Matching for Relational Data	Accuracy-0.946, AUC-0.926	The MovieLens, Wikipedia
[95]	SVM	Behavior-aware user response modeling in socialmedia is learning from diverse heterogeneous data	AUC-67,20	Rec-log-training, User Profile, Item and User SNS
[96]	LBM, Spice framework	Socially-Driven Learning-Based Prefetching in Mobile SM	Accuracy-84.5%	Twitter

Weighted Extreme Learning Machine and Signal Tendency Index. To achieve balancing normalized fingerprints, they applied the Signal Tendency Index and a similarity metric by transforming the received signal strength to location and fingerprint-based system. Working with interdisciplinary concepts like bioinformatics gives state of the art challenges to the researchers. Various studies are going on in bioinformatics through social media. It constitutes extra energy to SM and validating it has an extensive spectrum of data ready for researchers to explore its applications. Bioinformatics data could manage to develop the vital tools to intensify our regular days more significant.

3.4. Business intelligence

BI(Business Intelligence) is an useful methodology, architectures, and techniques that turn data into useful information that influences efficient business activities. It is indeed a collection of applications and services that convert data into experience and wisdom that can be used. Business intelligence (BI) concept applying on SM data improves the businesses of many domains. Business intelligence that captured from reviews, comments, likes, tweets, and shares of specific and related business data, that distributed among social media will help the companies in delivering good products and services [109,110]. The graphical abstraction of Business intelligence from SM using Machine learning are explained in Fig. 11.

For example, authors in [19], developed business intelligence methods using opinion mining or sentiment analysis of social network analysis. They gather information from three websites, such as Yelp, Expedia, and Tripadvisor hospitality and tourism domain sites. Based on the user's comments from these sites, they analyze the opinions of the users called opinion mining. They quote that several companies used there data to improve their businesses. The authors applied the LDA algorithm for the classification of words that are posted by the user as comments on these sites. They discussed how a particular star hotel improved its business by considering customer's reviews on them in these mentioned sites. They said LDA delivers good result for analyzing than that of other algorithms for their study. Customer satisfaction and trust are essential factors for business development. It is a known factor in business intelligence that if the companies do not know about their customers, they will not satisfy them with their service [111]. In [112], the authors implemented an algorithm based on firm trust over the statement that, the business will depend on customer's satisfaction. Based on customer reviews, business people have to think of their business strategies. The authors found a way to analyze the customer's reports on services provided by the companies to enhance their business plans. They concluded that their study leads to understand the customers in best possible way. Understanding and considering the user's knowledge or experiences are beneficial in identifying the business flaws. Rectification of these flaws will increase companies' business. Authors in [113], presented a way to discover the experience of a user using their social media information data by recalling and inferring information. These users Knowledge discovery leads to the top benefits of business intelligence. They presented their work using Ontologies based Artificial Immune Systems and an abstract global knowledge representation model. Authors said ther study helps to learn co-learning scenarios along with educational and competence situations. In their work, authors used datasets of Twitter and LinkedIn. They also quoted that concerning their study, one can enhance their work using even more diversified knowledge data by giving expertise users and also applying compatibility degrees among users and jobs [114].

We can use media data to identify abnormal events and make decisions promptly is a beneficial topic. Authors in [115], proposed a method for developing the business by understanding the user's emotional mood. Their paper proposed a method for unusual sentiment detection on social networks that obtains the correlation between different peculiarities of the sentiments and maintains a specific value of time by group estimation of the collective probability density of data sets. They utilized the multivariate Gauss distribution with the power-law distribution to model and examined the social media users' emotions and unstable emotional mood. Their results showed that individual users' unbiased, comfortable, and disturbing sentiments accept the normal distribution through the distribution test, but the upset and angry sentiments do not. Besides, sentiments from social media groups agree with the power-law distribution model they proposed, but personal sentiments were not affected. They said their method could help companies and state bodies to do arrangements according to the user's affective nature, interrupt early, or adopt proper strategies where required.

The cutting knowledge of business intelligence rose to develop BI analytics [116]. Recognizing such expertise from social media is a hectic task. It requires lots of manual resources, so the solution to this is applying machine learning techniques or developing social media analytics for business intelligence [117]. Social media analytics reduces the workload and improves the business by doing useful work for companies.In [118], the authors explored the importance of SM analytics. They demonstrated different processing stages of analytics and the general techniques available for SM analytics. They also presented how these analytics increases the value of the business. In [119], authors developed an approach called Supply Chain Management u sing Twitter tweeters(users who tweet), which itself consists of descriptive analytics, content analytics, and network analytics (NA), to identify the tweeter flows on specific topics. To spread information and share, one should use these techniques. Authors in [120] developed a dashboard for business intelligence to track the quality of a topic or news source that shared to social media sites such as Twitter and Facebook. Social media topical output is the number of times a post is liked, shared, commented, etc., on social media. They used text classification techniques such as Naïve Bayes, SVM, and Decision Tree to explain a news post of a topic on social

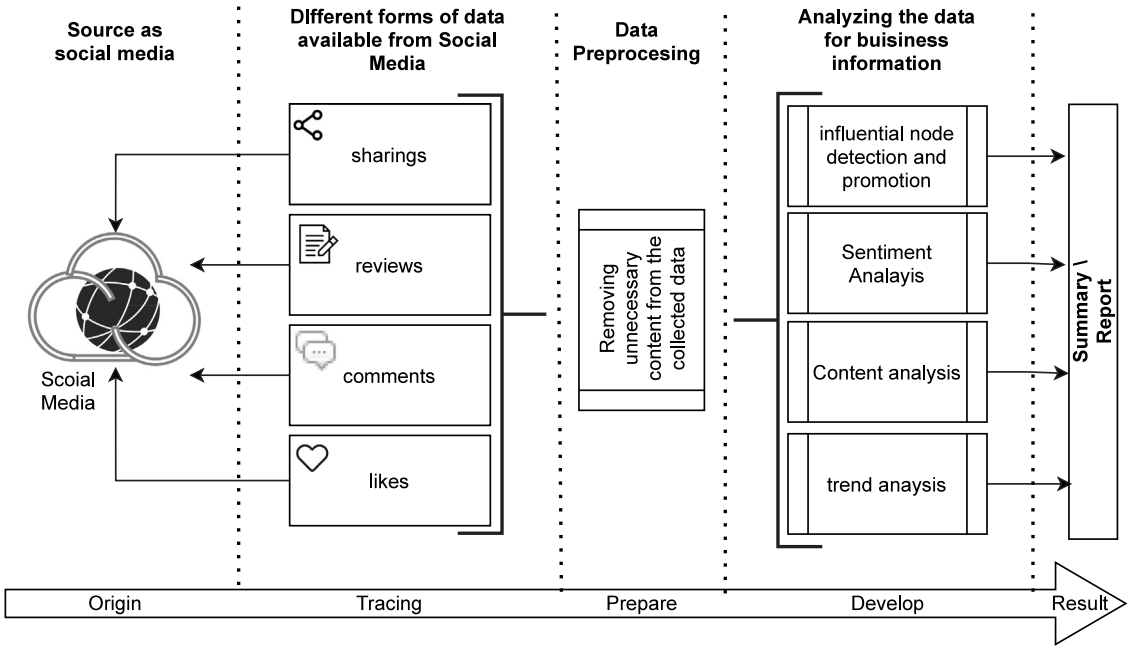

Fig. 11. A graphical abstraction of Business intelligence from SM.

media. They said they got the best accuracy, 77.99%, with the SVM algorithm.

To attract the customers, companies offers discounts on products to improve the business. These business discounts of vendors sometimes does not knew to the customers, who loves to grab such. If we have a system which finds the deals and inform the customers with a decision to pick that item at this time is good or bad are helpful to the users. Identifying these discounts manually is a hard job, so solution to this is applying machine learning techniques. We can also use social media broadcasting of services [121] or developing social media analytics for business intelligence. In [122], authors developed a new data processing concept called Decision learning. Decision learning is the combination of learning the data and making a decision from that learned data. For data leaning, they use social media users behavioral study data. This behavioral study includes the study of sequential behavior for data sharing and evolutionary behavior for deal selection like restaurant deals, deals on products. They concluded that their study gave satisfactory results using SVM algorithm.

In many cases, the decisions of a deviated customers influence the companies future. To overcome these customer deviations or churn rating of customers, its good to use ML techniques to predict such customers. Predicting the churn rating of customers gives a solution to the companies about their product health. for example, In India, Reliance Jio a telecom start-up attracts an enormous number of customers with incomprehensible discounts. That does that many low profiled or newly entered telecom companies to shut their doors. So churn prediction is a significant prediction system. In [92], they have surveyed how to prevent customer churning in the telecommunication industry by observing few social media data. This business intelligence fact explains to an industry that the customer holding costs are more than the customer addition. Authors developed an approach to stop customer churning using machine learning techniques. They found that under the boosted network settings, SVM gave the best score to apply. Hence machine learning in customer churning prediction advances the business.

In [123], they developed a framework called VOZIQ which work on sentiment benchmarks for business intelligence using social media data. They created sentiment benchmark using sentiment analysis, text mining, and other SM strategies on user-generated data on user accounts. In reality, applying the valid BI method in association with SM marketing can improve the business. Companies can accumulate profound insight toward target audiences' behaviors and decisions, as well as enhance support control, boost exchange rates and predict coming drifts and issues. Organizations contemplating to fund in BI for SM marketing want to ensure they wish a BI solution that is manageable, scalable, simple to manage, and has the capability of instantly analyzing online sources for its brand's related SM platforms [125, 126]. The summary of articles surveyed in business intelligence of SM presented in Table 4.

3.5. Crime detection

Detecting Crime in social media generated data, such as tweets, likes, posts, comments are one of the major approaches in these eras. There are several work done in SM using machine learning and those are described as follows. Authors in [127], developed an algorithm COMPA, which detects compromised accounts on social media. The compromised accounts give pressure to the user, sometimes criminals left no option to the user and compelled them to death. Hence, COMPA algorithm develops the individual behavioral profile of SM user based on their postings. It compares every new action done by the user with the datasets already prepared. If any activities not in support of the dataset, then the algorithm triggers the anomaly in it. So, it can detect compromised high-profile social media accounts, and compromises of regular accounts. However, Their algorithm can be compromised if the attacker knew the behavior of the user.

As per the recent statistics, 83% of all messages are spams. Spam detection is a promising technique, but almost all are not biased, means they have no prejudice concerning the incoming data. Authors in [128], focused on social spam detection using a Novel methodology that integrates word, topic, and user-based features and applies labeled latent Dirichlet allocation (L-LDA) and incremental learning to improve the accuracy of spam detection on social media sites. Authors in [129], developed a framework called anti cyberstalking text-based systems for detecting

Table 4
Summary of Business Intelligence in SM using ML.

Study	ML Used / Model/ Framework	Description	Evaluation	Datasets / Data utilized
[92]	RF, Naïve Bayes, DT, LR, SVM	Predicting the churn rate of a customer.	accuracy - 97% and F-measure -84%.	UCI Machine Learning Repository Churn dataset
[113]	Gaussian Naïve Bayes	Crowdsourcing and collaborative learning environments based on SM.	reward accuracy mechanism sampling probability -1	Twitter, Facebook, LinkedIN
[122]	DT	Data analytic learning with strategic decision making.	–	Twitter hashtag, Memetracker, and Yelp
[123]	Social media competitive analytics framework, VOZIQ, N-Gram, VOC3	Social media competitive analytics framework with sentiment benchmarks that can be used to glean industry-specific marketing intelligence.	N-Gram accuracy - 0.8242	Twitter, Facebook, Youtube
[124]	SVM, data-driven ITS model, Unlicensed Taxi Identification Model (UTIM)	Identifying unlicensed taxi	Accuracy equal to 86.667%,	Electronic Traffic Bayonet Device (ETBD) collected data
[120]	SVM, NB and DT	Developed a BI model using SM data	Accuracy(SVM)-77.99%	Facebook& Twitter

and avoiding cyberstalking. The framework uses algorithms of textual analysis and Machine Learning techniques to detect cyberstalking and harassment. The results of the works given by the authors never give any law abide proofs for cyberstalkings and harassment. We can enhance their work by developing an evidence-based advice on how to respond to harassment

The Health of an SM relies upon the conduct of clients in its [130]. We cannot generally expect all clients are honest to goodness and great clients. A few clients are ill-conceived, and their conduct could hurt multiple client's assumptions. Authors in [131], Categories the cyberbullying into direct and indirect cyberbullying and gave the solution for detecting these direct and indirect cyberbullying using Machine Learning classifiers algorithms. They also developed a framework which did not only classifies the type of bullying but also gives credibility to the user who participates in the bullying. By considering that users' trustworthiness, other users should identify his/her personality and take safety measures while dealing with them. As per our study, their research focused only on finding the credibility of the user based on the category of bullying he/she does. It does not reveal any status of the relationship between the users.

In [132], the authors presented a set of techniques for predicting aggressive comments in SM. In a time when cyberbullying has, unfortunately, made its entrance into society and the Internet, it becomes essential to find ways for preventing and overcoming this aspect. They applied machine learning techniques for automatically detecting cases of cyberbullying. Their primary task in cyberbullying detection consisted of aggressive text detection. In [91], the authors addressed the text-based cyberbullying discovery. They introduced a semantic dropout noise and enforcing sparsity, and developed a semantic-enhanced marginalized denoising auto-encoder knowledge model for cyberbullying discovery. Furthermore, word embeddings used to build and establish bullying word lists that are initialized by domain expertise. The performance of their methods verified through Twitter and MySpace datasets. Their work can be improved by improving the recognition of bullying using word order in messages.

Cybercrime is all over, and the minimum you can do is perused up about it. You may think the Internet and email tricks, influence the individuals who are not well informed or do not stay aware of the day by day news, yet that is not valid. From IT experts to instructors, to writers, individuals from all parts of society and expert fronts have fallen for these impeccably arranged online cheats, Business Compromise Scams, Pharming, and so forth, which may befuddle anyone. What is more, the idea of hazard taking does not work in a unique circumstance. With the appearance of informal organization and the broad utilization of messaging, these tricks obtained a significant decent footing. In [124], authors have taken a real-time scenario for the identification of fake taxi's as well as unlicensed drivers. Nowadays, society's safety is dragged down by fake taxis, which will impact the economy of public transport. To overcome this issue, the authors developed the Data-Driven Intelligent Transportation System by using the candidate selection model and candidate-refined model.

For instance, authors in [20], provided a detailed survey on detecting fake profiles available in SM. They also discussed how previously machine learning algorithms used in identifying fake profiles. They surveyed algorithms like DEPOLD, Markov clustering, which is used in detecting fake profiles of other authors. In [133], authors presented many ways we are compromised to vulnerabilities using mobile phones. They attempted to work on both social media data and data available through the internet as static and dynamic. They applied machine learning techniques to sentiment analysis. They developed an application that can automatically detect vulnerabilities in the Android system. In [134], authors conveyed the process for identification of frauds taken place in the unauthorized use of credit cards. Depending upon the usage of customers, the author's developed two algorithms, APATE (Anomaly Prevention using Advanced Transaction Exploration), to identify the spending behavior of the customer and the other one at merchants' end to determine the time-dependent suspicious score for the network object.

Obtaining the models and trends in crime is a challenging part. To identify a type, crime analysts necessitate plenty of time, looking for data to discover whether an appropriate crime fits into a known model. If the crime does not fit into a current model, then the data must be classified as a new pattern. After identifying a model, it can be used to prevent, anticipate, and predict crime. The summary table for Crime detection and application of SM presented in Table 5.

Table 5
Summary of Crime detection strategies in SM using Machine learning approaches.

Study	ML Used	Type of crime detection	Performance	Datasets
[20]	MRF	Fake profile detection.	Integro-Precision -95% Sibil belief - Accuracy -100%	Facebook
[91]	Bow, SVM, LDA, Semantic-Enhanced Marginalized Denoising Auto-Encoder (smSDA)	Cyberbullying Detection based on Semantic-Enhanced Marginalized Denoising Auto-Encoder.	Twitter (Accuracy-84.9, F1-71.9), MySpace(Accuracy-89.7, F1-77.6)	Twitter, Myspace
[127]	SVM	Compromised accounts	Based on anomaly score (1- online live frequency)	Twitter, Facebook
[128]	LDA	Spam detection	Accuracy —91.17% F1-measure - 78.43%	Youtube comments
[129]	CNN	Cyberstalking	–	Twitter
[131]	SVM, ANN	Cyberbullying	Accuracy-70.5%	Twitter

3.6. Epidemics

Epidemics in the physical world are frequent, but identifying it takes time, leading to many undesirable effects most of the time. So, for the quick response, it is the best we can analyze SM data, such as user's comments, posts, tweets, and so on. From that, we can get information and how to diagnose and identify the symptoms. In emergencies, the public receives panic; these panic/stress situations disturbs the governments. Identifying these panic/stress situations helps the governments to solve these situations quickly. There is another scary word we hear now is pandemic. A pandemic is a type of epidemic related to the wide-area spread and represents a disease that touches a whole country or the entire world. An epidemic becomes a pandemic when it spreads over notable terrestrial areas and affects many of the group. In short, a pandemic is an epidemic on a nationwide or global level.

Authors in [135] discussed the detection of influenza epidemics by analyzing Twitter tweets using machine learning techniques. They said they used multiple linear regression methods and simple linear regression techniques in their study. They mentioned their model got a 78% correlation with the Centers for Disease Control and Prevention(CDC). Authors in [136], presented a paper on COVID-19 pandemic. They discussed whether the weather influences the spread of the virus across the globe. For their study, they took data from Twitter and applied machine learning algorithms to it. Based on the short period tweet collection, they got the results as 40.4% registered uncertainty about weather's impact, 33.5% indicate no effect, and 26.1% indicate some effect. They concluded that considering the weather are misconceptions of people only most of the time.

Authors in [137], discussed how to Characterize the spread of situational messages in social media during the covid-19 epidemic. For that study, the authors collected data from Weibo, a Chinese social media app. They used SVM, NB, and RF machine learning algorithms to classify the collected data. Their main goal is to set quick identification and response to the situational epidemics. Authors in [138] discussed how to use Twitter to catch the flu based epidemics. They used an SVM classifier for their study and achieved a 0.97 correlation between Infection Disease Surveillance Center (IDSC), Japan, and Twitter data. Authors in [139] discussed how to use Twitter to predict the flu based epidemics. They used an auto-regression with exogenous inputs (ARX) model developed based on machine learning techniques for their study and achieved a 0.98 correlation between the data collected from Disease Control and Prevention (CDC) and Twitter. Authors in [140], discussed about a surveillance system framework that detects dengue outbreaks in Malaysia. They proposed Dengue Active Surveillance System (DASS) model by using machine learning techniques for their study.

Identifying root cause that participates in Epidemics gives awareness on information on the problem that is spreading, to the users of SM. For example, if some local people are discussing how some *x* virus is affecting their area and its dissemination capability. If any of their friends do not belong to their area, watch this information and warns other locality users by sharing that information. It makes the *x* virus information grabs the highest priority. It alerts the authorized people to stand-in against that. The analyzed data can wake up the public and government authorities to act prominently on these matters. Authors in [141], developed a unique bottom-up approach sentiment analysis for detecting latent infectious diseases. They used traditional unsupervised ML algorithms in a given location using user, textual, and temporal information in social media data. An unsupervised, primarily based methodology, is then presented to make symptom weight vectors for every individual. It is also provided a measure to support a person's sentiments and expressions. Finally, latent-infectious-disease-related information is known from individuals' symptom weight vectors. They valid their work by scrutiny the results that have taken from Twitter with the first electronic medical records (EMR). We can enhance their work by including notional approaches on how to progress and accuracy of identifying latent infectious diseases when considering SM user information (e.g., gender, age, posting frequency).

SM analysis is playing a significant role in detecting the specific systems successfully. Authors in [142], developed two types of a finite-state machine model, a) time-varying epidemic dynamics, and b) iterative local regression, which change concerning accuracy and complexity for de-individuation and anti-normative behavior in the online social network [143]. They verify the design with investigations, sorted by genre, and intimate how various epidemic patterns of action can be matched to specific interaction patterns among users for detecting abusive behavior. Authors in discussed the surveillance of influenza based on Twitter tweets. They said they were considered two seasons tweets for 2016–17 and 2017–18 in Italy. They correlated their work with the Italian InfluNet surveillance system. They said they show a strict association among the reports published on the InfluNet system and the tweets of Twitter users about their indications and fitness state. For this study, they used the ARX model on CDC and Twitter datasets. Authors in [144], presented a review paper on tracking pandemics using social networks. They reviewed papers from IEEE, ACM, Google Scholar, and Web of science. Authors in [145], provided the conceptual information using Sentiment analysis on Machine Learning technique to predict outbreaks and epidemics. Authors in [146], explained the usage of ML approaches to recognize potential persons for pre-exposure prophylaxis (PrEP) in healthcare environments in the USA and Denmark and a population-based research environment in Eastern Africa for HIV preventions. Although still in the concept

proof early stage, other applications include ML with mobile and social network data to advance real-time HIV risk minimization, virtual-reality agents to aid HIV serodisclosure, and chatbots for HIV guidance. ML has also been inherited as a common channel in HIV prevention considerations. Authors in [147], proposed a study that aims to get Twitter users' discussion and emotional reactions to COVID-19. They adopted the Latent Dirichlet Allocation (LDA) machine-learning technique to analyze about 1.9 million English language Tweets associated with coronavirus gathered from January 23 to March 7, 2020. For their study, they identified 11 topics and then categorized those topics into ten themes, such as updates about confirmed cases, COVID-19 related death, economic impact, Diamond Princess cruise, Preventive measures, cases outside China, COVID-19 outbreak in South Korea, early signs of the outbreak in New York, authorities, and supply chain. Their Results did not expose remedies and signs related messages as common topics on Twitter. Their analysis reveals the fear of the people for the coronavirus is predominant in all topics. Authors in [148], proposed a needful tweet analysis platform, which helps governments and municipal authorities to understand citizens' actual affective needs during the pandemic. Their platform nearly had four parts: data acquisition, storage, analysis, and visualization of data. It helps to identify users' needs based on their queries. They said they implemented this work based on the tweets collected from New York state to help the authorities during the pandemic time to study people's response in New York State. For their platform implementation, they adopted a machine learning technique SVM for user tweet classification. They said they used the SVM because it gives the best scores with it.

Hence, the rise in power of SM raises the availability of data to the researchers, particularly in the field of epidemics, which helps to analyze information for identifying the disease that occurred geographically. Analyzing social media data by epidemic researchers will establish geo-specific systems to administrate the outbreaks to the authorities. SM analysis for epidemics related information sometimes gives solutions to deal with the disease. For example, suppose the occurrences of a particular drug appeared in users message cured that disease many times. In that case, authorities also suggest the people and physicians use that drug while dealing with such victims. SM analysis helps in drug discovery, discovering the outbreaks, providing rationalized and clinical solutions, and identifying the root causes of epidemics.

Apart from the advantages, it also has limitations while using social media data for consideration in outbreaks. The backbone data for analysis come from the basic users and not from the experts, which leads to uncertainty in believing the data results. Though detecting outbreaks from SM data is cost-effective compared to the surveillance system, the surveillance system is the most trustable data corresponding to social media information. The Summary of Epidemics using machine learning for SM are given in a Table 6.

3.7. Event detection

Social media is the best biggest stage to share our thoughts. We can quickly broadcast the outbreak information in SM without extra efforts. We use SM analysis information to alert the people while detection of any events such as natural disasters or any dangerous disease disseminating in the country. Detecting events from social media could be used to disaster detection and management, traffic, pandemic outbreak, and news or topic detection. Social media is the best input for immediate crisis management resources. It has received much attention from academic researchers and practitioners to analyze event detection from SM data. This paper presents essential insights existing in the event detection area of research. In [150], authors applied machine learning techniques to real-time application disaster management. Geographical Information System (GIS) models and simulation capabilities are used to exercise response and recovery plans during the non-disaster time. By observing this information, someone will keep track of information and responses immediately at the time of disaster. The authors developed a participatory sensing-based model for mining spatial information of urban emergency events.

Getting up-to-date traffic report seeks traffic monitors information on roads and traffic authorities report. It delays the information posting to the user. Using SM, we can get traffic status by analyzing the people's comments regarding traffic status through ML approaches. In [151], authors described how effectively users could get active traffic jamming reports directly from traffic police using social media, so the users avoid those roads to escape inconvenience. Authors used the Linguistic Dynamic system's theories, fuzzy comprehension evaluation, the time-varying universe, and social transportation for dealing with the traffic of China's Shenzhen city. Their traffic analysis not fixed to a single flow factor; they considered three elements as a non-motorized vehicle, motor vehicle, and pedestrians. These assorted traffic flow gives a better result and to fabricate the supervision rule and strategy and evaluated miscellaneous traffic flows.

In [152], authors developed a real-time traffic detection using traffic-related events from Twitter. This service-oriented architecture system can fetch and classify Twitter tweets and informs the same to the users about traffic events due to an external objective like the cricket match, parade, and an exhibition or not. Their data extraction from twitter tweets just limited to the Italian language only, unlike others, practiced differently than the Italian language. For the best fit, authors adopted the SVM based classification system for real-time traffic events apprehension. This system detects traffic before online news websites and regional newspapers.

Spam detection is a promising technique which we discussed in Section 3.5, which means they have no prejudice concerning the incoming data, but almost all are not biased. In [153], the authors studied the performance, stability, and scalability of the ML algorithm on the datasets used for twitter spam detection. They measure nine popular machine learning algorithms and noticed that two decision tree-based algorithms — Random Forest and C5.0 gained solid performance in assorted states. The steadiness of each algorithm was deliberate by repeating the tests with dissimilar dimensions of datasets to inspect the performance oscillation, intending to discover a vigorous algorithm concerning performance. Their analysis confirmed that Boosted Logistic Regression (BLR) scored a relatively consistent detection performance in spite of the size of training samples, and tree-based algorithms such as random forest and C5.0 performed steadily under the situation where training samples are arbitrarily selected. They check the scalability of the algorithm analyzed by counting training time on while the CPU cores doubled. This observation gives them how deep learning they applied improved efficiency while they scale up CPU power. Link predictions, spam detection, and email filtering are several anomaly detection techniques that studies applied on SM through ML from ages. In [154], authors presented the state of the art research trends available for email filtering. They discussed what the different machine learning algorithms used by other researchers in email filtering are. The authors reported the trends, methodologies, performance of classifiers, challenges, and open issues available for the next generations who are interested. In [155], authors worked on link prediction technique. First, they started with identifying the Features of nodes, then utilized these node features to discover the model and algorithm. They combined both node-based features, co-clustering based features, local and global graph features, and described the Two-Level Learning algorithm to predict the links

Table 6
Summary of Epidemics using machine learning for SM.

Study	ML Used	Description	Performance	Datasets
[137]	SVM, NB, and RF	Identifying Epidemics	accuracies of SVM-0.54, NB-0.45, and RF-0.65	Weibo
[135]	Linear Regression, Multiple Regression	detection of influenza epidemics	accuracy-78%	Twitter
[138]	SVM, NB, and RF	detection of flu epidemics	0.97 correlation	Twitter and IDSC
[139]	ARX model	detection of flu epidemics	correlation-0.98	Twitter and CDC
[140]	DASS framework	Detecting dengue outbreak in Malaysia	–	–
[147]	LDA	Identifying user emotions on covid-19 pandemic	–	Twitter
[148]	SVM	Identifying users Needs during pandemic	accuracy-93%	Twitter
[149]	Regression	the surveillance of influenza	correlation-0.9	Twitter and Influnet, Italy

Table 7
Summary of Event detection on SM using ML.

Study	ML algorithm/ model/framework	Description	Dataset
[150]	GIS model	Disaster management using SM	Satellite images
[151]	Linguistic Dynamic system and fuzzy	Realtime traffic reporting	Twitter
[152]	SVM	Realtime traffic reporting by analyzing Italian language tweets.	Twitter
[153]	Boosted Logistic Regression, RF and C5.0	Spam detection	Twitter
[154]	NB, RF, DT and RF	Email or message filtering and its challenges	
[155]	RA, CRW, SRW, and AA	Link Prediction in Social Network	NetFlix, MovieLens, CiteSeer, Cora, WebKb, Conflict, and Protein-network
[156]	MLDS tool	Money laundering detection using SM	Twitter
[157]	SVM	Spam detection in SM	Sina Weibo
[158]	SVM	intusion detection	SCADA Network
[159]	NER and NEL model	Information extraction from SM	Twitter

available in the dataset. They worked with seven datasets as NetFlix, MovieLens, CiteSeer, Cora, WebKb, Conflict, and Protein-network for their developed link prediction algorithm. Their algorithm outperforms RA, CRW, SRW, and AA. In [160], authors developed an approach for the identification of social events in social media. The existing trend in the enhancement of hardware and technology leads to the usage of SM as handy.

Using the Event detection system, we can analyze the patterns participating in Money Launderings. It is another way to stop defaulters using SM. In [156], the authors presented a system to identify the money laundering using SM analysis. To detect such crimes, they used a tool called Money Laundering Detecting System (MLDS). The model presented in [156] builds a network using the bank statements and National Court Register. In conjunction with Clustering technique and frequent pattern mining in the SM analysis, they detected suspicious transactions and uncovered groups of offenders. So their automated criminal circle analysis and visualization identified the interaction 459 patterns of offenders and their roles in criminal circles. In [157], authors developed a new way of detecting spams that spread across social media. First, they collected data from Sina Weibo, one of the biggest social media platforms in China, using web crawlers. Such a web crawler gathers data of unauthorized users from SM. Then they apply SVM classification algorithms on it using their own constructed features. Lather, they explained how their system is very accurate and generating good performance compared to others. Their method for training the model takes more than one hour; we can use extreme machine learning to reduce the time taken to prepare the model; moreover, we can get more accuracy. They just classified the data only on ten existed features. We can use DL methodologies for automatic feature generation.

Customers and their published contents are critical ingredients to social media analysis. This analyzation gives information to improve the services offered by the companies. Some companies boost their product promotions using multiple social accounts, and biasing these accounts belongs to real customers who used these companies products and rated that product with all-stars. One solution to this is, identifying the users who are maintaining multiple accounts to publish the same information on SM. In [161], authors developed a solution to the real-time scenario of having multiple accounts in different social media, and 'me' edges is a technique used to define a solution to avoid misuse

of data with standard account credentials. It is challenging to maintain sensitive information not to reveal to all social media account. In [158], authors have taken a real-time application of power delivery infrastructure. Supervisory Control and Data Acquisition (SCADA) systems used by the electric utility industry to monitor and control electric power generation, transmission, and distribution, in which there is a possibility of real-time intrusion. To overcome such situations, a distributed intrusion detection system, IT-OCSVM, was developed.

In [159], the authors proposed Named entity recognition and Named entity linking approaches and its limitations for information extraction from Twitter. The authors revealed the cause of poor performance in Named Entity Recognition(NER) and Named Entity Link (NEL) methods annotation lacking in training microblog content. They identified that language identification, microblog-trained POS tagging, and normalization methods reduce the noise in NER and NEL performances. So we find that it would be a useful resource for initial training on web and forum content, although microblog-specific data will be critical.

So the researchers who are working on applications of SM analysis exposed another exciting application area of SM, such as event detection, which has been used to extract information to give next-level skills. Detecting events from social media could be applied to disaster detection [162] and management, traffic identification, pandemic outbreak, and news or topic detection. There might also be some event that has not been analyzed using social media yet. Other alternative methods may also be considered to improve the results. We discussed various event detection techniques in this section as a share of expressing SM intensity.

Social media report almost everything from daily life stories to the latest local and global events. This rich and continuous flow of social media data may lead to the difficulty of data overload. Such information is designated by substantial data challenges like volume (data quantity), velocity (the speed of development), noise, and variety (various natures of data). Consequently, mining such social media information has become a more challenging task than conventional text [163]. That is due to social media's dynamic features and the existence of both text content and network structure within the SM. Thus, data filtering, data analysis, and mainly detecting and monitoring social media text's exciting events have become the most complex task. Detecting the event from different events occurred like Planned, Unplanned, Breaking News, Local Events, and Entity Related is another challenging task that grabs significant attention from the researchers.

One extreme advantage of using social media data for event detection is that it reduces the overall maintenance costs because of no need to use sensors and physical equipment to detect the events. Apart from the cost and maintenance, there are still limitations in using the SM data to detect the event. The major limitation is with the accuracy of the detected event. False detection of the event to the power holders could lead to making wrong decisions. The performance measurements of methods used in event detection require high precision and extreme timelines. Hence, It concludes that reliable event detection is a pervasive problem. The summary table of event detection are presented in Table 7.

3.8. Image analysis

Since image analysis has immense application scope as SM and the web as an absolute and becomes added visual. It is not reliable for the brands to stay on text alone while analyzing SM information to learn from their customers properly. We can expect multimedia posts are predominant in SM. More than half of the uploading to the SM are images and videos only. Even text added to those multimedia posts is getting smaller and shorter, which creates curiosity for companies to understand those posts to analyze their customers. The internet is full of symbols that are continually evolving in a sense and being added too.

Recent advancements in machine learning and computer vision [164] have allowed image analysis [165,166] to be effective at scale. Image analysis acts as the dominant source of information while taking any crisis to relief agencies by adopting crisis maps with satellite images and images from the affected area social media users [167]. Disaster management can be made easy for the authorities by using social media available image analysis. In [168], the authors proposed a method to analyze people's stress levels in a crisis. The authors developed the crisis map for crisis detection, which they developed based on the social media user's information in the affected area. Authors quoted that these social media-assisted crisis maps can help the volunteers and disaster management agencies for better disaster control. They said that it could combine the ground-level images with satellite images to analyze both horizontal and vertical viewpoints and said it got outstanding results and scope to future technologies with multi-view scene applications.

Videos are a series of images. Video analysis on social media such as analyzing YouTube, Facebook, and Twitter videos are part of image analysis. Among enormous information, the extraction of user desired valid data is a challenging issue. In [169], authors pointed out the real-time problem encounter in the vast flow of videos in web portals. The authors developed a system that will take web users' interaction over videos in the SM as a community-driven web-video topic discovery model. The authors said their system would find the event in real-time by analyzing a series of collected videos from social media sites. In [124], the authors presented an SVM based Data-Driven Intelligent Traffic System (D2ITS) algorithm for taxi detection. They said D2ITS is an extension to the original, intelligent traffic systems (ITS) algorithm. It identifies the unlicensed taxi by analyzing the vehicles' images in the traffic collected from the social media sites. This image analysis includes vehicle type, color, number of the vehicle, model of the vehicle, and name of the manufacturer.

Many e-commerce companies try to deliver their products through drones for faster and safe delivery using image analysis. One more intelligent application of Image analysis is it can be useful for driver-less cars while traveling; they can process their 360-degree images and find the objects and act according to that. These driver-less cars require both satellite information for navigation and ground-level information like the present visual scene for avoiding impediments to traffic and accidents. Authors in [170], consolidate satellite image with the street level image for analyzing the scenes. They took satellite images from Google maps and street-level pictures from social media sites like Flicker, Facebook, and Instagram. Social media sites these days allow the user to upload photos with Geo-tagging and also 360-degree visualization as providing a virtual tour to the user. This multi-view scene analysis not just combining and classifying the pictures; they also try to detect few objects in the scene. They employed machine learning algorithms to identify the objects and classifying them. Applied algorithms in it are CNN, CRF, and k-NN.

Does anyone has the doubt, how Facebook knew our friends from uploaded selfie?. It is because they are applying image analysis on our photograph and compared it with the information they had using ML techniques on SM data. Individual Face recognition from a group photo is fun and had lots of machine learning techniques are applied to it, to complete the task. For example, in [171], the authors implemented two methods to recognize similar faces. They evaluated different shapes of faces like Eigenface [16], Gabor wavelets [1], Local Binary Patterns,[2]

[2] https://www.statista.com/statistics/264810/number-of-monthly-active-Facebook-users-worldwide/.

and Fisherface to obtain their effort done more adequately. Their unique methods recognize the human face even in many unimaginable conditions too. Their approaches recognize similar images, even in wild circumstances such as dark to extra light conditions. For their study authors used PubFig83 datasets for testing their work. The second method they used are User-specific application of partial least squares (PS-PLS) method used to gather data from social media to identify similar face groups. They applied SVM for avoiding outlier and pattern recognition.

In another scenario, authors in [172], developed a framework called deep aging face verification (DAFV), which used to identify individuals of any age photograph on SM. They constructed a cross-age face dataset starting from the kid, to youthful, to mature, to aged groups as input to DAFV. This framework discovers and cluster images of the same person with different ages. They applied machine learning algorithms such as CNN for their DAFV. They compare their DAFV with FG-NET, MORPH, and CACD to expose their framework is best. Authors in [173], developed a cascaded search method suitable for the big library of faces like Instagram. Authors developed an approach that worked on CASIA, a publicly available dataset to represent the face using deep learning approaches. These deep learning features are utilized in recognition of user faces bases on the KNN algorithm. Authors used COTS face matcher framework, it again re-ranked the recognized faces to find the best relational measures. They evaluate the aimed search scheme on an 80 million face gallery and show that the intended scheme grants fair trade-off between recognition-efficiency and run-time.

Most of the recommendations in SM happen because of analyzing user-uploaded images. Every user has a specific interest. Researchers are developing robust recommendation systems using machine learning on SM [174,175]. Authors in [176], proposed location-oriented clothes recommendations based on images uploaded by the users on social media. They developed a hybrid multi-label convolutional neural network combined with the support vector machine (mCNN-SVM) for both garment selection and area-specific preference information. They collected the data from three travel inspired social media sites, such as tripadvisor.com, mafengwo.com, and chanyouji.com. The attributes they selected for this study as 'Climate: Season,' 'Climate: Weather,' 'Attraction: Type,' and 'Attraction: Color', for area-based cloths recommendation. They trained there model with three fashion datasets (i.e., Fashionista, CCP, Colorful-Fashion) and best results for this study.

One hurdle for the images gathered from social media for analysis is feature identification and selection. [177]. Applying deep learning on features gives automatic feature selection [178]. There are many approaches for this; Authors in [179], described the SM image analysis through an exceptionally designed deep latent feature learning system that learns the feature itself. They claimed that their work can greatly help to the researchers for feature extraction and selection. In [180], authors developed extreme feature learning for automatic feature identification to utilize in classifications. For this extreme feature learning, they applied images with word combined and processed it called collective intelligence that extracted from social media images embedded with the word. They achieved collective intelligence based on SBU(Stony Brook University) flickr Image-caption labeled dataset with evaluated SBU learned features. It provides the best learning among text and pictures using a deep learning approach [181] applied to millions of labeled SBU datasets. Then, they introduced a two-step visual deep learning method to draft images into the embeddings. As a result, they can extract features that can cross modalities. They presented a consolidated approach for DL latent feature learning framework in SM. For this, they design a framework called Relational Generative Deep Belief Nets (RGDBN) model and employ it on the analysis of links. They confirmed that it captures both heterogeneous and homogeneous data. They described the learned latent feature is more delegate in user recommendation and image annotation jobs. Moreover, they create a lambda rank model toward deep multi-modal learning for image retrieval within the latent feature learning framework.

Furnishing travelers with tourism information in social media has gained dominant consideration from tourism researchers. However, very little practical evidence also increases the relationship between tourism quality factors and end image formation in social media [182]. Authors in [183], discussed a study that provides a more in-depth insight into the elements of tourism data leading the development of tourism destination image in social media. Their study collected data on Sina Weibo social media application and applied ML algorithms to it. For analyzing the data, they used IBM SPSS Statistics 20.0 and SmartPLS 3.0. By examining and researching tourism information quality factors in social media, authors found value-added, relevancy, completeness, interestingness, and web page design are tourism information quality factors affecting travelers' destination image formation. They said their study of social media's role in the tourism industry and information quality factors increase user's knowledge about the impacts of information quality factors and website design on target image formation by visitors, presenting changeable perspicacity to tourism marketers.

Hence, with limited image analysis usage, brands are juggling out on a massive piece of social media discussions about their products, competitors, customers, and brands. The improved usage of images leads to the development of image analytics. The improved image analytics helps researchers give results quicker than the traditional approach [184]. The SM image analysis is also used to track peoples' feelings to find the right place of advertisements necessary to the companies. So the image analysis urges the designers to design their products visually pleasing and influencing. Image analysis is not just limited to developing the business, and it has other meaningful utilization such as in public health, emergency, recommendation systems, automatic feature extraction, facial recognition, pattern recognition [27], crisis mapping, and more. All these applications' scope on SM's image analysis pushes the researchers to work on this trending topic. We presented the summary table of image analysis using SM data, in Table 8.

3.9. Recommendations

social media gives the best suggestions using Recommendations or Recommender systems [185]. The recommendation system uses machine learning algorithms on the datasets provided to extract the best Recommendations to the user, according to user interest [186]. Recommender systems work on the principle, which identifies the users with specific interests and suggesting the products to those particular users based on their concerns [174]. For constructing such a Recommender system, it requires legitimate data about products and their information that will be available to the qualified user. Detecting skilled users for the Recommender system is a challenging task for the social media analysts. Trusted users enhance the performance of Recommender systems [187]. A trusted user is a regular user who has a high trust factor among all other users in their community. Detecting Trusted users for the Recommender system achieve through machine learning concepts [188]. Recommender systems sometimes helps the users to improve their behaviors while interacting in social media. The Graphical abstraction of Recommender system using SM given in Fig. 12.

Table 8
Summary of image analysis using machine learning for SM.

Study	ML Used / Model used / used Framework	Description	Performance	Datasets
[168]	TRIDEC project	Real time crisis mapping of natural disasters using social media.	precision-0.78 recall-0.76 F1-0.77	Twitter, Google Earth
[171]	SVM, LDA	Generating person-specific representations used to boost face recognition performance	Aligned image filters – 92.28	PubFig83
[172]	Deep aging face verification (DAFV) framework, CNN	Deep Aging Face Verification with Large Gaps	Accuracy – 0.7739	Cross-age face (CAFE)
[173]	CNN, k-NN	Face recognition through COTS framework using SM.	Mugshot (93.7%vs. 98.6% TAR@FAR of 0.01%), performance (99.5% TAR@FAR of 0.01%), accuracy −98.20%	Mugshot
[176]	SVM	Location oriented clothing recommendation	Accuracy-0.8013, 0.8969, and 0.7405 on three datasets	Fashionista, Colorful-Fashion, CCP
[179]	k-NN	Developing latent feature learning	AUC- 0.7	MSR-Bing Image Retrieval Challenge dataset
[180]	CNN	Learning from Collective Intelligence like Feature Learning Using Social Images and Tags	Perfor-mance(mAP%) of classification: NUSWIDE-42.9, CVV-67.8	SBU, NUSWIDE, CCV, Flickr30K, and MEN
[183]	IBM SPSS Statistics 20.0 model and SmartPLS 3.0 model	Identifying tourism quality factors for website design	correlation 0.9	Sina Weibo

Authors in [89] developed a concept, which discovers the group of peoples who will get excited and fancy to share interesting content. They introduced a model called the influential activity maximization model, which finds the preeminent and trusted user from the available online users. They use this superior user to spread their information. This activity maximization model uses the NP-hard's diffusion models, such as the independent cascade (IC) model and the linear threshold (LT) model.

In [189], authors introduced a new way of a Recommender System using Collaborative filtering. They applied a collaborative filtering technique by coupling the sparse rating data and cold-start user's rating. They developed a Recommendations framework to produce trusted users by measuring their ratings. This framework has trustier, and the other is a trustee list. Authors not just using trusted user, they also predict the user as trusted by considering some trustees to be particular trustier. This gives more precise information for generating a trusted user list for making a Recommender system. They applied matrix factorization methods on these lists. Authors worked on are Epinions, Douban, and Flixster datasets to develop their model. Checked their matrix factorization methods using numerous performance measures techniques like a MAP, SVD++. , precision, F1-score. Their developed TrustPMF model was adequate to present a probabilistic view for understanding the Truster, Trustee, and TrustMF models, and it also improved Recommender value by fusing both browsing and review references in an amenable way. We can increase the trust of the user by combining both positive and negative comments while classifying and measuring data performance to get a best-trusted user for product Recommendations.

In [122], authors developed a new data processing concept called Decision learning. Decision learning is the combination of leaning the data and making a decision from that learned data. For data leaning, they use SM user behavioral study data. This behavioral study includes the study of sequential behavior for data sharing and evolutionary behavior for deal recommends to users for gaining more. Their Recommender system used to discover the restaurant deals on products to benefit from regular billings. In [190], authors examined issues in claiming social client quality induction by exploring the core relations of client qualities that are extracting both lexical and visual characteristics of internet user-generated substance. They efficiently consider six sorts of client attributes like gender, age, relationship, occupation, interest, and emotional orientation. In perspective of methodology, they developed a relational latent SVM (LSVM) model by consolidating a rich set from claiming client features, quality inference. Furthermore, quality relations in the framework. Their investigations gathered dataset starting with Google+ with full quality of the recommended approach in user attribute inference and attribute-based user recovery.

Trusted users are sometimes referred to as best users for Recommender systems. In [191], authors presented a Recommender system. This Recommender system identifies the best user based on the name of the online user given in the SM. In that article, they collected these names of the users by python based web crawling program across three sites, namely Facebook, Twitter, and Foursquare. Then they classify the data based on their usernames redundant feature. They developed ten features to classify this data and applied LRCV, a machine learning method, to identify the best user. Authors in [192] developed a social regularization method that involves social media data to avail to Recommender systems. For predicting missing values in the user–item matrix, they applied ratings and the number of friends among users. These two aspects of social media data are utilized to construct social regularization terms. To measure the similarities, they clustered the datasets with the similar favored friend's group.

In another case, authors in [193] presented the Recommender system to identify rising stars (best user) in the co-author network using machine learning techniques. In that, they explored

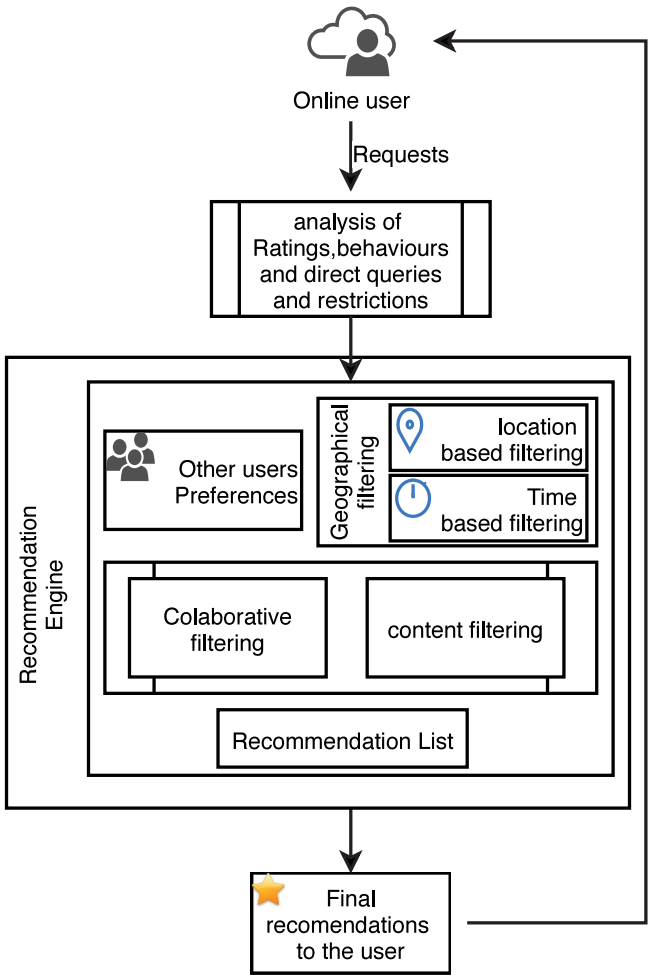

Fig. 12. Graphical abstraction of Recommender system using SM.

three classes of features, i.e., Author, Co-author and Venue type observed most effective. To analyze the results, they tested the ML-based Maximum Entropy Markov Model (MEMM), CART, Naïve Bayes, and Bayesian network algorithms. Two varieties of data sets are constructed for measuring the system. This performance measure using total citations and an average relative increase in citations. They identified MEMM resulted as the best as compared to other algorithms for total citations, and CART performs better as compared to different algorithms indicated by the authors for an average relative increase in citations.

In most of the cases in this modern era, one should depend on social media to purchase any product. Hence the incremented features of social media make more availability of information to the users. For example, Authors in [194] presented the survey of Facebook for the attainability of foreseeing increments in Facebook utilization. They figure out which predictors are essential. They standardized the execution of ML algorithms such as LR, RF, Stochastic Adaptive Boosting, Kernel Factory, NN, and SVM utilizing five times twofold cross-validation. Stochastic Adaptive Boosting with cross-validation gave the best performance. Most Facebook predictors forget to consider, like frequencies, group annotations, photo privacy, usage of templates, recency in comments. Facebook and other SM sites alike could utilize predictions of increments in the use of recurrence to redo its services. For example, pacing the rate of promotions and friend Recommendations, or adjusting News Feed content out and out are few identified services. The fundamental commitment of this investigation is to survey the forecast of increments in utilization recurrence in an interpersonal organization. They worked on the data which they had in 2014; we can improve with more accessible Facebook data and can get better results.

In indigent situations, we expect some people to help to overcome the situation decently. These people are called volunteers. This principle also applies to social media too. Detecting volunteer users for job completion from the existed users requires ML techniques. Authors in [195], presented a scheme to infer users' volunteerism tendency based on user-generated content (UGC) from multiple social networks, using a binary classification problem. As per the conceptual volunteer decision model, they measured users' volunteerism tendency by the user- and network-centric analysis in different ways. From the user's social data, they established and received a set of application-centric features for user-centric analysis. To represent the relationship among social users, they use a graph-based model for the network-centric analysis. Finally, they learned and developed some strategies on how to collect user's data from social media. Further, they explained the expansion and categorization of the system that identifies volunteerism among the users using the AQV dataset. This volunteerism tendency was very useful in developing applications such as age group prediction.

Recommender systems have different shades like authors in [196], described the prediction of negative links on SM. They developed a framework NeLP based on machine learning classification algorithm SVM to detects negative links. This negative link prediction was very useful to generate positive link prediction and recommender systems too. In [197], the authors defined a solution to the problem of constructing new cities. As population as well as commercial needs growing day by day, it is mandatory to identify the comfortable living location by collecting moments of citizens, as well as available transportation facilities and environmental conditions. The information collected, consolidated, arranged in a hierarchical structure, and cross-checked with the help of implementing online volunteered geographic information.

Apart from the discussed solutions to various problems of Recommender systems, it has few issues too. Few such issues are cold-start, gray sheep, early-rater, sparse evaluations, serendipity, and scalability [198]. In this paper, we presented different Recommender systems concerning its participation by utilizing SM. We noticed that the data related to users of SM influence the Recommender system. It forces the researchers to develop tools to assist the users of SM in handling the data they receive. So recommender system is another crucial application in the study of social media analysis. The summary table of Recommender systems using SM presented in Table 9.

3.10. Relationships and reputations

Using social media, we can improve our relations and reputations like increasing online friends list, making us popular in the virtual world. The increased friends can easily publicize or boost your ideas or business [199]. Authors in [200] did research on the same ecological people and how they communicate and what is their relationships with others in the group. They used Machine learning algorithms and applied it to their social media data grabbed from twitter. The authors explained the performances of three classifiers, which are SVM, Naïve Bayes, and Standard LR. They also test the performance of the classifiers using SM contents. They assessed it using critical factors like accuracy, recall, precision, the area under the precision–recall curve, and Krippendorff's Alpha. We can enhance this using more sophisticated trained datasets for diversified results.

Online community detection gives many solutions to real-world problems, and we can understand the dynamic topology

Table 9
Summary of machine learning algorithms for Recommenders systems.

Study	ML used / model / framework	Sentiment description or prediction	Performance	Datasets
[89]	Independent cascade (IC) model and the linear threshold (LT) model	Improving information diffusion in SM	–	Douban, AMiner, DBLP, and LiveJournal
[122]	Chinese-Restaurant-Game framework	Data analytic learning with strategic decision making	–	Twitter hashtag, Memetracker, and Yelp
[189]	SVM, Matrix Factorization	Constructing recommender systems using Collaborative Filtering by Trust	Epinion: NDCG-0.9371, Douban: NDCG-0,9388	Epinions, Douban and Flixster
[190]	Relational LSVM	Attribute based user retrieval	Age : 0.727, Gender: 0.798, Relationship: 0.624, Occupation : 0.25, Interest : 0.617, Emotional Orientation: 0.41	Google+
[192]	Biclustering Algorithm, SVM	Identifying the most suitable group of friends for generating different final recommendations.	Accuracy-16.73	Del.icio.us
[193]	MEMM, CART, Naïve Bayes and Bayesian network	Predicting future rising stars is to find brilliant scholars/researchers in co-author networks	–	Digital Bibliography and Library Project (DBLP) and Arnetminer databases
[194]	Stochastic Adaptive Boosting	Predicting Increases in Facebook Usage Frequency	cross-validated AUC - 0.66 and accuracy - 0.74	Facebook
[195]	SVM, RF, Gradient Boosted Regression Tree (GBRT)	Identifying volunteers to help in need using SM	Follow: F1-measure-71.75, Hashtag: Recall-94.39, Follow: Precision-63.25	Facebook, Twitter, Quora, LinkedIn
[196]	NeLP Framework	Negative link prediction in SM	Epinions: F1-measure-0.324 and Precision-0.286, Slashdot: F1-measure-0.2441 and precision-0.213	SlashDot, Epinions

of networks [201]. Traditionally most of the approaches follow undirected graph methodologies for identifying the user community. Authors in [84] developed a new concept "sentiment communities" for improving the online marketing's and business strategies by finding a stable relationship between users. Sentiment communities are users group who have an independent mindset on particular products or services on SM. In [202], the authors developed a new methodology for detecting the online community in social media networks. They used a standard probability graph and clustering coefficient together and developed a new vector influence based clustering coefficient for measuring the directed graphs. They said their Triangle method is a new way in the presumption of community detection.

Companies plot much interest in trusted users as their brand ambassadors to promote their products. If they are not legitimate users and spreading rumors about product information, it can spoil the company's reputation in the market [203,204]. To overcome this situation, we can detect rumors using this social media analysis through ML techniques. For example, in [205], the authors developed two new approaches to pin down the rumors happen at social media networks. Their self-possessed methodologies include one is to stop rumors at an essential user or to clarify the truth behind the rumor. Authors in [206] discussed a Synthetic Minority oversampling Technique Boost (SMOTE Boost) algorithm to compare two types of features such as raw reputation and probabilistic reputation features. Developed probabilistic reputation feature gave the best results for identifying the trusted user in SM. They tested these methods on the datasets taken from Wiki, Epinions, and Slashdot.

Reputations and relations in an SM identified through an influential degree. Influential degrees are constructed on users of social media. Based on their influential degree, companies selected them to influence other users for promoting their products. Influential degree identification in social media can leads to relational intelligence. Relational intelligence used to regulate emotions, values, interest and demands between each other in an online social media [207]. In [208], the authors developed two methods, Observed Aspect Influence Model (OAIM) and Latent Aspect Influence Model (LAIM), to find the aspect level influenced graphs. Graphs can be used in many applications like visualizing the topology of the network, data communication in the social web, personal node communication among them. Some graphs nodes persuade other nodes. This persuasion can be considered based

on their persuasion strength using persuasion degrees (influential degree). The authors developed OAIM will identify the influential degree. Their LAIM will deal with aspect level influential nodes in the graphs.

To enhance Relations and Reputations, we can use social media analytics [209]. In [210], the authors developed a different application crowd++ to detect the number of speakers talking in a meeting called crowdsensing. The first recorded the speech of people and processed it to propose the crowdsensing. Their used datasets contain a different correlated audio file for classification. Crowd ++ is a study application of wireless sensor networks using the Mel frequency cepstral coefficient via speech processing. Using this application, they represented the total number of people talking in a particular area and adjusted devices according to the noise level without adding extra infrastructure amenities. Their work saves not only the infrastructure cost but also time.

Network failure anomaly causes a reduction in information measure to the users of social media. Authors in [70], presented their article on root cause analysis to detect network failures by applying machine learning techniques. They discussed statistical methods for root cause analysis, such as applications of random forests for diagnosis and training and study of system properties. They applied different methodologies to find errors that lead to network failures. Failures among nodes give weak relationships. Detecting causes to eliminate this weak link using root cause analysis. In [211], the authors main motto is to identify the factors to get more popularity and togetherness in social events. A social gathering will mainly depend on reliable information exchange. The authors developed an algorithm for SM user groups to find out the uniqueness and reliability of other participants.

Social media has evolved to become an indomitable beast. Although today's modern public relations teachers could successfully twist things in the media, today, a single act of malfeasance will attract millions of views in minutes, creating an Instagram and Twitter trial for a company or user. Suppose weak popularity handling can lead to adverse reports exhibiting on social networks and rating websites. In that case, it can influence a companies image on websites, affect the firm's share prices, and hit its place inside Google's search engine's results pages. For persons, it degrades their credibility on social media platforms. Hence, it is observed that reputation and relations also play a crucial role, and companies requesting the researchers to develop new methods for its extended serviceability.

3.11. Sentiment, opinion and emotions analysis

Sentiment analysis is the manner of predicting the emotions, opinions, attitudes, and feelings expressed in the text data [212]. It can be used to classify subjective statements as positive, negative, or neutral to determine opinions or sentiment about a subject. Machine learning techniques are used in sentiment analysis to develop models that can predict sentiment in parts of the text [213,214]. The authors in [215], articulated their work on discovering sentiment analysis in dynamic events because the analyzing data are volatile so that output result may change during the time being. They have analyzed the Twitter data for predicting the results of elections in the USA. They gathered ten thousand labeled tweets, and they segregated them for five candidates named Bernie Sanders, Donald Trump, Hillary Clinton, John Kasich, and Ted Cruz. These segregation's are performed based on budget, finance, education [216], energy, environment, healthcare, immigration, gun control, and civil liberties issues. They avoid re-tweets and also checked tweet similarity using Levenshtein distance.

Sentiment Analysis can be described in another way as contextual mining of content which recognizes and extricates subjective data in the source material and helping business to comprehend the common supposition of the brand, item or administration while checking on the web discussions. It is similar to merely touching the most superficial layer and passing up an excellent opportunity for those high esteem bits of knowledge that are holding up to be found. Authors in recorded general tasks of affective computing and sentiment analysis and presented a broad categorization for them. Their research focused only on finding the credibility of the user based on the category of bullying he/she does. It does not reveal any status of the relationship between the users. The A graphical abstraction of Sentiment or opinion analysis from SM given in Fig. 13.

Predicting the outcome using Sentiment Analysis from social media generated data are the best practice for the researchers through Machine Learning. There are many prediction systems available in SM using Sentiment Analysis. In [217], authors predicted the outcome of Elections happen in India and Pakistan in 2014 and 2013. They predicted these results using Twitter tweets specific to Indian elections called India Election Tweet database (IET-DB). They applied machine learning concepts such as opinion polling and text mining strategies to get the results. Authors in [218] presented a report on the Link Prediction (LP) technique exercising in SM. They surveyed ten years of papers, starting from 2000–2014, collected from renowned international journals databases. Their article is handy to the researchers who have thought in mind to begin work applying LP. They explained algorithms utilized in LP, and then they illustrate recognized methods available for LP. Authors also acknowledged good articles and challenges of link prediction too.

We are presenting multiple sentiment analysis systems through Machine Learning techniques. This Sentiment Analysis we are mentioning here uses different approaches and methods using Machine Learning by numerous authors. Every author used a specific way to achieve the intended study — Twitter, a microblogging site that provides rich datasets to the researchers to research multiple aspects. Some applications are finding useful like sarcasm detection [219], sentiment evaluation, user mental health detection and more. For example, In [220], the authors presented the sentiment analysis on Twitter. Semantic evaluation forum shortly SemEval works on Twitter data to analyze the sentiment factors. Their primary task is to gather the tweets, messages, LiveJournal messages, and sarcastic tweets. The SemEval is a group of 44 to 46 members. They divide their tasks as collecting the data, creating datasets, applying classification techniques such as SVM, Maximum Entropy, and Naïve Bayes and performance measurements of classifiers using methodologies like F1-measures. They worked only on textual data generated across Twitter and LiveJournal. If we want more precised sentiment analysis, we can consider the processing of images too.

Authors in [221], proposed a new model for detecting sarcasm. They used a two-level memory network for sarcasm detection. In the first level, they captured the sentiment semantics, and in the second level, they took the difference between sentiment semantic and sentence situation. They also presented the effectiveness of their proposed model In their paper. Authors in [222] presented a work that describes a classification schema for irony detection in Greek political tweets.

The diversified data of the Twitter dataset will be challenging if it is not labeled. Labeled datasets finish the job quickly. By utilizing the Twitter dataset, we can identify the spams. In [223], the authors developed a novel approach called Lfun(Learning From Unlabeled tweets) approach to solving the twitter spam drifting. The Twitter spam drifting is the expiration of the twitter spam data set. This expiration means outdated spam data set. In that, the authors compared the labeled spam tweets over unlabeled non-spam tweets. They classify new spam tweets using machine

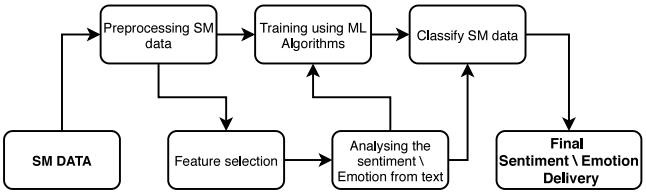

Fig. 13. A graphical abstraction of Sentiment or opinion analysis from SM.

learning classifiers algorithms. This new unlabeled spam tweet updates the twitter dataset so one can find the spam on Twitter very quickly. Authors said their Lfun(Learning From Unlabeled tweets) can detect spam using account age, some followers or friends, and the number of characters in a tweet. And they also said their spam detection method is dynamic.

In [224], authors developed a novel approach for sentimental analysis using the twitter dataset. They classified the data into seven classes, but their approach works with more or fewer classes too. They used a tool called SENTA (Sentiment analysis), which identifies the word and group them according to the labels to the tweets. Their sentimental analysis can work with other social data sets too. They first preprocessed the data and classified the data into negative tweets or positive tweets. This approach is called a binary classification approach. Their approach gives the best accuracy while using binary or ternary classification. In the future, they like to work with the ternary classification for sentiment analysis in Twitter data.

It is not possible to share all kinds of content in all groups. Every group has specific intentions like kids group talk about rhymes, youth, or other age people do not show much interest in these. So identifying the perfect group for appropriate information share is also a concern in SM. In [225], the authors projected an approach to identify the age groups of online users in social media. Some online sites never ask about age otherwise, users individually not interested in confirming their age. But using machine learning algorithms, we can classify the teenagers and adult age group users. This classification happens based on the use of punctuation which includes the emoticons, the number of letters, or the number of lines, slang, URL, fans, supporters, the total number of tweets posted and approached topics are appropriate to boost the boldness and exactness for classifying the age group. In this method, the authors use Deep Convolutional Neural Networks(DCNN) for classification and sentiment analysis. This DCNN gives the best results over many other algorithms. We summarized the sentiment analysis of SM using ML in Table 10.

Sentiment analysis also used to identify the essential users in SM. For instance, in [232], the authors developed a newfangled method Information Flow Model(IFM) enhanced from the Barabasi–Albert model (BA) and GMG. IFM used to describe the properties of SM to organize the beliefs. IFM controls the beliefs of SM without using additional overheads. Beliefs occur because of an essential user, who was identified using their essential user degree. So controlling essential users can limit the dissemination of beliefs. In [233], authors presented different methods and algorithms to examine corporations' informal social networks using social media. They grabbed the data from an organization's specific Facebook posts using a self-designed web crawler. They employed machine learning techniques to recognize community detection and measures its performance over the group. Their article gave an insight into the organization's structure and communication between the network of each company. It demonstrates to identify the most influential leaders in the community. Their research helps the organizations to expose their strengths and weaknesses in their organizational structure and corrected potential problem areas also.

In [234], authors applied natural language processing to identify a sentimental analysis constructed using user reviews is not the exact message they convey. In [235], authors considered YouTube and Flicker sentiment analysis to determine the precise choice of interest. Traditionally we all follow the concept of One class collaborative filtering (OCCF) method to analyze the users choice of usage. That traditional way cannot provide exact results as a customer choice are diverse. To overcome such situations, authors developed a new algorithm as Sentiment-aware one-class collaborative filtering as SA_OCCF to handle the challenge.

Text extraction leads to analyze the sentiment in groups. In [226], the authors developed a sentiment analysis methodology by applying the polarity level to the sentences called SentiNet(sentiment network). According to the polarity level of each phrase taken from the data-sets, they analyzed the sentiment of a data. They took a few data-sets and compared these data-sets with machine learning classifier algorithms along with their advanced methods. Authors in [236] discussed an automated program that designed to answer accordingly to the human conversation with it called bot. Business personalities are promoting their business using these intelligently developed bots. They also presented twitter bombs, faulty information spreading, and excluding minimal information verification confers the tricky face of Bots. They delivered multiple vulnerabilities of bots in that article and depicted them as the master of puppets.

Sentiment analysis not just used to predict behavior but also used as a tool for Stock market prediction [227,237]. In [227], the authors proposed a sentiment analysis using social media data for stock market prediction. They collected 18 message boards from yahoo finance for predicting the sentiment. For sentiment analysis, they combined the only topic based and the only sentiment-based and proposed joint sentiment-topic (JST) model. They compare JST with SVM and LDA machine learning algorithms to show how effective this JST for extracting sentiment. Authors failed to represent a non parametric topic model that can gather more sentiments and topics and do the automatic extraction of sentiments for stock market prediction. In this model, the prediction was made using two factors, such as past data and sentiments from social media, and it is a limitation. This limitation can overcome by applying more factors like real-time data, stocks co-variance, macroeconomic indicators, and companies' financial status gives a good stock prediction model.

In [238], authors intended to include crowdsourcing methods for mobile sensing and offer several crowdsourcing methods for assessing the weight of trust among agents. The direct and indirect mobile sensing network models are reviewed. A model of social-network-inspired mobile multimedia and sensing applications is explained toward the unification of social network and mobile sensing. Numerical investigations on real-world datasets show that mobile sensing can help from crowdsourcing methods for performance enhancements. In [228], authors presented the article on sentiment analysis using twitter tweet data. For analyzing the data, they applied the SVM algorithm. Authors combined both text mining and conversations to deliver an adequate system. They interpreted the sentiment as tension if their model shows the spike in the data visualization, which means there would be tension occurred in the group. Their tension representation includes text mining, opinion mining, and sentiment detection. They applied Cardiff Online Social Media Observatory (COSMOS) to gather twitter information for achieving their feat.

In [239], the authors have presented a Defense Advanced Research Projects Agency(DARPA) Twitter bot challenge competition approaches. The competition conducted for detecting influential bots on Twitter. In their report, they submitted the methodologies and techniques applied by the top three winners in the contest. Authors explained different bot detection approaches

Table 10
Summary of Sentiment Analysis over SM using ML.

Study	ML used	Sentiment description or prediction	Evaluation	Datasets
[215]	SVM	Challenges of sentiment analysis for dynamic events	precision - 0.66%, recall - 0.63%, F-measure - 0.64%.	Twitter.
[220]	SVM, Naïve Bayes	Semantic evaluation using Sentiment evaluation and NLP	–	Twitter.
[224]	SENTA framework	Sentiment Analysis in Twitter.	Accuracy - 81.3%	Twitter.
[225]	CNN	Age Groups Classification in Social Network Using Deep Learning	F-Measure - 0.94% Precision - 0.95%	Twitter.
[226]	SenticNet framework	Sentiment Data Flow Analysis by Means of Dynamic Linguistic Patterns	accuracy - 71.32%	Movie review[a], Blitzer Dataset, and Amazon product review dataset.
[227]	SVM, LDA	Stock market Prediction through sentiment	Accuracy - 0.544%	Twitter
[228]	SVM	Forecasting spikes in social tension	Precision - 0.74% Recall - 0.54% F-Measure - 0.63% Accuracy - 0.94%	Twitter
[229]	SVM	Multimodal information extraction to infers and aggregates the semantic and affective information	Accuracy - 87.95%	eNTERFACE dataset
[230]	SmartSA Algorithm	Opinion mining	Twitter: F1-measure - 77.22%, Digg: F1-measure - 67.85%, and Myspace: F1-measure −72.94	Twitter dataset, Myspace, and Digg datasets
[231]	LDA	Sentiment detection in SM	–	Twitter and Facebook

[a] http://rottentomatos.com.

Table 11
Critical datasets adopted by papers cited in our survey.

No.	Data Name	Dataset URL	Content
1.	The MySpace	http://arnetminer.org/lab-datasets/multi-sns/myspace.tar.gz	854,498 user profiles as well as 6,489,736 directed connections
2.	NetFlix	https://www.kaggle.com/netflix-inc/netflix-prize-data	2,649,429 users and 17,770 movies ratings
3.	MovieLens	https://datahub.io/dataset/movielens	6040 users and 3952 movies
4.	CiteSeer		3312 scientific publications and the citation network consists of 4732 links
5.	Cora	https://relational.fit.cvut.cz/dataset/CORA	2708 scientific publications and the citation network consists of 5429 links
6.	WebKb	www.cs.cmu.edu/afs/cs/project/theo-20/www/data/	877 scientific publications and the citation network consists of 1608 links
7.	Conflict	https://www.prio.org/Data	network of military disputes between countries, denoted Conflict
8.	Protein-network	https://link.springer.com/article/10.1186/1479-7364-3-3-291	76 dimensional feature vector
9.	Epinions	https://snap.stanford.edu/data/soc-Epinions1.html	49,289 users, 139,738 items, 664,823 ratings and 487,183 trust relations
10.	Douban	https://www.kaggle.com/utmhikari/doubanmovieshortcomments	129,490 users, 58,541 items, 16,830,839 ratings and 1,692,952 friend relations
11.	Flixster	http://socialcomputing.asu.edu/datasets/Flixster	1,049,511 users, 66,726 items, 8,196,077 ratings and 11,794,648 social relations
12.	Twitter	www.gnip.com	tweets
13.	LinkedIn	https://github.com/linkedin/databus	LinkedIn's data processing pipeline
14.	SBU		million images with associated descriptive text.
15.	NUSWIDE	lms.comp.nus.edu.sg/research/NUS-WIDE.htm	269,648 images across 81 general noun concepts like "cat" and "boat".
16.	CCV	www.ee.columbia.edu/dvmm/CCV/	9317 YouTube videos (210 h in total) over 20 semantic concepts
17.	Flickr30K	https://github.com/BryanPlummer/flickr30k_entities	31,783 images collected from Flickr. Five high-quality sentences for each image

used by the winners; this includes training dataset, constructing tweet syntax and semantics, feature identification (Temporal Behavior, user profile, network features), and feature analysis. Their article also discussed several security threats caused because of the bots. In [240] proposed a multimodal sentiment analysis framework, that is produced using features for text, and visual data. They have attained significant improvement in the performance of the textual sentiment analysis system. Authors discussed the gaze — and smile-based facial expression features are commonly perceived to be quite beneficial for sentiment classification in this article. Hence their work multimodal sentiment analysis we can use in social, image, and video analysis.

In [231], presented the use of the Mixed graph of terms structures (MGT) for textual document's sentiment classification. The authors applied the Latent Dirichlet Allocation (LDA) based probabilistic approach for grabbing the sentiment. They tested and evaluated their work using public datasets. In [229], developed a multimodal affective data analysis framework that comprised of sets of appropriate features for text, audio, and visual data, for combining the features extracted from distinctive modalities. They improved the textual emotion analysis module by semantic-computing-based features, which offered notable development in the performance of the textual emotion analysis system. They developed a lexical resource, EmoSenticSpace, for emotion and sentiment detection from the text. Their two-stage emotion disclosure classifier from facial images also improved the system's accuracy. They worked on the eNTERFACE dataset — the only publicly accessible dataset on which multiple systems have been analyzed, enabling for a favorable comparison.

In [230], authors presented a model for sentiment classification system for social media data. Their system incorporated the approaches to consider for contextual polarities of terms to increase classification accuracy. They proposed a method to capture global context by the generation of a hybrid lexicon that improves a general-purpose lexicon with domain knowledge for sentiment classification. They demonstrated the distant supervision to capture global context by the generation. Their system achieved better classification performance than a state-of-the-art lexicon-based classifier, SentiStrength. They combined distant-supervised data from the three domains (Twitter, MySpace, and Digg) leads to overall significant performance improvement with the hybrid lexicon, confirming the transferability of the lexicon across social media platforms. They combined distant-supervised data from multiple social media platforms grabbed if there is not sufficient data from a target platform.

So sentiment analysis is not an easy job, and its difficulty increases with the nuance and complexity of opinions expressed. We learned that we could overcome this challenge by adopting machine learning in sentiment analysis. We surveyed many articles, and these articles give contributions to sentiment analysis oriented fields that utilize ML in SM for various real-world applications. It helps in multiple areas such as political sentiment analysis [241], applications in business intelligence, law, policymaking, sociology, and psychology.

3.12. Mentioned datasets

We always heard the word datasets while learning or reading about machine learning that is available in many publications. Datasets are available in multiple formats, and it is not that easy to use instantly, we need to make it usable. It requires a standard format to make them simple to use in the ML research. Selecting a suitable datasets means finishing our research work with 70% accuracy. For making things easy for young researchers who are yet to start with the datasets, we include datasets in a tabular form, that is adopted in articles that we are collected to write this survey paper in Table 11.

4. Consequences and statistical analysis

This section provides the consequences of ML on SM analysis along with statistical analysis of this survey paper.

4.1. Consequences

Consequences of SM using ML, this gives both positive and negative sides of SM analysis that adopting machine learning [242]. Machine learning on social media guarantees to convey expanding knowledge to programming worldwide and empower the systems to develop new abilities that were thought to be impossible. While the majority of the core algorithms that drive machine learning have been around for years. In recent years, what has increased in these algorithms' exceptional development are data selection and its processing power. Both algorithms and data keep on growing at exponential rates, proposing that machine learning is toward the start of a long and beneficial run. Machine learning on social media not just enhances the reliability of the answers accuracy, but also its increase in speed to respond. However, it can join unique sets of preliminary information to answer new inquiries. ML grants to understand the people in a meaningful manner and protect the privacy of an individual. ML gives a chance to develop more robust machines trying to resolve and prevent criminal activities from happening. It has plenty of positive corners to praise its existence to this era, and at the same time, it has darker sides too.

The dark side of machine learning on SM: It is the healthy times to machine learning as new algorithms, clusters of strong GPU's, and Tensorflow are joining to create robust systems. But the mighty use of these machine learning techniques leaves us to live with its darker shades. Let us explain these with example scenarios, the first scenario, recommender systems used by many companies suggests users about the products based on their purchases and interests. So companies try to apply machine learning algorithms for heavy user-oriented personalization that leads to a new problem, said Filter-bubble. Filter-bubble keeps recommending suggestions on the same interests always and freezes the recommender system that seems the user feels like they stuck in a bubble. Some times the heavy adoption of machine learnings systems get compromised and do greater damage than the help it actually given. Considering the attack on such system is also a challenging tasks to the researchers [243].

In another scenario, developing self-driving cars are the dream to us eight years back. But as an initial step researcher-developed self-driving cars in games and said it could not master humans in the real world. But things had changed, many tech giants like Google, Tesla, and many others now shifting their focus to develop self-driving cars. These driver-less cars came into reality because of the extensive development in extreme machine learning techniques like the use of neural networks. If any cyber terrorists understand networking causalities of these driver-less cars, they can easily breach into the systems and take these systems into their control to do damages.

Another terror example that we can quote here the backdrops of massive ML implementation on the systems, think that if the system we are training to build, such as a weapon control system or power grid system, were compromised, its destruction rate to society would be unimaginable. We can add one more example if we develop a system that automatically posts the contents related to a particular political party on social media, which significantly influences the public over other political parties. It is just not an end to ML, and it may also mimic us and can skip the facial recognition by automatically matching images, which downloads it from our social media. It is not the meaning that we are asking everyone to be professional if one wants to use the internet

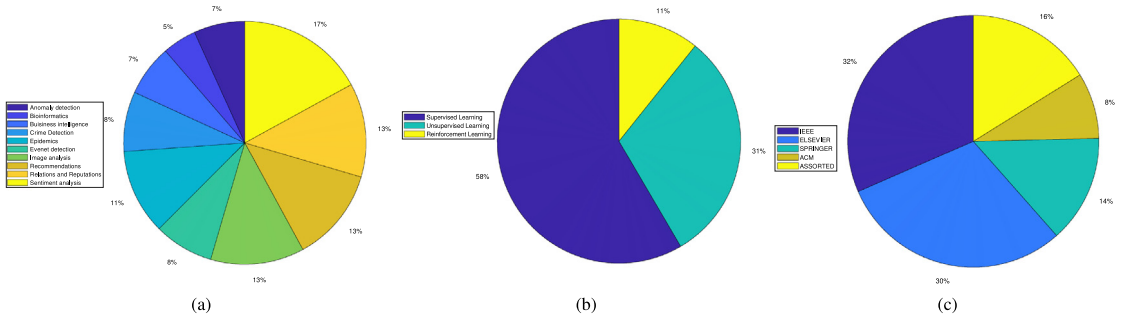

Fig. 14. Statistical Summary (a) Percentage of articles collected according to applications of SM analysis. (b) Percentage of articles surveyed according to the type of learning utilized. (c) Percentage of articles used in this study from major Databases.

here. Moreover, it is also a wrong thought to stop the use of the advancement in technology. However, we can make decisions to keep the technology we develop in the right hands.

So finally we can say that SM analysis encourages us to understand our audience. They show us whatever our most helpful social networks are. Social media analysis Supports us to Understand the contenders. Social metrics can serve us to design a better strategy. And social media analytics explains to us how a social media campaign is operating. By practicing social media analysis, we can evade inability about our life or appearance. We can beat the Fear of missing out from society. One can enjoy the fun of analytics during Isolation. Treatments for depression, anxiety, and Self-absorption are getting comfortable.

4.2. Statistical analysis

Finally, we are exhibiting the statistical analysis of our paper based on the articles we collected to write this survey paper. We surveyed this article with 130 papers collected from various reputed journals like Elsevier, IEEE, Springer and ACM. We neglected the proceedia papers to write this survey paper, and collected papers up to 2020. All papers are selected for this survey are based on their number of citations. Fig. 14(a), we displayed a pie chart on projected applications of SM analysis through machine learning. This pie chart visualizes the share of each application of SM, based on the sum of papers used for surveying this article. As discussed in Section 2.1., there are three types of learnings, such as SL, UL, and RL. We presented a pie chart that describes the percentage of papers we surveyed according to the type of learning used is shown in Fig. 14(b). In Fig. 14(c), we describe the Journal wise collection of articles that we used in this paper for surveying the Machine learning algorithms for social media analysis. We presented word cloud of our survey paper shown In Fig. 15 based on the words used while we are presenting our survey.

5. Open issues

This section provides various open issues for future studies of SM analysis using various ML approaches. All these issues which we are discussed in this section seek essential research to advance the applications of SM analysis usability.

Augmented And Virtual Reality: Augmented and virtuality are trends in the social media uploads [244]. Almost all leading social media platforms are supporting it. Many companies are unveiling their products using virtual reality concepts to their customers. For example, OnePlus, a Chinese mobile company, launched its mobile OnePlus 3 using the VR concept globally. Augmentation reality is also another concern in marketing the business on social media. Virtual reality gives the feel we are there as a first-person, and Augmented reality includes us/object as another person/object in the scenario. During the covid-19 pandemic, many page-3 people used this VR and AR for their weddings, and invited guests appeared virtually. IIT, Bombay, conducted their convocation ceremony using VR and AR technology. Such new techniques increase the business and improve user attachment, sentiment, and behavior with the products. The VR concept has adversaries too, which lead users to participate in crime with full 3-dimensional detail. This is where social media researchers have an extreme challenge to analyze such new types of data to optimize the SM analysis. AR and VR help treat the patients and get the awareness of the disease they are affected by without inquiring the person physically. It helps a lot to the medical practitioners while training and making them learn and valuable in medical marketing. VR is for image procurement and precise communications to give information concerning the illness. It is beneficial for the physical recovery and pain control of an infected patient during the treatment process. The image procured during the treatment can be uploaded by the practitioner in social media expert group for getting additional analysis from the professional people.

Bonding with Brands: The youth in the world passionate about wearing branded things [109]. This attitude makes of customers makes the companies to think them as their promoters in a modern style of using social media. Social media gives excellent interaction and commitment rates between user and brands. This engagement with brands by the users making the branded companies to retain the old and attract newbie's using SM. Researchers can use this data for presenting new business intelligence to the companies and can study the user behavior concerting about this.

Social Listening: Social listening is the state-of-the-art in the analysis of social media. Companies are transforming themselves smarter using this social listening. Social listening is abstractly helping them to double their profits. Social listening is the manner of following the discussions about distinct catchwords, terms or trademarks and then leveraging them to create possibilities and catch the trend. People using social media to check brand health by asking questions or poling. Sentiment analysis applying to users comments supplies the brand health to the companies and the other interested users. Hence, researchers can work on social listening to understand the user behavior, trending business intelligence solutions, and sentiment analysis.

Tailored Chat-Bots: Many online service providers are using customized or tailored chat-bots for improving customer relations and reducing maintenance cost. There are almost one million chat bots available on Facebook, a social media for making their customers stay happy by companies. Chat-bots communicate like a human with the user to fix the recorded issues from the customer side, it provides information about products, and take orders to purchase their product. These chat-bots not just saving the time of a company but also increases organizational productivity too. This is another exciting area to the

Fig. 15. Word cloud of journal papers used in writing our survey paper.

SM analysis researchers working to analyze the behavior of bots and analyze the sentiment between bots and persons. It tosses another challenging issue to the researchers that this chat-bots extensive personalization and improved efficiency leads add-on security issue too. So while making them smart, researchers need to consider their ability to participate in unintended purposes.

Share of Ads Impressions: Advertisements on social media are getting aggravated day by day on mobile and desktop platforms. Though the mobile ads grabs leading share, we can say almost 91% of ads are targetted on mobile users. The versatility of these advertisements may affect the user's behavior and progression of sales, and trust, which we yet to study.

Face Filters: Usage of taggings, stickers and emoticons are common in communicating between the social media users. There is one more trend appended to the list of trends, that is using filters. Uploading filtered images using Prisma or Instagram's face filters are the fashion to SM users. There are several types of filters we are using in the SM, such as face swap, face filters, age filters, and beautify options to images. Uploading a personal picture with the dog ears are very famous in social media, but at the same time, these kinds of pictures give confusion to the algorithms while processing to determine whether its a person or an animal. Analyzing this kind of pictures requires the development of extreme machine learning techniques. Filtered images lead to fraudulent information, hence SM analyzers have to develop more innovative methodologies to overcome such issues

Natural Language Processing: We can say that text is the new talk, but in the future digital speech is the latest trend. The rise of personal digital assistants is trending today from the release of Amazon Alexa. Personal digital assistants are new gadgets that are occupying the desks of geeks. They work on Natural Language Processing (NLP) techniques and maintains individuality for every customer. In the future, these devices will communicate social media without human intervention. They may become powerful as we are as with the increasing processing power. Most of the tech giants working on these devices, such as Google's Assistant, Apple Siri, Amazon Alexa, and Microsoft Cortana, are one of the best personal digital assistants available today. Hence social media analysts need to research these devices and their behavior with SM.

Recommender Systems: In social media apps, the image and video uploads are increased in the past decade. The users are uploading their images or videos on various occasions. In this context, extracting the brands from the images to recommend the items related to those brands will help to increase the business or online purchases. In the case of identifying the location of the image, we can recommend further visiting places from the same city or other visiting areas to the users. To extract and predict this king of features, big data analytics with the help of machine learning algorithms are enormously useful. Especially the deep learning or deep reinforcement learning are address these issues very efficiently.

Digital Garbage: Digital garbage is the data that is not useful in any manner. It is also considered the information which is unnecessarily replicated over the storage systems and involved during the processing. This data increases the unnecessary computations over the online data analysis and also occupies the storage systems. In this context, ML algorithms are the best solutions to identifies such garbage data. Notably, reinforcement learning, deep learning, dimensionality reduction, or outlier analysis techniques are the most suitable algorithms to avoid digital garbage in social media data. Identifying the garbage data generated sources also a challenging task, and here too, ML provides the best solutions.

COVID-19 related Issues: The pandemic covid-19 changes almost every human behavior [245] around the world. This pandemic gives a novel and never went challenges to the researchers. Undoubtedly social media researchers are one of the top participants during the pandemic for discovering new information. Unquestionably, social media greatly impacted the users and made researchers busy investigating how the users and affected people manage the pandemic [246]. This pandemic gives much scope to the researchers, such as identifying early experiences and analyzing such information shared by the patients on SM. Data scientists can research communication-related health crisis using social media data during pandemic times. Using SM data, one can explore urban spatial features of covid-19 transmission across the world. One more exciting behavior we can find from social media data is analyzing the new world of scientific communication during the Covid-19 outbreak.

It is evident that rather than affecting to the coronavirus, the fear of covid-19 traveled faster than its outbreak. So, fake news dissemination prevention and identification, prevention methodologies of fake characteristics and fake symptoms spread of covid-19, and creating alert models to fake drug [247] proposals for covid-19 are also severe research topics and gave a handful of work to the researchers.

6. Conclusion

Social media is a new human communication platform that has increased in popularity over the past few years. Applying machine learning on social media analysis is used in many applications. This paper gives a comprehensive survey of various applications using social media analysis through machine learning approaches. Our paper explores the prominent approaches to several applications, such as anomaly detection, behavioral analysis, bio-informatics, business intelligence, crime detection, epidemics, event detection, image analysis, Recommenders, relationships and reputations, and sentiment analysis. We start with machine learning algorithms and approaches used for social media analysis. Next, we summarize the benefits and shortcomings of various methods applied by the authors in their articles for the revolutionary social media data using machine learning. Later, we discuss the consequences of social media analysis using machine learning along with the statistical analysis. The addiction of the people to social media in the world is getting bigger, which pushes new research scopes. So, we finished with aggressive future directions for researchers as open issues in this paper.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We appreciate the time and efforts made by the editor during reviewing this manuscript. We pay our sincere thanks to the esteemed reviewers for their valuable comments and suggestions to improve the quality of this survey paper. We also thank Mr. Praveen Kumar Donta (IIT Dhanbad) and Mr. R. Satya Rajendra Singh (IIIT SriCity) for suggestions and comments on the manuscript.

References

[1] J.C. Flack, R.M. D'Souza, The digital age and the future of social network science and engineering, Proc. IEEE 102 (12) (2014) 1873–1877.
[2] V.U. Wanniarachchi, A. Mathrani, T. Susnjak, C. Scogings, A systematic literature review: What is the current stance towards weight stigmatization in social media platforms? Int. J. Hum.-Comput. Stud. 135 (2020) 102371.
[3] Z.M. Obeidat, R.S. AlGharabat, A.A. Alalwan, S.H. Xiao, Y.K. Dwivedi, N.P. Rana, Narcissism, interactivity, community, and online revenge behavior: The moderating role of social presence among Jordanian consumers, Comput. Hum. Behav. 104 (2020) 106170.
[4] T. Chen, E.C. Hui, J. Wu, W. Lang, X. Li, Identifying urban spatial structure and urban vibrancy in highly dense cities using georeferenced social media data, Habitat Int. 89 (2019) 102005.
[5] J.R. Quinlan, C4. 5: Programs for Machine Learning, Elsevier, 2014.
[6] E. Alpaydin, Introduction to Machine Learning, MIT Press, 2014.
[7] E. Whelan, A.N. Islam, S. Brooks, Applying the SOBC paradigm to explain how social media overload affects academic performance, Comput. Educ. 143 (2020) 103692.
[8] S. Buhan, Y. Özkazanç, et al., Wind pattern recognition and reference wind mast data correlations with NWP for improved wind-electric power forecasts, IEEE Trans. Ind. Inf. 12 (3) (2016) 991–1004.
[9] T. Kasakowskij, J. Fürst, J. Fischer, K.J. Fietkiewicz, Network enforcement as denunciation endorsement? A critical study on legal enforcement in social media, Telemat. Inform. (2019) 101317.
[10] Y. Sun, C. Liang, C.-C. Chang, Online social construction of Taiwan's rural image: Comparison between Taiwanese self-representation and Chinese perception, Tour. Manag. 76 (2020) 103968.
[11] Y. Gao, Y. Zhen, H. Li, T.-S. Chua, Filtering of brand-related microblogs using social–smooth multiview embedding, IEEE Trans. Multimed. 18 (10) (2016) 2115–2126.
[12] M. Egmont-Petersen, D. de Ridder, H. Handels, Image processing with neural networks–a review, Pattern Recognit. 35 (10) (2002) 2279–2301.
[13] C. Orellana-Rodriguez, M.T. Keane, Attention to news and its dissemination on Twitter: A survey, Comp. Sci. Rev. 29 (2018) 74–94.
[14] D. Praveen Kumar, A. Tarachand, A. Chandra Sekhara Rao, Machine learning algorithms for wireless sensor networks: A survey, Inf. Fusion 49 (2019) 1–25.
[15] M. Li, J. Wei, X. Zheng, M.L. Bolton, A formal machine–learning approach to generating human–machine interfaces from task models, IEEE Trans. Hum.-Mach. Syst. 47 (6) (2017) 822–833.
[16] J. Kim, M. Hastak, Social network analysis: Characteristics of online social networks after a disaster, Int. J. Inf. Manage. 38 (1) (2018) 86–96.
[17] K. Ravi, V. Ravi, A survey on opinion mining and sentiment analysis: tasks, approaches and applications, Knowl.-Based Syst. 89 (2015) 14–46.
[18] J. Bobadilla, F. Ortega, A. Hernando, A. Gutiérrez, Recommender systems survey, Knowl.-Based Syst. 46 (2013) 109–132.
[19] Z. Xiang, Q. Du, Y. Ma, W. Fan, A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism, Tour. Manag. 58 (2017) 51–65.
[20] D. Ramalingam, V. Chinnaiah, Fake profile detection techniques in large-scale online social networks: A comprehensive review, Comput. Electr. Eng. (2017).
[21] A. Sarker, R. Ginn, A. Nikfarjam, K. O'Connor, K. Smith, S. Jayaraman, T. Upadhaya, G. Gonzalez, Utilizing social media data for pharmacovigilance: a review, J. Biomed. Inform. 54 (2015) 202–212.
[22] B. Batrinca, P.C. Treleaven, Social media analytics: a survey of techniques, tools and platforms, AI Society 30 (1) (2015) 89–116.
[23] G. Bello-Orgaz, J.J. Jung, D. Camacho, Social big data: Recent achievements and new challenges, Inf. Fusion 28 (2016) 45–59.
[24] G. Harshvardhan, M.K. Gourisaria, M. Pandey, S.S. Rautaray, A comprehensive survey and analysis of generative models in machine learning, Comp. Sci. Rev. 38 (2020) 100285.
[25] B. Kavšek, N. Lavrač, APRIORI-SD: Adapting association rule learning to subgroup discovery, Appl. Artif. Intell. 20 (7) (2006) 543–583.
[26] S. Muggleton, L. De Raedt, Inductive logic programming: Theory and methods, J. Log. Program. 19 (1994) 629–679.
[27] M. Paolanti, E. Frontoni, Multidisciplinary pattern recognition applications: A review, Comp. Sci. Rev. 37 (2020) 100276.
[28] Y. Bengio, A. Courville, P. Vincent, Representation learning: A review and new perspectives, IEEE Trans. Pattern Anal. Mach. Intell. 35 (8) (2013) 1798–1828.
[29] B. McFee, L. Barrington, G. Lanckriet, Learning content similarity for music recommendation, IEEE Trans. Audio Speech Lang. Process. 20 (8) (2012) 2207–2218.
[30] J. Schmidhuber, Deep learning in neural networks: An overview, Neural Netw. 61 (2015) 85–117.
[31] I. Goodfellow, Y. Bengio, A. Courville, Y. Bengio, Deep Learning, Vol. 1, MIT Press Cambridge, 2016.
[32] P. Dixit, S. Silakari, Deep learning algorithms for cybersecurity applications: A technological and status review, Comp. Sci. Rev. 39, 100317,
[33] S. Ting, W. Ip, A.H. Tsang, Is Naive Bayes a good classifier for document classification, Int. J. Softw. Eng. Appl. 5 (3) (2011) 37–46.
[34] M.M. Saritas, A. Yasar, Performance analysis of ANN and Naive Bayes classification algorithm for data classification, Int. J. Intell. Syst. Appl. Eng. 7 (2) (2019) 88–91.
[35] R. Kohavi, Scaling up the accuracy of naive-bayes classifiers: A decision-tree hybrid, in: Kdd, Vol. 96, 1996, pp. 202–207.
[36] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al., Scikit–learn: Machine learning in Python, J. Mach. Learn. Res. 12 (Oct) (2011) 2825–2830.
[37] G. Mountrakis, J. Im, C. Ogole, Support vector machines in remote sensing: A review, ISPRS J. Photogramm. Remote Sens. 66 (3) (2011) 247–259.
[38] S.S. Istia, H.D. Purnomo, Sentiment analysis of law enforcement performance using support vector machine and K-nearest neighbor, in: 2018 3rd International Conference on Information Technology, Information System and Electrical Engineering, ICITISEE, IEEE, 2018, pp. 84–89.
[39] Y. Guo, M. Wang, X. Li, Application of an improved Apriori algorithm in a mobile e-commerce recommendation system, Ind. Manage. Data Syst. (2017).
[40] S. Asur, B.A. Huberman, Predicting the future with social media, in: 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, Vol. 1, IEEE, 2010, pp. 492–499.
[41] A. Criminisi, J. Shotton, E. Konukoglu, et al., Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi–supervised learning, Found. Trends® Comput. Graph. Vis. 7 (2–3) (2012) 81–227.
[42] K. Zhang, Y. Cheng, Y. Xie, D. Honbo, A. Agrawal, D. Palsetia, K. Lee, W.-k. Liao, A. Choudhary, SES: Sentiment elicitation system for social media data, in: 2011 IEEE 11th International Conference on Data Mining Workshops, IEEE, 2011, pp. 129–136.
[43] S.R. Safavian, D. Landgrebe, A survey of decision tree classifier methodology, IEEE Trans. Syst. Man Cybern. 21 (3) (1991) 660–674.
[44] C. Robert, Machine Learning, A Probabilistic Perspective, Taylor & Francis, 2014.
[45] M.I. Jordan, T.M. Mitchell, Machine learning: Trends, perspectives, and prospects, Science 349 (6245) (2015) 255–260.
[46] J. Chen, K. Li, Z. Tang, K. Bilal, S. Yu, C. Weng, K. Li, A parallel random forest algorithm for big data in a spark cloud computing environment, IEEE Trans. Parallel Distrib. Syst. 28 (4) (2016) 919–933.
[47] J. Snoek, H. Larochelle, R.P. Adams, Practical bayesian optimization of machine learning algorithms, in: Advances in Neural Information Processing Systems, 2012, pp. 2951–2959.
[48] K. Yu, L. Ji, X. Zhang, Kernel nearest-neighbor algorithm, Neural Process. Lett. 15 (2) (2002) 147–156.
[49] J.M. Keller, M.R. Gray, J.A. Givens, A fuzzy k-nearest neighbor algorithm, IEEE Trans. Syst. Man Cybern. (4) (1985) 580–585.
[50] R. Vatrapu, R.R. Mukkamala, A. Hussain, B. Flesch, Social set analysis: A set theoretical approach to big data analytics, IEEE Access 4 (2016) 2542–2571.
[51] C.K. Emani, N. Cullot, C. Nicolle, Understandable big data: a survey, Comput. Sci. Rev. 17 (2015) 70–81.
[52] N.A. Ghani, S. Hamid, I.A.T. Hashem, E. Ahmed, Social media big data analytics: A survey, Comput. Hum. Behav. 101 (2019) 417–428.
[53] S.J. Qin, L.H. Chiang, Advances and opportunities in machine learning for process data analytics, Comput. Chem. Eng. 126 (2019) 465–473.
[54] R.B. Rutledge, A.M. Chekroud, Q.J. Huys, Machine learning and big data in psychiatry: toward clinical applications, Curr. Opin. Neurobiol. 55 (2019) 152–159.
[55] I.H. Witten, E. Frank, M.A. Hall, C.J. Pal, Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann, 2016.
[56] R. Bost, R.A. Popa, S. Tu, S. Goldwasser, Machine learning classification over encrypted data, in: NDSS, Vol. 4324, 2015, p. 4325.
[57] J. Davis, M. Goadrich, The relationship between Precision-Recall and ROC curves, in: Proceedings of the 23rd International Conference on Machine Learning, 2006, pp. 233–240.

[58] M. Sokolova, G. Lapalme, A systematic analysis of performance measures for classification tasks, Inf. Process. Manage. 45 (4) (2009) 427–437.
[59] D. Chicco, G. Jurman, The advantages of the matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation, BMC Genomics 21 (1) (2020) 6.
[60] M. Sokolova, N. Japkowicz, S. Szpakowicz, Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation, in: Australasian Joint Conference on Artificial Intelligence, Springer, 2006, pp. 1015–1021.
[61] P. Branco, L. Torgo, R.P. Ribeiro, Relevance-based evaluation metrics for multi-class imbalanced domains, in: Pacific-Asia Conference on Knowledge Discovery and Data Mining, Springer, 2017, pp. 698–710.
[62] S. Cheong, S.H. Oh, S.-Y. Lee, Support vector machines with binary tree architecture for multi-class classification, Neural Inf. Process.-Lett. Rev. 2 (3) (2004) 47–51.
[63] V. Labatut, H. Cherifi, Evaluation of performance measures for classifiers comparison, 2011, arXiv preprint arXiv:1112.4133.
[64] D.J. Navarro, J.I. Myung, Model evaluation, in: Wiley StatsRef: Statistics Reference Online, Wiley Online Library, 2014.
[65] E. Costa, A. Lorena, A. Carvalho, A. Freitas, A review of performance evaluation measures for hierarchical classifiers, in: Evaluation Methods for Machine Learning II: Papers from the AAAI-2007 Workshop, 2007, pp. 1–6.
[66] C.N. Silla, A.A. Freitas, A survey of hierarchical classification across different application domains, Data Min. Knowl. Discov. 22 (1–2) (2011) 31–72.
[67] C. Maes, L. Schreurs, J.M. van Oosten, L. Vandenbosch, #(me) too much? The role of sexualizing online media in adolescents' resistance towards the metoo-movement and acceptance of rape myths, J. Adolesc. 77 (2019) 59–69.
[68] C. McClure, Y.-K. Seock, The role of involvement: Investigating the effect of brand's social media pages on consumer purchase intention, J. Retailing Consum. Serv. 53 (2020) 101975.
[69] Y. Qin, L.R. Men, Exploring negative peer communication of companies on social media and its impact on organization-public relationships, Public Relations Rev. (2019) 101795.
[70] J.M.N. Gonzalez, J.A. Jimenez, J.C.D. Lopez, et al., Root cause analysis of network failures using machine learning and summarization techniques, IEEE Commun. Mag. 55 (9) (2017) 126–131.
[71] S. Shahrampour, A. Rakhlin, A. Jadbabaie, Distributed detection: Finite-time analysis and impact of network topology, IEEE Trans. Automat. Control 61 (11) (2016) 3256–3268.
[72] J. Zhao, N. Cao, Z. Wen, Y. Song, Y.-R. Lin, C. Collins, # Fluxflow: Visual analysis of anomalous information spreading on social media, IEEE Trans. Vis. Comput. Graphics 20 (12) (2014) 1773–1782.
[73] L. Oneto, F. Bisio, E. Cambria, D. Anguita, Statistical learning theory and ELM for big social data analysis, IEEE Comput. Intell. Mag. 11 (3) (2016) 45–55.
[74] N. Cao, C. Shi, S. Lin, J. Lu, Y.-R. Lin, C.-Y. Lin, Targetvue: Visual analysis of anomalous user behaviors in online communication systems, IEEE Trans. Vis. Comput. Graph. 22 (1) (2016) 280–289.
[75] M. Stewart, U. Schultze, Producing solidarity in social media activism: The case of my stealthy freedom, Inf. Organ. (2019).
[76] M. Leban, Y. Seo, B.G. Voyer, Transformational effects of social media lurking practices on luxury consumption, J. Bus. Res. (2019).
[77] S. Sun, C. Luo, J. Chen, A review of natural language processing techniques for opinion mining systems, Inf. Fusion 36 (2017) 10–25.
[78] H. Arshada, A. Jantana, G.K. Hoon, I.O. Abiodun, Formal knowledge model for online social network forensics, Comput. Secur. (2019) 101675.
[79] L. Thomas, E. Orme, F. Kerrigan, Student loneliness: The role of social media through life transitions, Comput. Educ. (2019) 103754.
[80] D. Marengo, I. Poletti, M. Settanni, The interplay between neuroticism, extraversion, and social media addiction in young adult Facebook users: Testing the mediating role of online activity using objective data, Addict. Behav. 102 (2020) 106150.
[81] X. Luo, C. Jiang, W. Wang, Y. Xu, J.-H. Wang, W. Zhao, User behavior prediction in social networks using weighted extreme learning machine with distribution optimization, Future Gener. Comput. Syst. 93 (2019) 1023–1035.
[82] M. Singh, D. Bansal, S. Sofat, Behavioral analysis and classification of spammers distributing pornographic content in social media, Soc. Netw. Anal. Min. 6 (1) (2016) 41.
[83] L. Jin, Y. Chen, T. Wang, P. Hui, A.V. Vasilakos, Understanding user behavior in online social networks: A survey, IEEE Commun. Mag. 51 (9) (2013) 144–150.
[84] D. Wang, J. Li, K. Xu, Y. Wu, Sentiment community detection: exploring sentiments and relationships in social networks, Electron. Commer. Res. 17 (1) (2017) 103–132.
[85] Y. Wang, A.V. Vasilakos, J. Ma, N. Xiong, On studying the impact of uncertainty on behavior diffusion in social networks, IEEE Trans. Syst. Man Cybern.: Syst. 45 (2) (2015) 185–197.
[86] F.E. Ayo, O. Folorunso, F.T. Ibharalu, I.A. Osinuga, Machine learning techniques for hate speech classification of twitter data: State-of-the-art, future challenges and research directions, Comp. Sci. Rev. 38 (2020) 100311.
[87] E. Politou, E. Alepis, C. Patsakis, A survey on mobile affective computing, Comp. Sci. Rev. 25 (2017) 79–100.
[88] E. Cambria, Affective computing and sentiment analysis, IEEE Intell. Syst. 31 (2) (2016) 102–107.
[89] Z. Wang, Y. Yang, J. Pei, L. Chu, E. Chen, Activity maximization by effective information diffusion in social networks, IEEE Trans. Knowl. Data Eng. 29 (11) (2017) 2374–2387.
[90] Y. Wu, N. Pitipornvivat, J. Zhao, S. Yang, G. Huang, H. Qu, egoslider: Visual analysis of egocentric network evolution, IEEE Trans. Vis. Comput. Graph. 22 (1) (2016) 260–269.
[91] R. Zhao, K. Mao, Cyberbullying detection based on semantic–enhanced marginalized denoising auto-encoder, IEEE Trans. Affect. Comput. 8 (3) (2017) 328–339.
[92] T. Vafeiadis, K.I. Diamantaras, G. Sarigiannidis, K.C. Chatzisavvas, A comparison of machine learning techniques for customer churn prediction, Simul. Model. Pract. Theory 55 (2015) 1–9.
[93] T. Iwata, J.R. Lloyd, Z. Ghahramani, Unsupervised many–to–many object matching for relational data, IEEE Trans. Pattern Anal. Mach. Intell. 38 (3) (2016) 607–617.
[94] D.M. Jacoby, R. Freeman, Emerging network–based tools in movement ecology, Trends Ecol. Evol. 31 (4) (2016) 301–314.
[95] Z.-Y. Chen, Z.-P. Fan, M. Sun, Behavior–aware user response modeling in social media: Learning from diverse heterogeneous data, European J. Oper. Res. 241 (2) (2015) 422–434.
[96] C. Wu, X. Chen, W. Zhu, Y. Zhang, Socially–driven learning-based prefetching in mobile online social networks, IEEE/ACM Trans. Netw. 25 (4) (2017) 2320–2333.
[97] D. Ruths, J. Pfeffer, Social media for large studies of behavior, Science 346 (6213) (2014) 1063–1064.
[98] A. Lesk, Introduction to Bioinformatics, Oxford University Press, 2019.
[99] G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J.A. van der Laak, B. van Ginneken, C.I. Sánchez, A survey on deep learning in medical image analysis, Med. Image Anal. 42 (2017) 60–88.
[100] J. Fogel, M. Adnan, Trust for online social media direct-to-consumer prescription medication advertisements, Health Policy Technol. (2019).
[101] H. Müller, N. Michoux, D. Bandon, A. Geissbuhler, A review of content-based image retrieval systems in medical applications–clinical benefits and future directions, Int. J. Med. Inform. 73 (1) (2004) 1–23.
[102] M. Młyńczak, E. Migacz, M. Migacz, W. Kukwa, Detecting breathing and snoring episodes using a wireless tracheal sensor–A feasibility study, IEEE J. Biomed. Health Inform. 21 (6) (2017) 1504–1510.
[103] A. Budd, M. Corpas, M.D. Brazas, J.C. Fuller, J. Goecks, N.J. Mulder, M. Michaut, B.F. Ouellette, A. Pawlik, N. Blomberg, A quick guide for building a successful bioinformatics community, PLoS Comput. Biol. 11 (2) (2015) e1003972.
[104] C.A. Goble, J. Bhagat, S. Aleksejevs, D. Cruickshank, D. Michaelides, D. Newman, M. Borkum, S. Bechhofer, M. Roos, P. Li, et al., myExperiment: a repository and social network for the sharing of bioinformatics workflows, Nucleic Acids Res. 38 (suppl_2) (2010) W677–W682.
[105] A. Smiti, When machine learning meets medical world: Current status and future challenges, Comp. Sci. Rev. 37 (2020) 100280.
[106] C.D. Smyser, N.U. Dosenbach, T.A. Smyser, A.Z. Snyder, C.E. Rogers, T.E. Inder, B.L. Schlaggar, J.J. Neil, Prediction of brain maturity in infants using machine-learning algorithms, Neuroimage 136 (2016) 1–9.
[107] J. Torous, P. Staples, J.-P. Onnela, Realizing the potential of mobile mental health: new methods for new data in psychiatry, Curr. Psych. Rep. 17 (8) (2015) 61.
[108] H. Zou, B. Huang, X. Lu, H. Jiang, L. Xie, A robust indoor positioning system based on the procrustes analysis and weighted extreme learning machine, IEEE Trans. Wireless Commun. 15 (2) (2016) 1252–1266.
[109] M.M. Mostafa, More than words: Social networks' text mining for consumer brand sentiments, Expert Syst. Appl. 40 (10) (2013) 4241–4251.
[110] C. Townsend, D.T. Neal, C. Morgan, The impact of the mere presence of social media share icons on product interest and valuation, J. Bus. Res. 100 (2019) 245–254.
[111] X. Xu, X. Wang, Y. Li, M. Haghighi, Business intelligence in online customer textual reviews: Understanding consumer perceptions and influential factors, Int. J. Inf. Manage. 37 (6) (2017) 673–683.
[112] E. D'Avanzo, G. Pilato, Mining social network users opinions' to aid buyers' shopping decisions, Comput. Hum. Behav. 51 (2015) 1284–1294.
[113] V. Stantchev, L. Prieto-González, G. Tamm, Cloud computing service for knowledge assessment and studies recommendation in crowdsourcing and collaborative learning environments based on social network analysis, Comput. Hum. Behav. 51 (2015) 762–770.
[114] L.A. Johnson, N. Dias, G. Clarkson, A.M. Schreier, Social media as a recruitment method to reach a diverse sample of bereaved parents, Appl. Nurs. Res. (2019) 151201.

[115] X. Sun, C. Zhang, G. Li, D. Sun, F. Ren, A. Zomaya, R. Ranjan, Detecting users' anomalous emotion using social media for business intelligence, J. Comput. Sci. 25 (2018) 193–200.
[116] H. Chen, R.H. Chiang, V.C. Storey, Business intelligence and analytics: From big data to big impact, MIS Q. (2012) 1165–1188.
[117] J. Choi, J. Yoon, J. Chung, B.-Y. Coh, J.-M. Lee, Social media analytics and business intelligence research: A systematic review, Inf. Process. Manage. 57 (6) (2020) 102279.
[118] W. Fan, M.D. Gordon, The power of social media analytics, Commun. ACM 57 (6) (2014) 74–81.
[119] B.K. Chae, Insights from hashtag# supplychain and Twitter analytics: Considering Twitter and Twitter data for supply chain practice and research, Int. J. Prod. Econ. 165 (2015) 247–259.
[120] P.F. Kurnia, et al., Business intelligence model to analyze social media information, Procedia Comput. Sci. 135 (2018) 5–14.
[121] H. Rui, A. Whinston, Designing a social-broadcasting-based business intelligence system, ACM Trans. Manage. Inf. Syst. 2 (4) (2012) 1–19.
[122] Y. Chen, C. Jiang, C.-Y. Wang, Y. Gao, K.R. Liu, Decision learning: Data analytic learning with strategic decision making, IEEE Signal Process. Mag. 33 (1) (2016) 37–56.
[123] W. He, H. Wu, G. Yan, V. Akula, J. Shen, A novel social media competitive analytics framework with sentiment benchmarks, Inf. Manage. 52 (7) (2015) 801–812.
[124] W. Yuan, P. Deng, T. Taleb, J. Wan, C. Bi, An unlicensed taxi identification model based on big data analysis, IEEE Trans. Intell. Transp. Syst. 17 (6) (2016) 1703–1713.
[125] S. Yu, Y. Hu, When luxury brands meet China: The effect of localized celebrity endorsements in social media marketing, J. Retailing Consum. Serv. 54 (2020) 102010.
[126] A. Dabbous, K.A. Barakat, Bridging the online offline gap: Assessing the impact of brands' social network content quality on brand awareness and purchase intention, J. Retailing Consum. Serv. 53 (2020) 101966.
[127] M. Egele, G. Stringhini, C. Kruegel, G. Vigna, Towards detecting compromised accounts on social networks, IEEE Trans. Dependable Secure Comput. 14 (4) (2017) 447–460.
[128] L. Song, R.Y.K. Lau, R.C.-W. Kwok, K. Mirkovski, W. Dou, Who are the spoilers in social media marketing? Incremental learning of latent semantics for social spam detection, Electron. Commer. Res. 17 (1) (2017) 51–81.
[129] I. Frommholz, H.M. Al-Khateeb, M. Potthast, Z. Ghasem, M. Shukla, E. Short, On textual analysis and machine learning for cyberstalking detection, Datenbank-Spektrum 16 (2) (2016) 127–135.
[130] M.A. Bryan, Y. Evans, C. Morishita, N. Midamba, M. Moreno, Parental perceptions of the internet and social media as a source of pediatric health information, Acad. Pediatr. (2019).
[131] G. Sarna, M. Bhatia, Content based approach to find the credibility of user in social networks: an application of cyberbullying, Int. J. Mach. Learn. Cybern. 8 (2) (2017) 677–689.
[132] L.P. Del Bosque, S.E. Garza, Prediction of aggressive comments in social media: an exploratory study, IEEE Lat. Am. Trans. 14 (7) (2016) 3474–3480.
[133] M. Spreitzenbarth, T. Schreck, F. Echtler, D. Arp, J. Hoffmann, Mobile–Sandbox: combining static and dynamic analysis with machine-learning techniques, Int. J. Inf. Secur. 14 (2) (2015) 141–153.
[134] V. Van Vlasselaer, C. Bravo, O. Caelen, T. Eliassi-Rad, L. Akoglu, M. Snoeck, B. Baesens, APATE: A novel approach for automated credit card transaction fraud detection using network-based extensions, Decis. Support Syst. 75 (2015) 38–48.
[135] A. Culotta, Towards detecting influenza epidemics by analyzing Twitter messages, in: Proceedings of the First Workshop on Social Media Analytics, 2010, pp. 115–122.
[136] M. Gupta, A. Bansal, B. Jain, J. Rochelle, A. Oak, M.S. Jalali, Whether the weather will help us weather the COVID-19 pandemic: Using machine learning to measure Twitter users' perceptions, Int. J. Med. Inform. (2020) 104340.
[137] L. Li, Q. Zhang, X. Wang, J. Zhang, T. Wang, T.-L. Gao, W. Duan, K.K.-f. Tsoi, F.-Y. Wang, Characterizing the propagation of situational information in social media during COVID-19 epidemic: A case study on weibo, IEEE Trans. Comput. Soc. Syst. 7 (2) (2020) 556–562.
[138] E. Aramaki, S. Maskawa, M. Morita, Twitter catches the flu: detecting influenza epidemics using Twitter, in: Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, 2011, pp. 1568–1576.
[139] H. Achrekar, A. Gandhe, R. Lazarus, S.-H. Yu, B. Liu, Predicting flu trends using twitter data, in: 2011 IEEE Conference on Computer Communications Workshops, INFOCOM WKSHPS, IEEE, 2011, pp. 702–707.
[140] M.K. Othman, M.S.N.M. Danuri, Proposed conceptual framework of dengue active surveillance system (DASS) in Malaysia, in: 2016 International Conference on Information and Communication Technology, ICICTM, IEEE, 2016, pp. 90–96.
[141] S. Lim, C.S. Tucker, S. Kumara, An unsupervised machine learning model for discovering latent infectious diseases using social media data, J. Biomed. Inform. 66 (2017) 82–94.
[142] C. Liao, A. Squicciarini, C. Griffin, S. Rajtmajer, A hybrid epidemic model for deindividuation and antinormative behavior in online social networks, Soc. Netw. Anal. Min. 6 (1) (2016) 13.
[143] S.A. Sumner, S. Galik, J. Mathieu, M. Ward, T. Kiley, B. Bartholow, A. Dingwall, P. Mork, Temporal and geographic patterns of social media posts about an emerging suicide game, J. Adolesc. Health (2019).
[144] M.A. Al-garadi, M.S. Khan, K.D. Varathan, G. Mujtaba, A.M. Al-Kabsi, Using online social networks to track a pandemic: A systematic review, J. Biomed. Inform. 62 (2016) 1–11.
[145] R. Singh, R. Singh, A. Bhatia, Sentiment analysis using machine learning technique to predict outbreaks and epidemics, Int. J. Adv. Sci. Res 3 (2) (2018) 19–24.
[146] J.L. Marcus, W.C. Sewell, L.B. Balzer, D.S. Krakower, Artificial intelligence and machine learning for HIV prevention: Emerging approaches to ending the epidemic, Curr. HIV/AIDS Rep. (2020) 1–9.
[147] J. Xue, J. Chen, C. Chen, C. Zheng, S. Li, T. Zhu, Public discourse and sentiment during the COVID 19 pandemic: Using latent Dirichlet allocation for topic modeling on Twitter, PLoS One 15 (9) (2020) e0239441.
[148] Z. Long, R. Alharthi, A. El Saddik, Needfull–a tweet analysis platform to study human needs during the COVID-19 pandemic in new york state, IEEE Access 8 (2020) 136046–136055.
[149] C. Comito, A. Forestiero, C. Pizzuti, Twitter-based influenza surveillance: An analysis of the 2016–2017 and 2017–2018 seasons in Italy, in: Proceedings of the 22nd International Database Engineering & Applications Symposium, 2018, pp. 175–182.
[150] Z. Xu, H. Zhang, V. Sugumaran, K.-K.R. Choo, L. Mei, Y. Zhu, Participatory sensing–based semantic and spatial analysis of urban emergency events using mobile social media, EURASIP J. Wireless Commun. Networking 2016 (1) (2016) 44.
[151] H. Mo, X. Hao, H. Zheng, Z. Liu, D. Wen, Linguistic dynamic analysis of traffic flow based on social media–A case study, IEEE Trans. Intell. Transp. Syst. 17 (9) (2016) 2668–2676.
[152] E. D'Andrea, P. Ducange, B. Lazzerini, F. Marcelloni, Real–time detection of traffic from twitter stream analysis, IEEE Trans. Intell. Transp. Syst. 16 (4) (2015) 2269–2283.
[153] G. Lin, N. Sun, S. Nepal, J. Zhang, Y. Xiang, H. Hassan, Statistical Twitter spam detection demystified: Performance, stability and scalability, IEEE Access 5 (2017) 11142–11154.
[154] G. Mujtaba, L. Shuib, R.G. Raj, N. Majeed, M.A. Al-Garadi, Email classification research trends: Review and open issues, IEEE Access 5 (2017) 9044–9064.
[155] A. De, S. Bhattacharya, S. Sarkar, N. Ganguly, S. Chakrabarti, Discriminative link prediction using local, community, and global signals, IEEE Trans. Knowl. Data Eng. 28 (8) (2016) 2057–2070.
[156] R. Dreżewski, J. Sepielak, W. Filipkowski, The application of social network analysis algorithms in a system supporting money laundering detection, Inform. Sci. 295 (2015) 18–32.
[157] X. Zheng, Z. Zeng, Z. Chen, Y. Yu, C. Rong, Detecting spammers on social networks, Neurocomputing 159 (2015) 27–34.
[158] L.A. Maglaras, J. Jiang, T.J. Cruz, Combining ensemble methods and social network metrics for improving accuracy of OCSVM on intrusion detection in SCADA systems, J. Inf. Secur. Appl. 30 (2016) 15–26.
[159] L. Derczynski, D. Maynard, G. Rizzo, M. van Erp, G. Gorrell, R. Troncy, J. Petrak, K. Bontcheva, Analysis of named entity recognition and linking for tweets, Inf. Process. Manage. 51 (2) (2015) 32–49.
[160] D.T. Nguyen, J.E. Jung, Real–time event detection for online behavioral analysis of big social data, Future Gener. Comput. Syst. 66 (2017) 137–145.
[161] F. Buccafurri, G. Lax, A. Nocera, D. Ursino, Discovering missing me edges across social networks, Inform. Sci. 319 (2015) 18–37.
[162] M.-A. Kaufhold, M. Bayer, C. Reuter, Rapid relevance classification of social media posts in disasters and emergencies: A system and evaluation featuring active, incremental and online learning, Inf. Process. Manage. 57 (1) (2020) 102132.
[163] N. Panagiotou, I. Katakis, D. Gunopulos, Detecting events in online social networks: Definitions, trends and challenges, in: Solving Large Scale Learning Tasks. Challenges and Algorithms, Springer, 2016, pp. 42–84.
[164] N.M. Oliver, B. Rosario, A.P. Pentland, A Bayesian computer vision system for modeling human interactions, IEEE Trans. Pattern Anal. Mach. Intell. 22 (8) (2000) 831–843.
[165] D.C. Duro, S.E. Franklin, M.G. Dubé, A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery, Remote Sens. Environ. 118 (2012) 259–272.
[166] M. Sonka, V. Hlavac, R. Boyle, Image Processing, Analysis, and Machine Vision, Cengage Learning, 2014.
[167] J.A. Richards, J. Richards, Remote Sensing Digital Image Analysis, Vol. 3, Springer, 1999.

[168] S.E. Middleton, L. Middleton, S. Modafferi, Real–time crisis mapping of natural disasters using social media, IEEE Intell. Syst. 29 (2) (2014) 9–17.
[169] J. Cao, Y. Zhang, R. Ji, F. Xie, Y. Su, Web video topics discovery and structuralization with social network, Neurocomputing 172 (2016) 53–63.
[170] S. Lefèvre, D. Tuia, J.D. Wegner, T. Produit, A.S. Nassaar, Toward seamless multiview scene analysis from satellite to street level, Proc. IEEE 105 (10) (2017) 1884–1899.
[171] G. Chiachia, A.X. Falcao, N. Pinto, A. Rocha, D. Cox, Learning person–specific representations from faces in the wild, IEEE Trans. Inf. Forensics Secur. 9 (12) (2014) 2089–2099.
[172] L. Liu, C. Xiong, H. Zhang, Z. Niu, M. Wang, S. Yan, Deep aging face verification with large gaps, IEEE Trans. Multimed. 18 (1) (2016) 64–75.
[173] D. Wang, C. Otto, A.K. Jain, Face search at scale, IEEE Trans. Pattern Anal. Mach. Intell. 39 (6) (2017) 1122–1136.
[174] S. Kulkarni, S.F. Rodd, Context aware recommendation systems: A review of the state of the art techniques, Comp. Sci. Rev. 37 (2020) 100255.
[175] S. Rantala, A. Toikka, A. Pulkka, J. Lyytimäki, Energetic voices on social media? Strategic niche management and finnish Facebook debate on biogas and heat pumps, Energy Res. Soc. Sci. 62 (2020) 101362.
[176] X. Zhang, J. Jia, K. Gao, Y. Zhang, D. Zhang, J. Li, Q. Tian, Trip outfits advisor: Location–oriented clothing recommendation, IEEE Trans. Multimed. 19 (11) (2017) 2533–2544.
[177] A. Tommasel, D. Godoy, A social-aware online short-text feature selection technique for social media, Inf. Fusion 40 (2018) 1–17.
[178] R. Zhang, F. Nie, X. Li, X. Wei, Feature selection with multi-view data: A survey, Inf. Fusion 50 (2019) 158–167.
[179] Z. Yuan, J. Sang, C. Xu, Y. Liu, A unified framework of latent feature learning in social media, IEEE Trans. Multimed. 16 (6) (2014) 1624–1635.
[180] H. Zhang, X. Shang, H. Luan, M. Wang, T.-S. Chua, Learning from collective intelligence: Feature learning using social images and tags, ACM Trans. Multimedia Comput. Commun. Appl. 13 (1) (2017) 1.
[181] S. Poria, E. Cambria, A. Gelbukh, Aspect extraction for opinion mining with a deep convolutional neural network, Knowl.-Based Syst. 108 (2016) 42–49.
[182] K. Stepaniuk, The relation between destination image and social media user engagement–theoretical approach, Procedia-Soc. Behav. Sci. 213 (2015) 616–621.
[183] S.-E. Kim, K.Y. Lee, S.I. Shin, S.-B. Yang, Effects of tourism information quality in social media on destination image formation: The case of sina weibo, Inf. Manage. 54 (6) (2017) 687–702.
[184] M.J. Enright, J. Newton, Tourism destination competitiveness: a quantitative approach, Tourism Manage. 25 (6) (2004) 777–788.
[185] M. Elahi, F. Ricci, N. Rubens, A survey of active learning in collaborative filtering recommender systems, Comp. Sci. Rev. 20 (2016) 29–50.
[186] J. Lu, D. Wu, M. Mao, W. Wang, G. Zhang, Recommender system application developments: a survey, Decis. Support Syst. 74 (2015) 12–32.
[187] S. Raza, C. Ding, Progress in context-aware recommender systems–An overview, Comp. Sci. Rev. 31 (2019) 84–97.
[188] K. Gibson, S. Trnka, Young people's priorities for support on social media: "It takes trust to talk about these issues", Comput. Hum. Behav. 102 (2020) 238–247.
[189] B. Yang, Y. Lei, J. Liu, W. Li, Social collaborative filtering by trust, IEEE Trans. Pattern Anal. Mach. Intell. 39 (8) (2017) 1633–1647.
[190] Q. Fang, J. Sang, C. Xu, M.S. Hossain, Relational user attribute inference in social media, IEEE Trans. Multimed. 17 (7) (2015) 1031–1044.
[191] Y. Li, Y. Peng, W. Ji, Z. Zhang, Q. Xu, User identification based on display names across online social networks, IEEE Access 5 (2017) 17342–17353.
[192] Z. Sun, L. Han, W. Huang, X. Wang, X. Zeng, M. Wang, H. Yan, Recommender systems based on social networks, J. Syst. Softw. 99 (2015) 109–119.
[193] A. Daud, M. Ahmad, M. Malik, D. Che, Using machine learning techniques for rising star prediction in co–author network, Scientometrics 102 (2) (2015) 1687–1711.
[194] M. Ballings, D. Van den Poel, CRM in social media: Predicting increases in Facebook usage frequency, European J. Oper. Res. 244 (1) (2015) 248–260.
[195] X. Song, Z.-Y. Ming, L. Nie, Y.-L. Zhao, T.-S. Chua, Volunteerism tendency prediction via harvesting multiple social networks, ACM Trans. Inf. Syst. 34 (2) (2016) 10.
[196] J. Tang, S. Chang, C. Aggarwal, H. Liu, Negative link prediction in social media, in: Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, ACM, 2015, pp. 87–96.
[197] S. Jiang, A. Alves, F. Rodrigues, J. Ferreira Jr., F.C. Pereira, Mining point–of–interest data from social networks for urban land use classification and disaggregation, Comput. Environ. Urban Syst. 53 (2015) 36–46.
[198] I. Konstas, V. Stathopoulos, J.M. Jose, On social networks and collaborative recommendation, in: Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, 2009, pp. 195–202.
[199] C. De Maio, G. Fenza, V. Loia, M. Parente, Time aware knowledge extraction for microblog summarization on twitter, Inf. Fusion 28 (2016) 60–74.
[200] W. van Zoonen, G. Toni, Social media research: The application of supervised machine learning in organizational communication research, Comput. Hum. Behav. 63 (2016) 132–141.
[201] P. Chunaev, Community detection in node-attributed social networks: a survey, Comp. Sci. Rev. 37 (2020) 100286.
[202] X. Deng, J. Zhai, T. Lv, L. Yin, Efficient vector influence clustering coefficient based directed community detection method, IEEE Access 5 (2017) 17106–17116.
[203] A. Kumar, S.R. Sangwan, Rumor detection using machine learning techniques on social media, in: International Conference on Innovative Computing and Communications, Springer, 2019, pp. 213–221.
[204] J. Wu, L. Dai, F. Chiclana, H. Fujita, E. Herrera-Viedma, A minimum adjustment cost feedback mechanism based consensus model for group decision making under social network with distributed linguistic trust, Inf. Fusion 41 (2018) 232–242.
[205] S. Wen, J. Jiang, Y. Xiang, S. Yu, W. Zhou, W. Jia, To shut them up or to clarify: Restraining the spread of rumors in online social networks, IEEE Trans. Parallel Distrib. Syst. 25 (12) (2014) 3306–3316.
[206] E.D. Raj, L.D. Babu, An enhanced trust prediction strategy for online social networks using probabilistic reputation features, Neurocomputing 219 (2017) 412–421.
[207] J. Zhang, L. Tan, X. Tao, T. Pham, B. Chen, Relational intelligence recognition in online social networks—A survey, Comp. Sci. Rev. 35 (2020) 100221.
[208] C. Hu, H. Cao, Aspect–level influence discovery from graphs, IEEE Trans. Knowl. Data Eng. 28 (7) (2016) 1635–1649.
[209] S. Stieglitz, M. Mirbabaie, B. Ross, C. Neuberger, Social media analytics–Challenges in topic discovery, data collection, and data preparation, Int. J. Inf. Manage. 39 (2018) 156–168.
[210] C. Xu, S. Li, Y. Zhang, E. Miluzzo, Y.-f. Chen, Crowdsensing the speaker count in the wild: Implications and applications, IEEE Commun. Mag. 52 (10) (2014) 92–99.
[211] P. De Meo, E. Ferrara, D. Rosaci, G.M. Sarné, Trust and compactness in social network groups, IEEE Trans. Cybern. 45 (2) (2015) 205–216.
[212] M.V. Mäntylä, D. Graziotin, M. Kuutila, The evolution of sentiment analysis–A review of research topics, venues, and top cited papers, Comp. Sci. Rev. 27 (2018) 16–32.
[213] Q. Ye, Z. Zhang, R. Law, Sentiment classification of online reviews to travel destinations by supervised machine learning approaches, Expert Syst. Appl. 36 (3) (2009) 6527–6535.
[214] P. Ducange, M. Fazzolari, M. Petrocchi, M. Vecchio, An effective decision support system for social media listening based on cross-source sentiment analysis models, Eng. Appl. Artif. Intell. 78 (2019) 71–85.
[215] M. Ebrahimi, A.H. Yazdavar, A. Sheth, Challenges of sentiment analysis for dynamic events, IEEE Intell. Syst. 32 (5) (2017) 70–75.
[216] S. Manca, Snapping, pinning, liking or texting: Investigating social media in higher education beyond Facebook, Internet Higher Educ. (2019) 100707.
[217] V. Kagan, A. Stevens, V. Subrahmanian, Using twitter sentiment to forecast the 2013 pakistani election and the 2014 indian election, IEEE Intell. Syst. 30 (1) (2015) 2–5.
[218] P. Wang, B. Xu, Y. Wu, X. Zhou, Link prediction in social networks: the state–of–the–art, Sci. China Inf. Sci. 58 (1) (2015) 1–38.
[219] E. Lunando, A. Purwarianti, Indonesian social media sentiment analysis with sarcasm detection, in: 2013 International Conference on Advanced Computer Science and Information Systems, ICACSIS, IEEE, 2013, pp. 195–198.
[220] P. Nakov, S. Rosenthal, S. Kiritchenko, S.M. Mohammad, Z. Kozareva, A. Ritter, V. Stoyanov, X. Zhu, Developing a successful SemEval task in sentiment analysis of Twitter and other social media texts, Lang. Resour. Eval. 50 (1) (2016) 35–65.
[221] L. Ren, B. Xu, H. Lin, X. Liu, L. Yang, Sarcasm detection with sentiment semantics enhanced multi-level memory network, Neurocomputing (2020).
[222] B. Charalampakis, D. Spathis, E. Kouslis, K. Kermanidis, A comparison between semi-supervised and supervised text mining techniques on detecting irony in greek political tweets, Eng. Appl. Artif. Intell. 51 (2016) 50–57.
[223] C. Chen, Y. Wang, J. Zhang, Y. Xiang, W. Zhou, G. Min, Statistical features–based real–time detection of drifted Twitter spam, IEEE Trans. Inf. Forensics Secur. 12 (4) (2017) 914–925.
[224] M. Bouazizi, T. Ohtsuki, A pattern–based approach for multi–class sentiment analysis in Twitter, IEEE Access 5 (2017) 20617–20639.
[225] R.G. Guimarães, R.L. Rosa, D. De Gaetano, D.Z. Rodríguez, G. Bressan, Age groups classification in social network using deep learning, IEEE Access 5 (2017) 10805–10816.
[226] S. Poria, E. Cambria, A. Gelbukh, F. Bisio, A. Hussain, Sentiment data flow analysis by means of dynamic linguistic patterns, IEEE Comput. Intell. Mag. 10 (4) (2015) 26–36.
[227] T.H. Nguyen, K. Shirai, J. Velcin, Sentiment analysis on social media for stock movement prediction, Expert Syst. Appl. 42 (24) (2015) 9603–9611.

[228] P. Burnap, O.F. Rana, N. Avis, M. Williams, W. Housley, A. Edwards, J. Morgan, L. Sloan, Detecting tension in online communities with computational Twitter analysis, Technol. Forecast. Soc. Change 95 (2015) 96–108.
[229] S. Poria, E. Cambria, A. Hussain, G.-B. Huang, Towards an intelligent framework for multimodal affective data analysis, Neural Netw. 63 (2015) 104–116.
[230] A. Muhammad, N. Wiratunga, R. Lothian, Contextual sentiment analysis for social media genres, Knowl.-Based Syst. 108 (2016) 92–101.
[231] F. Colace, L. Casaburi, M. De Santo, L. Greco, Sentiment detection in social networks and in collaborative learning environments, Comput. Hum. Behav. 51 (2015) 1061–1067.
[232] T. Wang, H. Krim, Y. Viniotis, Analysis and control of beliefs in social networks, IEEE Trans. Signal Process. 62 (21) (2014) 5552–5564.
[233] M. Fire, R. Puzis, Organization mining using online social networks, Netw. Spat. Econ. 16 (2) (2016) 545–578.
[234] J. Serrano-Guerrero, J.A. Olivas, F.P. Romero, E. Herrera-Viedma, Sentiment analysis: A review and comparative analysis of web services, Inform. Sci. 311 (2015) 18–38.
[235] J. Sun, G. Wang, X. Cheng, Y. Fu, Mining affective text to improve social media item recommendation, Inf. Process. Manage. 51 (4) (2015) 444–457.
[236] E. Ferrara, O. Varol, C. Davis, F. Menczer, A. Flammini, The rise of social bots, Commun. ACM 59 (7) (2016) 96–104.
[237] D.P. Gandhmal, K. Kumar, Systematic analysis and review of stock market prediction techniques, Comp. Sci. Rev. 34 (2019) 100190.
[238] P.-Y. Chen, S.-M. Cheng, P.-S. Ting, C.-W. Lien, F.-J. Chu, When crowdsourcing meets mobile sensing: a social network perspective, IEEE Commun. Mag. 53 (10) (2015) 157–163.
[239] V. Subrahmanian, A. Azaria, S. Durst, V. Kagan, A. Galstyan, K. Lerman, L. Zhu, E. Ferrara, A. Flammini, F. Menczer, The DARPA Twitter bot challenge, Computer 49 (6) (2016) 38–46.
[240] S. Poria, E. Cambria, N. Howard, G.-B. Huang, A. Hussain, Fusing audio, visual and textual clues for sentiment analysis from multimodal content, Neurocomputing 174 (2016) 50–59.
[241] A.Y.F. Zhu, A.L.S. Chan, K.L. Chou, Creative social media use and political participation in young people: The moderation and mediation role of online political expression, J. Adolesc. 77 (2019) 108–117.
[242] W. Akram, R. Kumar, A study on positive and negative effects of social media on society, Int. J. Comput. Sci. Eng. 5 (10) (2017) 351–354.
[243] N. Pitropakis, E. Panaousis, T. Giannetsos, E. Anastasiadis, G. Loukas, A taxonomy and survey of attacks against machine learning, Comp. Sci. Rev. 34 (2019) 100199.
[244] N. Vaughan, B. Gabrys, V.N. Dubey, An overview of self-adaptive technologies within virtual reality training, Comp. Sci. Rev. 22 (2016) 65–87.
[245] J. Fry, J.M. Binner, Elementary modelling and behavioural analysis for emergency evacuations using social media, European J. Oper. Res. 249 (3) (2016) 1014–1023.
[246] K.B. Habersaat, C. Betsch, M. Danchin, C.R. Sunstein, R. Böhm, A. Falk, N.T. Brewer, S.B. Omer, M. Scherzer, S. Sah, et al., Ten considerations for effectively managing the COVID-19 transition, Nat. Hum. Behav. 4 (7) (2020) 677–687.
[247] I. Alimova, E. Tutubalina, Automated detection of adverse drug reactions from social media posts with machine learning, in: International Conference on Analysis of Images, Social Networks and Texts, Springer, 2017, pp. 3–15.

Balaji TK received B.Tech. from JNTU Hyderabad and M.Tech. from JNTUA Ananthapuramu with the specialization of Computer Science and Engineering in 2005 and 2012 respectively. Currently, he is pursuing his Ph.D. in the Department of Computer Science and Engineering at Indian Institute of Information Technology, Sricity, Andhra Pradesh, India since 2021. He has 13 years of teaching experience in India. He is a student member of IEEE and ACM. His current research is in Machine learning algorithms for social media Analysis.

Chandra Sekhara Rao Annavarapu received B.Tech from KL University, Guntur, and M.Tech from Andhra University, Vizag in 2003 and 2005 respectively. He received Ph.D. from Indian Institute of Technology (Indian School of Mines), Dhanbad, India in 2016. Currently, he is working as Assistant Professor, Department of Computer Science and Engineering, Indian Institute of Technology (Indian School of Mines), Dhanbad, Jharkhand, India. His research interests are Bioinformatics, Evolutionary Algorithms , Data mining , Machine Learning, and Wireless sensor networks.

Annushree Bablani received the B.Tech. degree in computer science and engineering from the Rajasthan College of Engineering, Jaipur, India, in 2013, and the M.Tech. degree in software engineering from the Government Engineering College, Ajmer, India, in 2016. She received Ph.D. from Department of Computer Science and Engineering, National Institute of Technology Goa, India, in 2020. Currently, she is working as Assistant Professor, Department of Computer Science and Engineering, Indian Institute of Information Technology, Sricity, Andhra Pradesh, India. Her research interests include brain–computer interface, soft computing, fuzzy logic, and machine learning.