INTERNATIONAL RESEARCH JOURNAL OF HUMANITIES AND INTERDISCIPLINARY STUDIES
(Peer-reviewed, Refereed, Indexed & Open Access Journal)
DOI : 03.2021-11278686 ISSN : 2582-8568 IMPACT FACTOR : 6.865 (SJIF 2023)

Effectively using Semantic Similarity Learning for Mining Hidden Social Network Contents

Deepak R. Patil[1], Parag Bhalchandra[2], Santosh Khamitkar[2]

[1] Department of Comp.Sci, KRP Kanya Mahavidyalaya, Islampur, Dist. Sangli (Maharashtra, India)

[2] School of Computational Sciences, S.R.T.M.University, Nanded (Maharashtra, India)

E-mail: drpatil9@gmail.com, srtmun.parag@gmail.com, s.khamitkar@gmail.com

DOI No. 03.2021-11278686 **DOI Link ::** https://doi-ds.org/doilink/03.2023-84911696/IRJHISIC2302043

Abstract:

Individuals and organizations are increasingly relying on social media to communicate. Massive volumes of publicly accessible data are stored on social media platforms, making it a great source of knowledge and insight. Text mining may be useful for generating insights from language data; however, it can be difficult to effectively deduce sense using social media text based on a single social media account. The study presents a technique for mining brief text structures to deduce the user's overarching themes from commonly appearing terms in social media accounts. The cosine textual similarity approach is used to determine the degree of similarity between two texts. It uses a clustering label propagation approach for labeling the text. This approach may be beneficial for getting decision-making insights from social media or other online forms that include short or sparse language.

1. Introduction:

In the modern world, the usage of social networking sites is increasing at an accelerating pace. A more concerning piece of evidence is that these networks have grown to turn out to be a significant repository for unprocessed data from a wide range of disciplines involving business, government, and health. Spam has been a severe security issue in cyberspaces and artificial intelligence-based detection systems are being investigated extensively [1]. Data mining methods are required to help organize and reform unstructured data in light of the growing dependence on social networks. Bridges between individuals who live in different parts of the world are now being made through the Internet. Participatory, procedures, feedback, community forums, and social network sites are also employed on the Internet [2]. Social networks have evolved into virtual communities that have been thoroughly explored in terms of evaluating human interactions and finding essential structural trends [3].Data mining is a fascinating technique for obtaining predictive information from raw data, and it has a lot of potentials. In most cases, a data mining strategy consists of many steps such as data comprehension and preparation, modeling, and assessment [4]. Data mining is the approach to identifying invisible patterns of information from which meaningful knowledge may be extracted. Its origins may be traced back to both classical statistical analysis and machine learning/artificial intelligence disciplines to reap the benefits of both [5].

1.1. Data Mining algorithm for Social Network Data:

The globe is becoming more and more like a little village as a result of the palpable effect of social media. It brings together individuals from all over the globe of all ages, and all countries enabling them to express themselves via the exchange of thoughts, experiences, sentiments, interests. Public and commercial organizations from all sectors may now promote, profit from, analyze and learn from the data supplied by social media, and develop their organizations as a result of the information offered by social media [6]. As a result, the importance of social media for academe and business is evident in the number of studies conducted by these two avenues to find solutions to critical problems in their respective fields. There are various challenges in thematerial on social media platforms such as the contentbeing disorganized, the information may be found in a variety of formats such as text, voice, photos, and videos.Furthermore, the standard statistical approaches are inadequate for analyzing such massive amounts of databecause of the vast volume of real-time data generated by social media[7].

Data mining methods might play a significant role to solve these challenges. Despite the huge amount of empirical research that has been conducted on data mining methods and social media, only a small quantity of research has been conducted for the comparison of data mining approaches in terms of precision, performance, and applicability. It has been shown that the precision of particular machine learning approaches is measured in a variety of wayswhich makes it tough to ascertain the applicability of the data mining algorithms [8].

1.2. Machine Learning Versus Non-Machine Learning Methods in Mining social media Data for Discovering Invisible Patterns of Social Collaboration:

Data mining methods are the procedures of obtaining concealed information out of a set of data via statistical analysis. It may be accomplished in a variety of ways involving the use of machine learning methods such as K-Nearest Neighbors (KNN), K-means. In certain circumstances, statistical approaches are regarded as non-machine learning methods for discovering patterns. A statistical approach is driven by data and is used to uncover trends and construct prediction models [9].

Machine learning-based text mining approaches vary from non-machine learning-based text mining techniques for the following reasons:

A. Conventional quantitative analysis techniques draw results from a sample of the population, but machine learning approaches enable the investigator to get results out of the complete population.
B. Conventional quantitative approaches permit the investigator to evaluate data theoretically, but machine learning techniques let the investigator derive the data's true meaning from natural language text.
C. Unlike conventional quantitative approaches, machine learning techniques do not need the investigator to analyze the data before evaluating it [10].

In this way, data mining encompasses all statistical and tentative data analysis techniques for data that make use of computers' processing capacity to extract and discover patterns from large amounts of information. As a result, machine learning and non-machine learning data mining approaches including classic quantifiable techniques in statistics, are complimentary [11]. Additionally, the arrangement of information sharing, and integrated social collaboration has good effects on the social cohesiveness of teams and organizations, according to researchers [12].

2. Review of Literature:

This section contains the explanation of the Review of Literature in the field of data mining techniques for social collaboration.In the context of comparative analysis of the review of literature Table-1 is given below:

Table-1: Comparative Analysis of Literature Review.

Author [Ref. no.]	Technique Used	Outcome	Research Gap
Nanayakkara et al., (2021) [13]	K-means, ANN, SVM, and Neuro-Fuzzy Logic.	The approach was found to be extensively utilized in the review.	Social media analytics practice would be utilized as a starting point for future research projects.
Zhang et al., (2018) [14]	BIM	Evaluated and assessed the structural properties of the identified collaborative network based on BIM at the micro, meso, and macro levels.	The advancement of the social group would be examined in more depth in the design environment as time progresses.
Chen et al.,	ANGIS	Findings indicated that all	In the future, land-use

(2017) [15]		models had AUCgreater than 0.75.	planning might take advantage of the new ensemble data mining methods.
Paul-Hus et al., (2017) [16]	Credit Attribution	Assessed collaboration practices to make more accurate predictions.	A more realistic image of the present research would be provided by further study.
Moro et al., (2016) [17]	LPC	Analysis assisted in making choices about whether or not to publish a post.	In the future, social media would be the most essential media channel for firms to contact their customers.
Kaur et al., (2016) [18]	Spectral-based, behavior-based, and structure-based.	A variety of data mining techniques were reviewed in work that focused on anomaly detection.	There are several possible avenues for further study in the conclusion of the report.
Injadat et al., (2016) [19]	Criterion-based	Data mining approaches had been employed using social media information to meet nine distinct study goals.	Data mining approaches have varied benefits and disadvantages, affecting future research selection.
Meng et al., (2016) [20]	SMC	The results as well as their limits and policy implications were reviewed after the study.	Future studies would identify and eradicate bias with special effort and resources.
Anders et al., (2016) [21]	TCP	TCPs provided a range of options for multi-modal communication and attention allocation.	Future studies would examine how well businesses and individuals can balance the advantages and downsides of communication visibility in TCPs.

3. Background Study:

Online purchasing has grown quite popular and has increased dramatically in recent years because of the digital revolution. Providing recommendations to detect users' preferences is a vital necessity for all search engines. A method to identify images based on analysis of variance (ANOVA) Cosine Similarity was reported in the study, in which text and visual characteristics were used to bridge the semantic gap between the two images. Using the ANOVA p-value, visual synonyms of each phrase are calculated by taking into consideration image visual attributes on a text-based search. For image recommendations, the cosine similarity between two images is calculated. Images gathered from a domain-specific site, are used in experiments. The quality of ranking images is examined with the assistance of users, and the relevance score is utilized to determine this. ANOVA Cosine Similarity (ACS) has an accuracy of relevance score of 15.26 percent for the top-10 selected images. The findings of the experiments reveal that the ANOVA Cosine Similarity Image Recommendation (ACSIR) approach surpasses several methods in terms of presenting more relevant results to the user's query [22].

4. Research Methodology:

The concept of designed architecture is examined in the context of research methodology. There are two techniques used in Mining the Hidden Social Network Content using Semantic Similarity-based learning which is explained below:

4.1. Techniques Used:

There are two techniques are used in the proposed methodology which is discussed below:

I. Cosine Textual Similarity Technique:

The cosine similarity metric is presently one of the extensively used measures of similarity around the globe. Text similarity measurement seeks to determine the cohesion occurring across text documents that is crucial to the majority of extracting information, retrieval of information, and text mining problems. In the study, a cosine textual similarity-based novel similarity measure is utilized to compare two texts. Text categorization, clustering, and query search are some of the activities that have benefited from the cosine similarity technique. The measure of cosine similarity is a prominent and often used similarity metric. A bag of words model is often used to estimate the similarity between two words [23].

This approach considers a document to be a collection of words and does not take into consideration grammar or word order. Consider the situation in which it is necessary to calculate a similarity score betwixt 2 documents, t, and d. There are several ways to quantify the degree of a document's similarity, but one typical technique involves assigning weights to terms based on how many times each phrase is used, the vector space model is employed for calculating the similarity. This modelconsidersevery document to be a vector, with each word corresponding to one of the components of vector space scoring models.

A significant portion of the computing overhead in document interpretation tasks comes from similarity measurement, and cosine similarity is every so often employed in text-similarity methods. Manning and Raghavan presented an illustration that showed the utility of cosine similarity in a straightforward manner [24]. To illustrate, four words (affection, jealousy, gossip, and wuthering) from the novel's sense and sensibility (SaS), pride and prejudice (PaP), and wuthering heights (WH) by Jane Austen, as well as the novel Wuthering Heights (WH) by Emily Bronte, are taken from their respective works. In novel d, the log frequency weight of term t has been determined using Eq. 1 for simplicity and consistency.

$$w_{t,d}=\begin{cases} 1+\log_{10} tf_{t,d} & if\ tf_{t,d} > 0 \\ 0 & otherwise \end{cases} \quad (1)$$

Where, w- frequency weight of texts.

t, d- texts or documents, between them similarity score is to be calculated.

$f_{t,d}$ – the function of 2 texts t and d.

Tables 2 and 3 show the number of times these phrases appear in each of the novels, as well as the log frequency weight associated with each of those occurrences. Finally, the cosine similarity betwixt these novels is shown in Table 4. When two documents are substantially identical, the cosine similarity function yields one, and when the documents are entirely diverse, the cosine similarity function gives zero [25].

Table-2: Term frequencies of terms in each of the novels [26].

Term	SaS	PaP	WH
Affection	115	58	20
Jealous	10	7	11
Gossip	2	0	6
Wuthering	0	0	38

Table-3: Log frequency weight of terms in each of the words [26].

Term	SaS	PaP	WH
Affection	3.06	2.76	2.30
Jealous	2.00	1.85	2.04
Gossip	1.30	0	1.78
Wuthering	0	0	2.58

Table-4: Cosine similarity between novels [26].

	SaS	PaP	WH
SaS	1.00	0.94	0.79
PaP	0.94	1.00	0.69
WH	0.79	0.69	1.00

To determine cosine similarity betwixt 2 documents x and y, they must first standardize to one another in the L2 norm (2).

$$\sum_{i=1}^{m} x_i^2 = 1 \quad (2)$$

Where x and y are two documents between them cosine similarity is to be determined and m is an integer number.

When two standardized vectors x and y are present, the cosine similarity betwixt them is calculated by taking the dot product of the two vectors (Eq. 3).

$$\cos(x, y) = \frac{\sum_{i=1}^{m} x_i y_i}{\sqrt{\sum_{i=1}^{m} x_i^2}\sqrt{\sum_{i=1}^{m} y_i^2}} \tag{3}$$

A careful inspection of Eq. 3 reveals that cosine similarity may be obtained from Euclidean distance in a straightforward manner (Eq. 4).

$$d_{Euclid}(x, y) = [\textstyle\sum_i (x_i - y_i)^2]^{\frac{1}{2}} [2 - 2\textstyle\sum_i x_i y_i]^{\frac{1}{2}} \tag{4}$$

II. Label Propagation Method:

Semi-supervised learning is made possible by the use of label propagation, which is an efficient and effective strategy. After the homogeneous network has been established via the cosine textual similarity technique, the label propagation technique is to be used to spread the labels throughout. Every node in a homogenous network would be assigned a unique label at the beginning of the network. After that, label propagation and selection must be carried out and repeated for numerous iterations. All of the nodes would communicate their labels to their immediate neighbors at the end of each round. In addition, each node would be issued a new label that would be derived from the labels of its neighboring nodes. It would be ended the label propagation and selection procedures if the labels of each node are not modified within this time. As a result, objects with similar labels should be grouped [27].

Each node would get several labels from its neighbors as the label propagation process proceeds. Every label would be assigned a value when the label selection process is being carried out. As previously stated, each homogeneous network edge is given a weight value that is proportional to its size. It is determined the degree of closeness between two objects by comparing their weights. Each neighboring object's label should be weighed against the weight of the edge that connects them when calculating a target object's label value.Two cases should be taken into consideration while determining the value of each label. In the first example, a distinctive label is sent to a target object by a neighbor, which the target item recognizes. It indicates that the target object does not get a label that is identical to that of its neighbors. Assume that 'A' is a target object which has a total of t distinct neighbors. These neighbors are $A_{1...}$, *At.* The *i-th* neighbor A_i sends out a distinctive label $label_i$ to 'A'. $weight\_value_i$ is the edge weight value betwixt A and A_i. In sucha particular instance, $label_i$ must be equivalent to $weight\_value_i$. In the second situation, multiple neighbors of a targeted object have asimilar label. The *m* neighbors of 'A' are supposed to betransmitted $label_{use}$ to A. These neighbors are $A_j, A_{j=1}, \ldots, A_{j+m-1}$. The weight values of edges betwixt A and these neighbors are *weight value*$_j$, *weight_value*$_{j+1}$,…, *weight_value*$_{j+m-1}$[28]. The following is the formula for calculating the value held by label *label* $_{use}$:

$$value_{label_{use}} = \sum_{t=j}^{j+m-1} weight\_value_t \qquad (5)$$

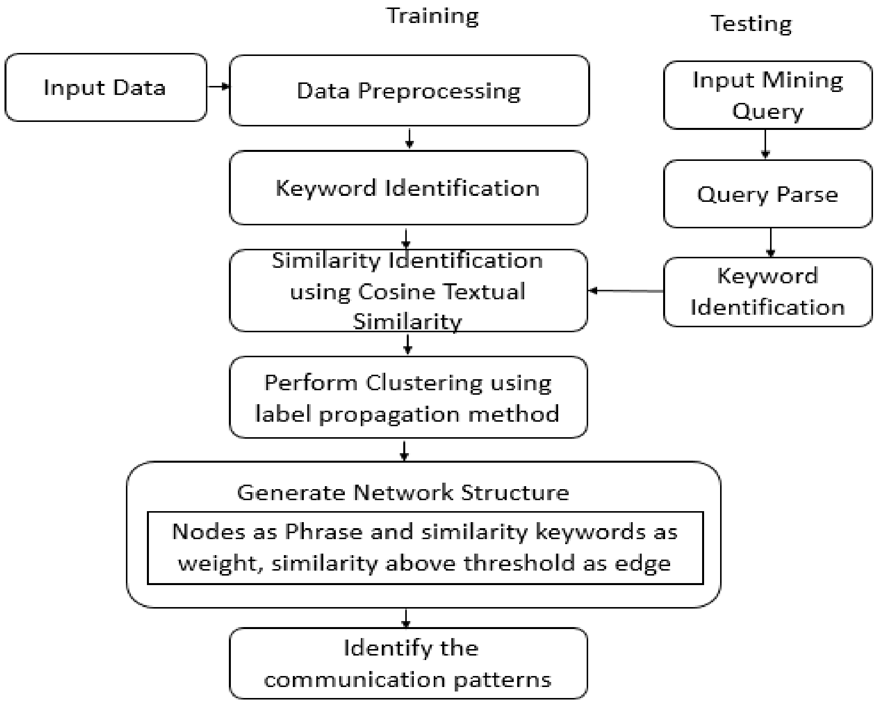
Where *label use* is the value held by *label, weight_valuet* is the assigned value to weight, *t* is the total number of neighbors, *j+m-1* is the limit of weight value.

Once all the neighbors' labels and their accompanying values are known, each target object must decide. It is determined by the monetary worth of the label. The label with the highest value would be selected. It may be inferred from the above statement that label selection takes place after label transmission has a place. At each node's request, the label transmission operation is carried out in an attempt to communicate the label to its neighbors. This procedure seems to be able to be run simultaneously. As soon as the label transmission mechanism is completed, every node would findnumerous labels. It is possible to run both processes at the same time because the label selection technique is segregated. As a result, by using a multi-thread approach, the label transmission and label selection tasks may be completed simultaneously [29].

4.2. Proposed Methodology:

This section contains an explanation of the proposed methodology which is shown in Figure 1 given below. The input data for the suggested approach is gathered from several social networking sites in the first instance. Following the gathering of data, data preparation is carried out to make it possible to identify keywords from the preprocessed data easily. Following that, the cosine textual similarity approach is used to determine the degree of similarity between two texts. After that, clustering is conducted on the data using the label preprocessing approach which allows for semi-supervised learning to take place. After that, a network structure is constructed by treating the node of the network as a phrase, similarity keyword as a phrase, and similarity above the threshold as an edge, among other considerations.In the end, the pattern is discovered, allowing the network to be organized more advantageously.The following framework is discussed in detail, step by step, in the context of research methodology:

Figure 1: Framework of Proposed Methodology.



Step-1: Training of the Input Data:

Throughout this part, the following steps are used to train the data:

I. Input Data:

In the first stage, data is gathered from numerous social networking sites to be processed further for mining for hidden social networking content.Two procedures are carried out on the obtained data for training and testing, which are detailed more below.

II. Data Preprocessing:

On the collected data, data preprocessing is performed because on the raw data work cannot be performed. Data preprocessing is the method of transforming unstructured data into a structure that could be accessed and used.

III. Keyword Identification:

In this step, keywords are identified from the preprocessed data. A keyword is a term or phrase that represents the meaning or primary concepts of a data collection. They are often used as a means of indexing the contents of a data source.

Step-2: Testing:

This section contains the testing of input data which is explained below:

I. Input Mining Query:

In this step, query mining is performed on the user query data on social networking sites. A text-based search is used to get the first set of results from expanded queries that have been entered by the user.

II. Query Parse:

In this section, parsing of the query is performed which is the method used to determine how many distinct ways a particular query may be executed before reaching a final choice. At the very least, each query should be parsed once. The Optimizer component is responsible for performing the parsing of a query inside a database environment.

Keyword Identification:

In this section, keywords are identified after performing the query parse process so that they can be used for processing.

Step-3: Similarity Identification using Cosine Textual Similarity Method:

In this step, a cosine textual similarity-based novel similarity measure is utilized to compare trained and tested texts. Text similarity measurement seeks to determine the cohesion occurring across text documents that is crucial to the majority of extracting information, retrieval of information, and text mining problems.

Step-4: Perform Clustering using the Label Propagation Method:

After finding the similarity between texts clustering is performed by the label propagation method. Clustering is the process of grouping the population of data points to make it easier to compare data points in the same group with those from other groups. A more straightforward explanation is that the objective is to divide groups with similar features and assign them to various clusters.

Step-5: Generate Network Structure:

After performing the clustering, the network structure is generated by considering nodes as phrases, similarity keywords as weight, and similarity which is above than threshold value as edge. An emerging sort of structure, network structure is considered to be less hierarchical (in other words, flatter), more decentralized, and more adaptable than other types of structures.

Step-6: Identify the Communication Pattern:

After generating the network structure communication pattern is identified in this step. Communication Patterns is a phrase used to describe a team's communication structure and how well it works.

Conclusion:

Most people can deduce meaning from brief texts like micro-blogs and Face-book status updates because they are familiar with the language employed. Text mining may be useful for generating insights from language data; however, it can be difficult to effectively deduce sense using social media text based on a single social media account. The conversation is a challenging task because of the restricted number of words available that may be gleaned from social media. It's also difficult to discern the context from a tiny selection of words. The study presents a technique for mining brief text structures to deduce the user's overarching themes from commonly appearing terms in social media accounts. The cosine textual similarity approach is used to determine the degree of similarity between two texts. It uses a clustering label propagation approach for labeling the text. This approach may be beneficial for getting decision-making insights from social media or other online forms that include short or sparse language. A step wise methodology is presented in the support of the hypothesis.

References:

[1]. Guo, Zhiwei, Jinhui Yang, and Keping Yu. "A Semantics and Behaviors-Collaboratively Driven Spammer Detection Method." In *International Conference on Autonomous Unmanned Systems*, pp. 1320-1329. Springer, Singapore, 2021.

[2]. Duffy, Brooke Erin, and Megan Sawey. "In/visibility in social media work: The hidden labor behind the brands." *Media and Communication* 10, no. 1 (2022): 77-87.

[3]. Mallavarapu, Aditi, Leilah Lyons, and Stephen Uzzo. "Exploring the utility of social-network-derived collaborative opportunity temperature readings for informing design and research of large-group immersive learning environments." *Journal of Learning Analytics* 9, no. 1 (2022): 53-76.

[4]. de Medeiros, Francisco Petrônio Alencar, and Alex Sandro Gomes. "An Approach Based on Social Network Analysis to Enhance Social Presence in a Collaborative Learning Environment." *IEEE Transactions on Education* (2022).

[5]. Santos, Ian Macedo Maiwald, Luciana de Oliveira Rech, and Ricardo Moraes. "A Topic Modeling Method for Analyzes of Short-Text Data in Social Media Networks." *EPiC Series in Computing* 82 (2022): 112-121.

[6]. Ghorbanalizadeh, Mojtaba, Nahideh Derakhshanfard, and NimaJafari Navimipour. "Introducing a new algorithm based on collaborative game theory with the power of learning selfish node records to encourage selfish nodes in mobile social networks." *Wireless Networks* (2022): 1-13.

[7]. Xiao, Xin, Zhaokun Yang, Qiuyan Yu, Dongjian Shi, Weifu Dong, Hongji Zhang, and Mingqing Chen. "Multiple Response Colors of Invisible Hollow Silica Photonic Crystals Patterns for Information Encoding." *Advanced Materials Interfaces* 8, no. 16 (2021): 2100814.

[8]. Pan, Mengyao, Chengyu Wang, Yuefan Hu, Xin Wang, Lei Pan, Yipeng Lin, Lei Shang, Hongbo Xu, Jiupeng Zhao, and Yao Li. "Dual Optical Information-Encrypted/Decrypted Invisible Photonic Patterns based on Controlled Wettability." *Advanced Optical Materials* 10,

no. 2 (2022): 2101268.

[9]. Mihaescu, Marian Cristian, and Paul Stefan Popescu. "Review on publicly available datasets for educational data mining." *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 11, no. 3 (2021): e1403.

[10]. Suriadi, Suriadi, Robert Andrews, Arthur HM ter Hofstede, and Moe Thandar Wynn. "Event log imperfection patterns for process mining: Towards a systematic approach to cleaning event logs." *Information systems* 64 (2017): 132-150.

[11]. Jeyapriya, A., and CS Kanimozhi Selvi. "Extracting aspects and mining opinions in product reviews using a supervised learning algorithm." In *2015 2nd international conference on electronics and communication systems (ICECS)*, pp. 548-552. IEEE, 2015.

[12]. Zhou, Changtong, Yong Qi, Shufen Zhang, Wenbin Niu, Suli Wu, Wei Ma, and Bingtao Tang. "Multicolor Invisible Patterns Encrypted in Double-Inverse Opal Films Based on Thermally Induced Structural Deformation." *physica status solidi (RRL)–Rapid Research Letters*.

[13]. Nanayakkara, A. C., B. T. G. S. Kumara, and R. M. K. T. Rathnayaka. "A Survey of Finding Trends in Data Mining Techniques for Social Media Analysis." *Sri Lanka Journal of Social Sciences and Humanities* 1, no. 2 (2021).

[14]. Zhang, Limao, and Baabak Ashuri. "BIM log mining: Discovering social networks." *Automation in Construction* 91 (2018): 31-43.

[15]. Chen, Wei, Mahdi Panahi, and Hamid Reza Pourghasemi. "Performance evaluation of GIS-based new ensemble data mining techniques of adaptive neuro-fuzzy inference system (ANFIS) with genetic algorithm (GA), differential evolution (DE), and particle swarm optimization (PSO) for landslide spatial modeling." *Catena* 157 (2017): 310-324.

[16]. Paul-Hus, Adèle, Philippe Mongeon, Maxime Sainte-Marie, and Vincent Larivière. "The sum of it all: Revealing collaboration patterns by combining authorship and acknowledgments." *Journal of Informetrics* 11, no. 1 (2017): 80-87.

[17]. Moro, Sérgio, Paulo Rita, and Bernardo Vala. "Predicting social media performance metrics and evaluation of the impact on brand building: A data mining approach." *Journal of Business Research* 69, no. 9 (2016): 3341-3351.

[18]. Kaur, Ravneet, and Sarbjeet Singh. "A survey of data mining and social network analysis based anomaly detection techniques." *Egyptian informatics journal* 17, no. 2 (2016): 199-216.

[19]. Injadat, Mohammad Noor, Fadi Salo, and Ali Bou Nassif. "Data mining techniques in social media: A survey." *Neurocomputing* 214 (2016): 654-670.

[20]. Meng, Yu. "Collaboration patterns and patenting: Exploring gender distinctions." *Research Policy* 45, no. 1 (2016): 56-67.

[21]. Anders, Abram. "Team communication platforms and emergent social collaboration practices." *International Journal of Business Communication* 53, no. 2 (2016): 224-261.

[22]. Sejal, D., T. Ganeshsingh, K. R. Venugopal, S. S. Iyengar, and L. M. Patnaik. "Image recommendation based on ANOVA cosine similarity." *Procedia Computer Science* 89 (2016): 562-567.

[23]. Turney, Peter D., and Patrick Pantel. "From frequency to meaning: Vector space models of semantics." *Journal of artificial intelligence research* 37 (2010): 141-188.

[24]. Tiwari, Prayag, Brojo Kishore Mishra, Sachin Kumar, and Vivek Kumar. "Implementation of n-gram methodology for rotten tomatoes review dataset sentiment analysis." In *Cognitive analytics: Concepts, methodologies, tools, and applications*, pp. 689-701. IGI Global, 2020.

[25]. Kumar, Sachin, and Durga Toshniwal. "A novel framework to analyze road accident time series data." *Journal of Big Data* 3, no. 1 (2016): 1-11.

[26]. Sohangir, Sahar, and Dingding Wang. "Improved sqrt-cosine similarity measurement." *Journal of Big Data* 4, no. 1 (2017): 1-13.

[27]. Hwang, Taehyun, and Rui Kuang. "A heterogeneous label propagation algorithm for disease gene discovery." In *Proceedings of the 2010 SIAM International Conference on Data Mining*, pp. 583-594. Society for Industrial and Applied Mathematics, 2010.

[28]. Shahreza, Maryam Lotfi, Nasser Ghadiri, Seyed Rasoul Mousavi, Jaleh Varshosaz, and James R. Green. "Heter-LP: A heterogeneous label propagation algorithm and its application in drug repositioning." *Journal of biomedical informatics* 68 (2017): 167-183.

[29]. Liu, Dongjiang, and Leixiao Li. "A Label Propagation Based Node Clustering Algorithm in Heterogeneous Information Networks." *IEEE Access* 9 (2021): 132631-132640.