

AN EFFICIENT SURVEY OF TEXT MINING TECHNIQUES FOR SOCIAL NETWORK ANALYSIS BASED ON FACEBOOK AND TWITTER PERSPECTIVES

Alampally Sreedevi¹, Kavitha Gopu²

¹²Assistant Professor, Dept of CSE, Sri Indu College of Engineering and Technology, Hyderabad, Telangana, India

ABSTRACT:

Text mining has turned out to be one of the in vogue handle that has been joined in a few research fields, for example, computational etymology, Information Retrieval (IR) and data mining. Natural Language Processing (NLP) methods were utilized to extricate learning from the textual text that is composed by people. Text mining peruses an unstructured form of data to give important information designs in a most brief day and age. Long range interpersonal communication locales are an awesome wellspring of correspondence as the vast majority of the general population in this day and age utilize these destinations in their everyday lives to keep associated with each other. It turns into a typical practice to not compose a sentence with remedy punctuation and spelling. This training may prompt various types of ambiguities like lexical, syntactic, and semantic and because of this kind of indistinct data; it is elusive out the genuine data arrange. As needs be, we are directing an examination with the point of searching for various text mining techniques to get different textual requests via web-based networking media sites. This review expects to depict how contemplates in online networking have utilized text investigation and text mining methods to identify the key topics in the data. This study concentrated on examining the text mining contemplates identified with Facebook and Twitter; the two prevailing web-based social networking on the planet. Aftereffects of this overview can fill in as the baselines for future text mining research.

Keywords: Text Mining, Social Media, Facebook, Twitter

I. INTRODUCTION

As we realize that there are different interpersonal interactions locales accessible, Face book and Twitter are considered as the most swarmed ones [1], [2]. These systems administration destinations have made it simple to speak with loved ones without trying [3], [4]. Individuals identified with various qualities come nearer to each other by sharing their thoughts, premiums, and knowledge [5].

These days, it turns out to be simple for anybody to meet the general population of their interests for learning and sharing valuable information [6], [7].

Text mining, also referred to as text data mining, roughly equivalent to text analytics, is the process of deriving high-quality information from text. High-quality information is typically derived through the devising of patterns and trends through means such as statistical pattern learning. Text mining usually involves the process of structuring the input text (usually parsing, along with the addition of some derived linguistic features and the removal of others, and subsequent insertion into a

database), deriving patterns within the structured data, and finally evaluation and interpretation of the output. 'High quality' in text mining usually refers to some combination of relevance, novelty, and interestingness. Typical text mining tasks include text categorization, text clustering, concept/entity extraction, production of granular taxonomies, sentiment analysis, document summarization, and entity relation modeling (i.e., learning relations between named entities).

Text analysis involves information retrieval, lexical analysis to study word frequency distributions, pattern recognition, tagging/annotation, information extraction, data mining techniques including link and association analysis, visualization, and predictive analytics. The overarching goal is, essentially, to turn text into data for analysis, via application of natural language processing (NLP) and analytical methods.

A typical application is to scan a set of documents written in a natural language and either model the document set for predictive classification purposes or populate a database or search index with the information extracted.

As individuals compose words or sentences with blunders, so as to give them a chance to compose or seek with legitimate punctuation and organized sentences, text mining approach [16] is utilized. Text mining implies the extraction of the data which isn't natural to anybody. On the off chance that we contrast web looking and text mining then both the terms are incomprehensibly not the same as each other. On the off chance that we discuss web looking, at that point you are completely mindful of what you will seek. Be that as it may, on account of text mining, the primary concentration is to draw out the most fitting data as per the composed text, regardless of whether it is organized or not. This system just requires a specific letters in order keeping in mind the end goal to uncover the data which is then additionally transformed into various proposals and desires. Text mining appears to get a handle on the entire programmed natural language processing. For example, investigation of linkage structures, references in scholarly written work and hyperlinks in the Web composing are imperative wellsprings of data that lie outside the regular region of NLP. NLP is one of the intriguing issues that worries about the interrelation among the colossal measure of unstructured text via web-based networking media [17], other than the examination and understanding of person languages [18], [19].

A few research articles were gathered from different databases with a specific end goal to be broke down and utilized as a part of this review. The inquiry terms incorporate "Text mining with web-based social networking", "Text mining with Facebook", and "Text mining with Twitter". This review is classified as takes after: area 2 gives a total foundation about the text mining field. Other related examinations are tended to by area 3. Conclusion and future points of view are introduced in area 4.

II. RELATED WORK

Organizations have recognized data-driven methodologies as the perfect outline for their development. It is less demanding to comprehend this hypothesis.

All things considered, wouldn't it profit an organization to get a thought regarding the view of its items in the market without consulting singular surveys from everybody? Wouldn't it be better in the event that they could check which political hopeful is perfect for their open picture without

analyzing them all independently? This is the reason advertise study and research are probably the most profoundly put fields on the planet at the present time. Person to person communication destinations like Twitter and Facebook are perfect for this reason. Posts or messages shared by individuals on these platforms with their companions remain unreservedly open or are kept secret. They allow organizations to gather up open assumptions [20], [21] about points that they are intrigued to share by a huge gathering of individuals.

2.1. Text Mining

Text mining makes it simple to acquire an important and organized data from the irregular data designs [24], [25], and [26]. It is extremely not a simple errand for the PCs to comprehend the unstructured data [27], [28] and make it organized. People can perform this assignment with no further endeavors because of the accessibility of various semantic methods. Be that as it may, individuals are restricted as far as speed and space as contrasting with PCs. That is, PCs are vastly improved than people to do these errands. A large portion of the current data in any association is spoken to in a text format, so on the off chance that we contrast data mining and text mining then text mining is more vital [29]. Be that as it may, as text mining is utilized for organizing the unstructured text data then this errand is all the more requesting when contrasted with data mining. By and large, the data identified with web-based social networking destinations isn't gathered for the exploration reason [30], it is compulsory to change the structure of the data originating from the web-based social networking. 80% of the accessible text on the web is unstructured while just 20% is organized [31].

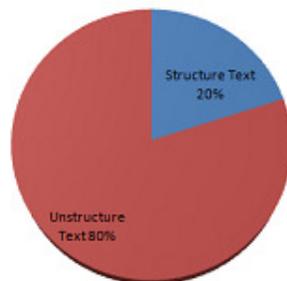


Figure 1: Text accessible on the web.

2.2. Text Mining versus Data Mining

On account of posting remarks on any post on various person to person communication locales, there isn't a solitary organized procedure accessible which causes issues in the direct use of the data.

Data accessible in the text format has significantly more significance and that is the reason text mining is producing much business esteem [32]. An investigation by [33] expressed that data mining speaks to the induction of a significant example or standards from a spatial database for determining a specific issue or issues. Data mining is not quite the same as text mining [34]. An investigation by [35] called attention to that text mining is significantly more intricate than data mining since it contains irregular and unstructured data designs, though data mining is managing the organized arrangements of data. The instruments that The processing of studies and open impressions utilizing extraordinarily composed computational frameworks is a mutual goal of between associated fields like subjectivity investigation, feeling mining [22], and supposition examination. Making critical thinking strategies or techniques to characterize the structure and priority or for outlining obstinate messages for specific subjects [23], events or items is another objective of the overview. For instance, these techniques could be utilized for measuring support for specific events or things, or determining thumbs down or thumbs up votes in favor of particular motion pictures in view of their surveys.

were utilized as a part of data mining were just managing organized data [34]. Text mining resembles a knowledge framework which is removing appropriate words or sentences from the uncalled for words and afterward transforming those words into the specific proposals. Text mining is fundamentally another field having the primary motivation behind data recuperation, machine learning, information mining and computational etymology [36].

2.3. Text mining in interpersonal organizations

The significance of text mining has been expanded because of the huge commitments in the field of innovation. Data mining as detailed by [37] is additionally critical yet because of the progression, text mining is having its spot. It is extremely a major push to pass on important information and learning [38] through effective dealing with and mining forms from the irregular form of information. In this period, organized data has lost its significance and the unstructured data has picked up the prominence. The vast majority of the associations are going towards text mining and overlooking the idea of data mining [39]. Researchers of [40] announced that all the long range interpersonal communication destinations are giving an awesome space to people to encourage connection and offer their perspectives and assessments. The best thing which these destinations are doing is that it has turned out to be simple for the people to comprehend a specific individual relying on his or her exercises. Through every one of these exercises, individuals identified with various traditions and qualities have come nearer to each other on account of having the better comprehension of each other's feelings, observations and zones of intrigue. As of now, UIs will be furnished with identity based characteristics [41]. Customized plans were utilized as a part of web based business [42], [43], e-learning, and information separating for upgrading distinctive styles and abilities.

III. Text mining endeavors in settling different NLP issues

An examination by [44] expressed that text mining is in charge of organizing the irregular data designs written in the human language. As a large portion of the general population interface with each other as text so for those individuals who are not ready to share organized form of data, text mining is the best method to deal with these circumstances. Among others, NLP is considered as the

most astonishing examination field. The fundamental objective of NLP is to look for information in regards to how the PC frameworks are examining and getting information from the languages of people to make utilizations of excellent [17]. The craft of imparting important information to the assistance of exceptional and unimportant data is genuinely something to be thankful for. Text mining method as depicted by [45] looks at the substance for separating the significant data which can be utilized for specific purposes. It would appear that text mining that will incorporate the general NLP plot [46] in its framework keeping in mind the end goal to successfully inspect the human language and to structure the unstructured data designs in like manner. As the innovation is propelling step by step, text mining framework will improve and better and this is the thing that all individuals are searching for.

3.1. Text mining in Facebook

The interpersonal organizations are developing at a fast rate without a break. In particular, the unstructured data is being put away on these systems as they go about as an expansive pool and this data relates to a large group of areas containing governments, organizations, and wellbeing. Data mining strategies have a tendency to transform the unstructured data for its situation inside an orderly plan [47]. These days, Facebook is a standout amongst the most well known online networking. This media is utilized by an expansive number of individuals on earth for communicating their thoughts, contemplations, distresses, joys and ballads [48]. Specialists had picked various Facebook factors that were required to build up the correct circumstance for completing our examinations. The important insights of client's identity are given by the Facebook profiles and exercises, which uncovered the real protests rather than anticipated or romanticized, character [49]. The advanced data has as of now seen a colossal development. The key territory of enthusiasm among experts is presently data mining and learning revelation. Also, a solid need has been felt to transform such data into helpful learning and information. Various applications like business administration and market examination have understood the advantages from the information and learning separated out of vast scale data. Information is put away in text form crosswise over different applications so one of the up and coming territories for look into is text mining. The difficult issue is separating the client required information. The learning revelation process has a vital advance which is accepted to be the Text Mining. The concealed information is separated from unstructured to semi-organized data in this procedure. Extricating information from various composed assets and its programmed disclosure is called as Text mining. In addition, PCs are likewise utilized for the needful and to meet this objective.

Researchers of [50] showed the text mining systems, techniques, and difficulties. These effective strategies would be depicted to give helpfulness over information obtaining amid text mining. The examination talked about the circumstances where every innovation could be useful for an alternate number of clients. Various business associations would be inspected by mining data that has been uncovered by their workers on LinkedIn, Facebook, and other transparently accessible sources. A system of informal social associations among representatives is removed through web crawler created for this reason. As per the discoveries, positions of authority can be recognized inside the association and this could be accomplished totally by utilizing machine learning methods other than centrality investigation. Bunching the interpersonal organization of an association and gathering accessible information inside each group can bring about the significant non-minor observations. A key resource or an extensive risk to the essential association can be the learning about the system of

informal connections. Other than examining interpersonal organizations of the associations, calculations and techniques used to accumulate data from openly accessible sources would be exhibited by this paper. A web crawler was created to acquire profiles of workers from six focused on associations and this was finished by gathering the Facebook data. An informal community topology was made for every association, and machine-learning calculations and centrality measures were actualized so the concealed administration positions inside each organization could be found. Besides, the social group bunches inside these associations were likewise uncovered by the calculations, which gave us understanding about the correspondence system of each organization notwithstanding the structure of the association.

As per an investigation by [51], it has turned out to be certain that online networking data is just defenseless to abuse. The plan incorporates organized approach and its application. Moreover, it involves performing a measurable bunch investigation notwithstanding the extensive examination of online networking remarks with the goal that analysts could decide the between connections among key elements. The subjective online networking data can be measured by these plans and therefore group them in view of their comparable highlights, and afterward they can be utilized as basic leadership instruments. The SAMSUNG Mobile Facebook page, where Samsung cell phones were presented, was utilized for the data procurement process. The remark distributed by Facebook clients on the subtitled Facebook page is alluded to as the "Data". In a time of 3 months, very nearly 128371 remarks were downloaded. The English remarks just were experienced through the investigation procedure. A while later, the calculated investigation was utilized by the substance examination and at last factual group investigation was performed via doing social investigation. Subsequently, web-based social networking data is incorporated by applying the factual group examination and it is performed in view of the yield of the reasonable investigation. The analysts are thus empowered to sort a huge dataset into numerous subsets, now and again, alluded to as articles. One of the controls of its application is showcasing. Components that can be reasonable sometimes are likewise limited by these kinds of methods.

An examination by [52] investigated the social data as a methodical data mining design. Discoveries showed that Facebook as a long range interpersonal communication site is the significant wellspring of data. Other than this approach, information on "my divider" post in regards to myself, age and remarks from the Facebook all are accentuated by the writer. It has been taken as a crude data, which is connected later to study and screen the explanatory strategies. What's more, the examination researched pictures for the commercial of their items and for the basic leadership process. Various data mining strategies go before the compulsion of scholarly learning from social data. Essentially, it sorts out the key information and other connected exercises in which clients are credited with respect to their associates on long range interpersonal communication destinations (i.e. Facebook). For the recuperation on Facebook client database, Facebook API performs Application Secret key and Facebook API Key are executed by Facebook API. Subsequently, WEKA records and data mining methods are bolstered to gather certain data into the auxiliary database, while the text data is spoken to by the disengaged data.

Specialists of [41] investigated the appropriateness of speaking to client's identity in light of the separated highlights from the Facebook data. The grouping strategies and their utilities were totally investigated with respect to the inspirational research results. An example of 250 client occasions from Facebook formed the examination study and this example was from around 10,000 notices,

which was conveyed by the My Personality venture [53]. The examination has the accompanying two interconnected goals: (1) knowing about the applicable identity connected markers that presents client data certainly or expressly in Facebook, and (2) distinguishing the achievability of prognostic character showing with the goal that up and coming canny frameworks could be bolstered. The investigation stressed on the advancement of appropriate highlights in a model, through which the improved yield of the classifiers under assessment could be watched.

3.2. Text mining in Twitter

A significant size of research has been occupied by the Twitter data analysis over the last couple of years [54]. Large spectrums of domains are using this data, some of which are using it for academic research and others for applications [55]. New improvements regarding twitter data are presented by this section. The document collection from various resources triggers the "Text Mining" process. A particular document would be retrieved by Text mining tool and this document is pre-processed by checking the character sets and format [56]. Subsequently, a text analysis phase would monitor the document. Semantic analysis is used to derive high-quality information from text; this is referred to "Text analysis". The market has a lot of text analysis techniques. Professionals can use combinations of techniques subject to the goal of the organization. Researchers tend to repeat the text analysis techniques till the time information is acquired. A management information system is capable of incorporating the resulting information, and as a result, significant knowledge is reduced for the user of that information system [57]. A key issue in text mining is intricacy of natural language. The ambiguity problem is much dense in the natural language. There are multiple meanings of a single word and multiple words can possess same meaning. Ambiguity is referred to as the understanding of a word which has more than one possible meaning. Noise has emerged in extracted information as a result of this ambiguity. Since usability and flexibility are the main parts of ambiguity, it cannot be removed from the natural language. One phrase or sentence can have multiple understandings, so there is a chance we can obtain a number of meanings. The work is still undeveloped and a particular domain is correlated with the suggested approach while the experts have attempted to resolve the ambiguity problem by performing a number of research studies. As there is uncertainty/vagueness in the semantic meanings of many discovered words, so it is very difficult to answer the requirements of the user.

Scholars of [58] developed and formulated an automatic classification technique through which potentially abuse-indicating user posts could be identified and evaluating the likelihood of social media usage as a source for automatic monitoring of drug medication abuse. In this regard, Twitter user posts (tweets) were collected and these were linked with three commonly abused medications (Oxycodone, Adderall, and Quetiapine). Besides interpreting a control medication (metformin), which is not the subject of abuse due to its process, nearly 6400 tweets were manually annotated, where these three medications were pointed out. The annotated data was qualitatively and quantitatively analyzed to determine as to whether or not signals of drug medication abuse are presented in Twitter posts. To sum up, Twitter's value was assessed in exploring the patterns of abuse over time and an automatic supervised classification technique was also designed, in which the purpose was to observe and separate the posts containing signals of medication abuse from those that do not. According to the findings of investigations, Twitter posts have yielded clear signals of medication abuse. As compared to the proportion for the control medication (i.e., metformin: 0.3 %), there is a very high ratio of tweets containing abuse signals for the three case medications

(Adderall: 23 %, oxycodone: 12 %, quetiapine: 5.0 %). In addition, almost 82 % accuracy (medication abuse class recall: 0.51, precision: 0.41, F-measure: 0.46) has been achieved through the automatic classification approach. The Study demonstrated how the abuse patterns over time can be analyzed by using the classification data and its goal is to illustrate the effectiveness of automatic classification. As a result, it is found that abuse-related information for medications can be significantly acquired from social media, and the research indicates that natural language processing and supervised classification are the automatic approaches that have potentials for future monitoring and intervention assignments. With respect to supervised learning, the lack of sufficient training data is believed to be the largest shortcoming of the study. Both annotation and automatic classification are hindered by the lack of context and ambiguity in tweets. During the course of annotations, many ambiguous tweets were found and services of pharmacology expert were hired to address these issues. As a result of these ambiguities, the undefined situation is observed in the binary classification process and this inadequacy will continue until the time fine-tuned annotation rules could be specified by the future annotation rules.

A study by [59] applied the text mining approach on a large dataset of tweets. The complete Twitter timelines of 10 academic libraries were used to collect the dataset for this research. Nearly 23,707 tweets formed the total dataset, where there were 7625 hashtags, 17,848 mentions, and 5974 retweets. Inconsistency among academic libraries is found in the distribution of tweets. "Open" was the most repeated word that was used by the academic libraries in different perspectives. It was observed that "special collections" was the most frequent bigram (two-word sequence) in the aggregated tweets. While "save the date" was the most recurrent tri-gram (three-word sequence). In the semantic analysis, words such as "insight, knowledge, and information about cultural and personal relations" were the most frequent word categories. Moreover, "Resources" was the most widespread category of the tweets among all the selected academic libraries. The significance of data and text-mining approaches are reported within the study and their purpose is to gain an insight with the aggregate social data of academic libraries so that the process of decision-making and strategic planning could become facilitated for marketing of services and patron outreach. The 10 academic libraries from top global universities have undergone the text mining approach. The study aimed to illustrate their Twitter usage and to examine their tweet content.

As far as social media is concerned, decision-making is supported and user-generated text is analyzed through text mining and content analysis [60]. By employing an archiving service (twimemachine.com) in December 2014, the complete Twitter timelines of 10 academic libraries were taken into account to collect the dataset for this research. The libraries of 10 highest-ranking universities from the global Shanghai Ranking were chosen for that purpose. The language of the university must be English-based, which was the condition for selection and selection was restricted to only one library if there was more than one library in the university. Certain weaknesses were found in the study, for example, all of the libraries are English-language libraries in the sample and only 10 academic libraries were considered for the analysis. This gap must be filled in future by applying the analysis to a dataset from diversified academic libraries, including non-English language libraries. Consequently, a complete understanding of tweet patterns would be acknowledged. The future inquiry can also incorporate the international or cross-cultural comparisons. Any discrepancy among libraries in their tweets' content affected by the number and interaction of followers could be highlighted by the analysis and its findings. The accuracy of the tweet categorization tool has

yielded the inadequate findings, and the said tool needs to be substantiated through other machine-learning models along with their applications.

Researchers of [55] demonstrated in a smoking cessation nicotine patch study an innovative Twitter recruitment system that is deployed by the group. The study aimed to describe the methodology and used to address the issue of digital recruitment. Furthermore, designing a rule-based system with the provision of system specification besides representing the data mining approaches and algorithms (classification and association analysis) using Twitter data. Twitter's streaming API captured two sets of streaming tweets, which were collected for the study. Ten search terms, (i.e. quitting, quit, nicotine, smoking, smoke, patches, cig, cigarette, ecig, cigs, marijuana) were used to gather the first set. The second set of tweets contains 30 terms, in which the terms from the first set were included. Moreover, the second set is a superset of the first one. A number of studies have been conducted to review the information gathering methods. As unstructured data sets are in the textual format, the use of various procedures of text mining has been tackled by many research studies. Nonetheless, the data sets on the social networking websites are not mainly discussed by these studies. A study by [50] applied various text mining techniques. The study would describe the application of these strategies in the social networking websites. In the field of intelligent text analysis, the latest improvements would also be examined in the survey. The study focused on two key techniques pertaining to the text mining field, namely classification and clustering. Usually, they are operated for the study of the unstructured text accessible on the extensive scale frameworks. Prior to the start of World Cup, a total of approximately 30,000 tweets were used by [61]. Moreover, an algorithm was used for integrating the consensus matrix and the DBSCAN algorithm. Consequently, the concerned tweets on those prevailing topics were available to him. Afterward, the clustering analysis was applied to seek the topics discussed by the tweets. The tweets were grouped utilizing the k-means [62], Non-Negative Matrix Factorization (NMF), and a popular clustering algorithm. After that, the results were compared. Similar results were delivered by both algorithms. However, NMF became faster and the researchers could easily interpret the outcomes.

A study by [1] initiated a workflow to gain an insight into both the large-scale data mining methods and qualitative analysis. Twitter posts of engineering students were the primary concern. The basic goal was to identify their issues in their academic experiences. The study conducted a qualitative analysis of samples obtained from around 25,000 tweets that were associated with the engineering students and their college life. The encounter troubles of engineering students were discovered during the study. For example, a large volume of study, sleep deprivation and lack of social engagement. Considering these outcomes, a multi-label classification algorithm was implemented to categorize tweets in lieu of students' queries. The algorithm was applied on approximately 35,000 tweets streamed at the geo-location of Purdue University. At the first instance, the concerned authorities have addressed the experiences and issues of the students and social media data was used to expose the issues. Moreover, a study by [1] also developed a multi-label classifier so that tweets founded within the content evaluation phase could be organized. A number of renowned classifiers are significantly consumed in machine learning domain and data mining process. With Comparison to other state-of-the-art multi-label classifiers, the Naïve Bayes classifiers were found proficient on the dataset.

A study by [63] discussed the clustering technique, the execution of correlation and association analyses to social media. The investigation of insurance Twitter posts was carried out to assess this

matter. Consequently, recognizing theories and keywords in the social media data has become an easy task, due to which the information by insurers and its application would be facilitated. After having a detailed analysis, client queries and the potential market would be proactively addressed with usefulness and the findings of the analysis are to be effectively implemented in suitable fields. According to this evaluation, the overall 68,370 tweets were utilized. Two additional kinds of evaluation need to be applied to the data. The first is the clustering analysis, through which the tweets depending on their similarities or dissimilarities would be merged. An Association Analysis is the second one whereas the occurrences of particular composed words were discovered.

Authors of [64] stated that sentiment analysis through social media usage has witnessed a huge interest from scholars in the last few years. In that, the authors discussed the influence of tweets' sentiment on elections and the impact of the elections' results on web sentiment.

Conclusion

The technique for correspondence with each different has now totally changed because of the advance in the field of web-based social networking. These days, modernization can be seen all over and in view of that; the information creation is touching the elevations. As of now, the new organizations are advancing to take a dynamic part in transforming the specialized technique [65]. The catchphrases and expressions' particularization can wind up accommodating to various organizations fit as a fiddle their future. In the present examination, we have featured the cutting edge look into work with respect to the execution of text mining in the most predominant web-based social networking (Facebook and Twitter). From the perspective of a few researchers, Text mining was clarified through different models. In addition, diverse real references are additionally given to help the examination work. Subsequently, text mining can be characterized into text bunching, text order, affiliation manage extraction and pattern examination as per applications. With the progression of time, text mining will be advanced well..We can see from the overviewed writing that Telugu text in online networking is ignored from the perspective of a few text mining considers. Subsequently, this hole opens the entryway for some, text mining researchers to connect that hole through directing different examinations in the field of text mining in the Telugu language context. An investigation by [66] contended that scientists breaking down the Telugu post are only here and there discovered, concentrating on the text mining of English, but the Telugu post via web-based networking media is available in mass sum. Researchers of [67] sketched out its odd and impossible to miss qualities as the explanations for this state of mind. From the overviewed writing, we have watched that specialists have given careful consideration to notion investigation in the Telugu text. The modern errands of parsing and sense disambiguation brace creation of target arrangements of the most intermittent linguistic structures and faculties of polysemous words, and the potential for syntactic and semantic equivocalness is observed to be high [68]. As a future work, we are very keen on examining the text mining procedures on Telugu textual data from Facebook and Twitter. What's more, future research should take feeling examination of Telugu text into thought. The Telugu language is convoluted morphologically, has free word arrange, accentuation only here and there found and short vowels are maintained a strategic distance from in the composed form of Standard Telugu. Thus, context is fundamental to kill winning equivocalness from obviously indistinguishable forms which is huge in perceiving sentiments.

References

- [1] Chen, X., Vorvoreanu, M., & Madhavan, K. (2014). Mining social media data for understanding students' learning experiences. *IEEE Transactions on Learning Technologies*, 7(3), 246-259.
- [2] Buettner, R. (2016). Predicting user behavior in electronic markets based on personality-mining in large online social networks. *Electronic Markets*, 1-19.
- [3] Baumer, E. P., Sinclair, J., & Tomlinson, B. (2010, April). America is like Metamucil: fostering critical and creative thinking about metaphor in political blogs. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1437-1446). ACM.
- [4] Stieglitz, S., Dang-Xuan, L., Bruns, A., & Neuberger, C. (2014). Social media analytics. *Wirtschaftsinformatik*, 56(2), 101-109.
- [5] Kasture, N. R., & Bhilare, P. B. (2015, February). An Approach for Sentiment analysis on social networking sites. In *Computing Communication Control and Automation (ICCUBEA), 2015 International Conference on* (pp. 390-395). IEEE.
- [6] Sorensen, L. (2009, May). User managed trust in social networking-Comparing Facebook, MySpace and LinkedIn. In *Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronic Systems Technology, 2009. Wireless VITAE 2009. 1st International Conference on* (pp. 427-431). IEEE.
- [7] Naaman, M. (2012). Social multimedia: highlighting opportunities for search and mining of multimedia data in social media applications. *Multimedia Tools and Applications*, 56(1), 9-34.
- [8] Evans, B. M., Kairam, S., & Pirolli, P. (2010). Do your friends make you smarter?: An analysis of social strategies in online information seeking. *Information Processing & Management*, 46(6), 679-692.
- [9] Li, J., & Khan, S. U. (2009, November). MobiSN: Semantics-based mobile ad hoc social network framework. In *Global Telecommunications Conference, 2009. GLOBECOM 2009. IEEE* (pp. 1-6). IEEE.
- [10] Liu, X., Wang, M., & Huet, B. (2016). Event analysis in social multimedia: a survey. *Frontiers of Computer Science*, 1-14.
- [11] Tsytarau, M., & Palpanas, T. (2012). Survey on mining subjective data on the web. *Data Mining and Knowledge Discovery*, 24(3), 478-514.
- [12] Al-Emran, M., & Malik, S. I. (2016). The Impact of Google Apps at Work: Higher Educational Perspective. *International Journal of Interactive Mobile Technologies (IJIM)*, 10(4), 85-88.
- [13] Al-Emran, M. N. H. (2014). Investigating Students' and Faculty members' Attitudes Towards the Use of Mobile Learning in Higher Educational Environments at the Gulf Region.
- [14] Al-Emran, M., & Shaalan, K. (2015, August). Learners and educators attitudes towards mobile learning in higher education: State of the art. In *Advances in Computing, Communications and Informatics (ICACCI), 2015 International Conference on* (pp. 907-913). IEEE.
- [15] Irfan, R., King, C. K., Grages, D., Ewen, S., Khan, S. U., Madani, S. A., ...& Tziritas, N. (2015). A survey on text mining in social networks. *The Knowledge Engineering Review*, 30(02), 157-170.
- [16] Berry Michael, W. (2004). *Automatic Discovery of Similar Words. Survey of Text Mining: Clustering, Classification and Retrieval*, Springer Verlag, New York, 200, 24-43.
- [17] Salloum, S. A., Al-Emran, M., & Shaalan, K. (2016). A Survey of Lexical Functional Grammar in the Arabic Context. *Int. J. Com. Net. Tech*, 4(3).
- [18] Al Emran, M., & Shaalan, K. (2014, September). A Survey of Intelligent Language Tutoring Systems. In *Advances in Computing, Communications and Informatics (ICACCI), 2014 International Conference on* (pp. 393-399). IEEE.
- [19] Al-Emran, M., Zaza, S., & Shaalan, K. (2015, May). Parsing modern standard Arabic using Treebank resources. In *Information and Communication Technology Research (ICTRC), 2015 International Conference on* (pp. 80-83). IEEE.
- [20] Alfaro, C., Cano-Montero, J., Gómez, J., Moguerza, J. M., & Ortega, F. (2016). A multi-stage method for content classification and opinion mining on weblog comments. *Annals of Operations Research*, 236(1), 197-213.
- [21] Yang, L., Geng, X., & Liao, H. (2016). A web sentiment analysis method on fuzzy clustering for mobile social media users. *EURASIP Journal on Wireless Communications and Networking*, 2016(1), 1.
- [22] Robinson, R., Goh, T. T., & Zhang, R. (2012). Textual factors in online product reviews: a foundation for a more influential approach to opinion mining. *Electronic Commerce Research*, 12(3), 301-330.
- [23] Rahmani, A., Chen, A., Sarhan, A., Jida, J., Rifaie, M., & Alhajj, R. (2014). Social media analysis and summarization for opinion mining: a business case study. *Social Network Analysis and Mining*, 4(1), 1-11.
- [24] Grimes, S. (2008). *Unstructured data and the 80 percent rule*. CarabridgeBridgepoints.
- [25] Hung, J. L., & Zhang, K. (2012). Examining mobile learning trends 2003– 2008: A categorical meta-trend analysis using text mining techniques. *Journal of Computing in Higher education*, 24(1), 1-17.
- [26] Feldman, R., & Dagan, I. (1995, August). Knowledge Discovery in Textual Databases (KDT). In *KDD (Vol. 95, pp. 112-117)*.
- [27] Rajman, M., & Besançon, R. (1998). Text mining: natural language techniques and text mining applications. In *Data mining and reverse engineering* (pp. 50-64). Springer US.

- [28] Gök, A., Waterworth, A., & Shapira, P. (2015). Use of web mining in studying innovation. *Scientometrics*, 102(1), 653-671.
- [29] Fan, W., Wallace, L., Rich, S., & Zhang, Z. (2005). Tapping into the power of text mining.
- [30] SØRENSEN, H. T., Sabroe, S., & OLSEN, J. (1996). A framework for evaluation of secondary data sources for epidemiological research. *International journal of epidemiology*, 25(2), 435-442.
- [31] Zhang, J. Q., Craciun, G., & Shin, D. (2010). When does electronic word-of-mouth matter? A study of consumer product reviews. *Journal of Business Research*, 63(12), 1336-1341.
- [32] Gupta, V., & Lehal, G. S. (2009). A survey of text mining techniques and applications. *Journal of emerging technologies in web intelligence*, 1(1), 60-76.
- [33] Zaza, S., & Al-Emran, M. (2015, October). Mining and Exploration of Credit Cards Data in UAE. In 2015 Fifth International Conference on e-Learning (econf) (pp. 275-279). IEEE.
- [34] Navathe, S. B., & Ramez, E. (2000). Data warehousing and data mining. *Fundamentals of Database Systems*, 841-872.
- [35] Akilan, A. (2015, February). Text mining: Challenges and future directions. In *Electronics and Communication Systems (ICECS), 2015 2nd International Conference on* (pp. 1679-1684). IEEE.
- [36] Sukanya, M., & Biruntha, S. (2012, August). Techniques on text mining. In *Advanced Communication Control and Computing Technologies (ICACCCT), 2012 IEEE International Conference on* (pp. 269-271). IEEE.
- [37] Piatetsky-Shapiro, G. (2007). Data mining and knowledge discovery 1996 to 2005: overcoming the hype and moving from “university” to “business” and “analytics”. *Data Mining and Knowledge Discovery*, 15(1), 99-105.
- [38] Weninger, T. (2014). An exploration of submissions and discussions in social news: mining collective intelligence of reddit. *Social Network Analysis and Mining*, 4(1), 1-19.
- [39] Chakraborty, G., & Krishna, M. (2014). Analysis of unstructured data: Applications of text analytics and sentiment mining. In *SAS global forum* (pp. 1288-2014).
- [40] Stieglitz, S., & Dang-Xuan, L. (2013). Social media and political communication: a social media analytics framework. *Social Network Analysis and Mining*, 3(4), 1277-1291.
- [41] Markovikj, D., Gievska, S., Kosinski, M., & Stillwell, D. (2013, June). Mining facebook data for predictive personality modeling. In *Proceedings of the 7th international AAAI conference on Weblogs and Social Media (ICWSM 2013)*, Boston, MA, USA.
- [42] Zhang, Y., & Yu, T. (2012). Mining trust relationships from online social networks. *Journal of Computer Science and Technology*, 27(3), 492-505.
- [43] Yan, B. N., Lee, T. S., & Lee, T. P. (2015). Analysis of research papers on E-commerce (2000–2013): based on a text mining approach. *Scientometrics*, 105(1), 403-417.
- [44] Witten, I. H. (2005). Text mining. *Practical handbook of Internet computing*, 14-1.
- [45] Steinberger, R. (2012). A survey of methods to ease the development of highly multilingual text mining applications. *Language Resources and Evaluation*, 46(2), 155-176.
- [46] Schoder, D., Gloor, P. A., & Metaxas, P. T. (2013). Special Issue on Social Media. *KI*, 27(1), 5-8.
- [47] Injadat, M., Salo, F., & Nassif, A. B. (2016). Data mining techniques in social media: A survey. *Neurocomputing*.
- [48] Kamal, S., & Arefin, M. S. (2016). Impact analysis of facebook in family bonding. *Social Network Analysis and Mining*, 6(1), 1-14.
- [49] Back, M. D., Stopfer, J. M., Vazire, S., Gaddis, S., Schmukle, S. C., Egloff, B., & Gosling, S. D. (2010). Facebook profiles reflect actual personality, not self-idealization. *Psychological science*.
- [50] Fire, M., & Puzis, R. (2012). Organization mining using online social networks. *Networks and Spatial Economics*, 1-34.
- [51] Chan, H. K., Lacka, E., Yee, R. W., & Lim, M. K. (2014, December). A case study on mining social media data. In *2014 IEEE International Conference on Industrial Engineering and Engineering Management* (pp. 593-596). IEEE.
- [52] Rahman, M. M. (2012). Mining social data to extract intellectual knowledge. *arXiv preprint arXiv:1209.5345*.
- [53] Celli, F., Pianesi, F., Stillwell, D., & Kosinski, M. (2013, June). Workshop on computational personality recognition (shared task). In *Proceedings of the Workshop on Computational Personality Recognition*.
- [54] Reips, U. D., & Garaizar, P. (2011). Mining twitter: A source for psychological wisdom of the crowds. *Behavior research methods*, 43(3), 635-642.
- [55] Hamed, A. A., Wu, X., & Rubin, A. (2014). A twitter recruitment intelligent system: association rule mining for smoking cessation. *Social Network Analysis and Mining*, 4(1), 1-19.
- [56] Dey, L., & Haque, S. M. (2009). Opinion mining from noisy text data. *International Journal on Document Analysis and Recognition (IJDAR)*, 12(3), 205-226.
- [57] Gaikwad, S. V., Chaugule, A., & Patil, P. (2014). Text mining methods and techniques. *International Journal of Computer Applications*, 85(17).

- [58] Sarker, A., O'Connor, K., Ginn, R., Scotch, M., Smith, K., Malone, D., & Gonzalez, G. (2016). Social media mining for toxicovigilance: automatic monitoring of prescription medication abuse from Twitter. *Drug safety*, 39(3), 231-240.
- [59] Al-Daihani, S. M., & Abrahams, A. (2016). A Text Mining Analysis of Academic Libraries' Tweets. *The Journal of Academic Librarianship*, 42(2), 135-143.
- [60] Abrahams, A. S., Fan, W., Wang, G. A., Zhang, Z. J., & Jiao, J. (2015). An integrated text analytic framework for product defect discovery. *Production and Operations Management*, 24(6), 975-990.
- [61] Godfrey, D., Johns, C., Meyer, C., Race, S., & Sadek, C. (2014). A case study in text mining: Interpreting twitter data from world cup tweets. arXiv preprint arXiv:1408.5427.
- [62] Velez, D., Sueiras, J., Ortega, A., & Velez, J. F. (2015). A method for K-Means seeds generation applied to text mining. *Statistical Methods & Applications*, 1-23.
- [63] Mosley Jr, R. C. (2012). Social media analytics: Data mining applied to insurance Twitter posts. In *Casualty Actuarial Society E-Forum*, Winter 2012 Volume 2 (p. 1).
- [64] Kermandis, K. L., & Maragoudakis, M. (2013). Political sentiment analysis of tweets before and after the Greek elections of May 2012. *International Journal of Social Network Mining*, 1(3-4), 298-317.
- [65] Wu, J., Sun, H., & Tan, Y. (2013). Social media research: A review. *Journal of Systems Science and Systems Engineering*, 22(3), 257-282.
- [66] Habash, N. (2010). Introduction to Arabic natural language processing. *Synthesis Lectures on Human Language Technologies*, 3(1), 1-187, Morgan & Claypool Publishers.
- [67] Shaalan, K., Abo Bakr, H., & Ziedan, I. (2007). Transferring Egyptian Colloquial into Modern Standard Arabic, *International Conference on Recent Advances in Natural Language Processing*, PP. 525-529, Bulgaria.
- [68] Farghaly, A., Shaalan, K. (2009). Arabic Natural Language Processing: Challenges and Solutions. *ACM Transactions on Asian Language Information Processing (TALIP)*, ACM, 8(4), 1-22.