DOI 10.12694/scpe.v20i2.1513
Scalable Computing: Practice and Experience ISSN 1895-1767
Volume 20, Number 2, pp. 223–235. http://www.scpe.org

ONTOLOGY DRIVEN SOCIAL BIG DATA ANALYTICS FOR FOG ENABLED SENTIC-SOCIAL GOVERNANCE

AKSHI KUMAR* AND ABHILASHA SHARMA†

Abstract. Conventional e-government has many practical infrastructure development and implementation challenges. The recent surge of SMAC (Social media, Mobile, Analytics, Cloud) technologies re-defines the e-governance ecosystem. Cloud-based e-governance has numerous operational challenges which range from development to implementation. Moreover, the contemplation and vocalization of public opinion about any government initiative is quintessential to be cognizant of how citizens perceives and get benefitted from cloud/fog enabled governance. This research puts forward a semantic knowledge model for investigating public opinion towards adaption of fog enabled services for governance and comprehending the significance of two s-components (sentic and social) in aforesaid structure that specifically visualize fog enabled Sentic-Social Governance. The results using conventional TF-IDF (Term Frequency-Inverse Document Frequency) feature extraction are empirically compared with ontology driven TF-IDF feature extraction to find the best opinion mining model with optimal accuracy. The results depict that the implementation of ontology driven opinion mining for feature extraction in polarity classification outperforms the traditional TF-IDF method validated over baseline supervised learning algorithms. An average of 7.3% improvement in accuracy and approximately 38% reduction in features has been reported.

Key words: Fog Computing, Government Intelligence Cloud , Opinion Mining, Feature Selection, Ontology

AMS subject classifications. 68M14, 97R50, 91D30

1. Introduction.

Government-to-Citizen (G2C) is a concept [1] that speaks about the connectivity between public administration organizations and citizens of a country. The relationship specifies ICT (Information and communication technologies) based solution that streamlines the interaction and association between governing bodies and citizens. The ultimate goal of G2C [2] governance is to serve the society by offering various ICT services with the use of cutting edge technologies in a more productive and profitable way. It also plays up to increase people participation in governance both in terms of quantity and quality. G2C interactions empower citizens by apprising them with the government policies, practices, rules, regulations, strategies and services. Several countries have developed their respective framework of G2C technology for enhancing public participation in governmental proceedings.

Digital India [3] is a flagship programme launched by Government of India to transform the nation into a digitally empowered society and knowledge economy. The objective is to explore innovative ideas and practical solutions by exploiting digital technologies such as cloud computing, mobile applications, Internet of Things (IoT) etc. It is a "citizen-centric" campaign that concentrates on three prime parameters [4]: digital infrastructure as a utility to every citizen, governance and services on demand, and digital empowerment of citizens. The initiative includes coupling of multiple government ministries and departments with several thoughts and ideas transforming into a single comprehensive vision but its practical implementation is very demanding and challenging. Large development and implementation cost, lack of internet accessibility, lack of computing skills among citizens, insufficient technology requirements, requisite of email addresses, lack of privacy are certain issues [5] that makes it difficult for governing bodies to reach out citizens. The nine growth pillars of Digital India [6] that contribute towards the economic and electronic dissemination of government information in public are represented in figure 1.1. Several policies, schemes, campaigns, and initiatives are undertaken by different government officials with the purpose to reinforce these growth pillars. Thus, the holistic development of these growth pillars is the need of the hour.

Electronic governance or e-governance [7], an application of ICT and framework for G2C communication, plays a significant role in information exchange of government services to citizens. It is an electronic execution of governance [8] that facilitates a simple, rapid, efficient and highly transparent process of information broadcasting, delivery of government services in order to promote good governance. It signifies the reformation of government with the implementation of technology in government processes and functions. This technological revolution in governance can modernize the modus operandi of any society.

*Department of Computer Science & Engineering, Delhi Technological University, Delhi-42, India (akshi.kumar@gmail.com).
†Department of Computer Science & Engineering, Delhi Technological University, Delhi-42, India (abhi16.sharma@gmail.com)

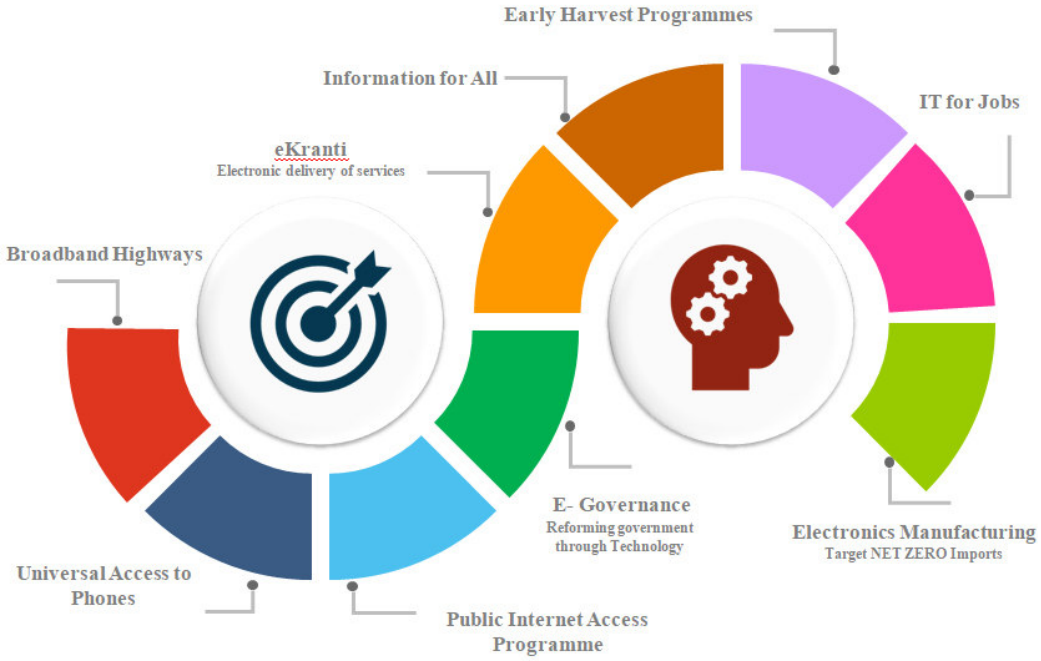

FIG. 1.1. *Nine Growth Pillars of Digital India [6]*

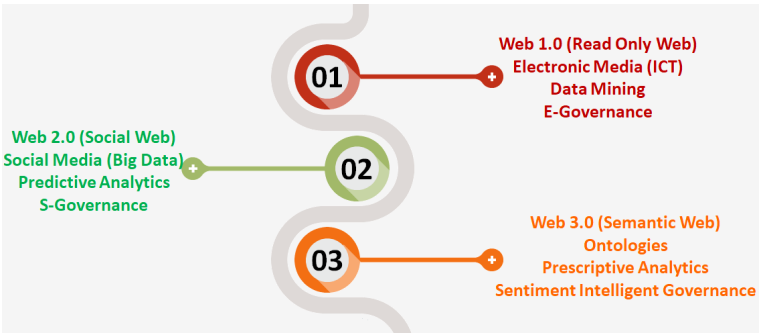

FIG. 1.2. *Evolution of Media, Analytics and Governance with Web*

Digital information and communication technologies have come forth as wave maker that can transform an entire economy, build up smart citizens and realize a well established form of socio-political organizations across globe. In view of this, convergence of four technologies namely Social media, Mobile, Analytics, Cloud (SMAC paradigm) [9] is the driving force of governance ecosystem. A voluminous amount of data (big data) has been generated through various sources of *social media* and *mobile* applications that require optimized techniques for data *analytics*. Cloud computing characterizes the on-demand delivery of computing services over the internet. This virtualization technology offers self-service capability, scalability, flexibility, and affordability. The evolving web has changed the traditional model of governance into digital governance model turning the [10] paradigm shift from e-governance to s-governance (social governance). It also turns the corners of social media as well as data analytics. The evolution timeline of all four factors has been depicted in figure 1.2.

Cloud computing [11] makes a big shift from the conventional methods of governance due to various reasons such as infrastructure cost reduction, quick provision of computing resources, scalability in terms of delivering right amount of IT services, improves productivity and performance, procuring more secure environment and much more. It proffers a simple way to access servers, storage, databases and a broad set of application services

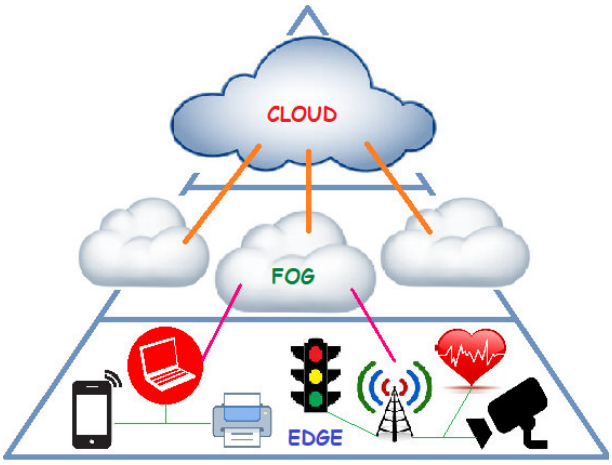

FIG. 1.3. *Big Data Processing Layer Stack [13]*

over the Internet. Cloud computing, the name is so because information accessed is found in "the cloud" and provides user a anywhere, anytime access. Fog computing, dew computing, cloudlet and edge computing are extended variants that lies at different levels in hierarchy of cloud computing architecture. Figure 1.3 represents how the [12] storage, accessibility and maintenance of huge amount of data i.e. big data available over web is being maintained and processed by cloud, fog and edge computing.

Fog computing brings the benefits, services and power of cloud to the edge of the network. It has closer proximity to end users i.e. performing short term data analytics at the edge and larger geographical distribution. It acts as an interface between cloud and edge device layers [14] deciding what data needs to be pushed to cloud and what needs to be analyzed locally at the edge. The goal of fogging is to improve efficiency, reducing data transportation to cloud, security and compliance. All these factors put accent on the adoption of fog layered architecture for adapting governance infrastructure. The role and need of cloud/fog based governance has been studied across literature. Various literary resources are available online which discuss the practical and potential application of cloud in the governance ecosystem. Certainly, fog enabled governance is the non-trivial buzzword within the contemporary paradigm for government intelligence. *MeghRaj: the national cloud* is one such the government intelligence cloud initiative [15] launched by Government of India.

The contemplation and vocalization of public opinion about any governmental processes or practices is quintessential as they are entitled to enjoy all the legal rights and privileges granted by the government. It is equally important to be cognizant of how citizens perceives and get benefitted from cloud/fog enabled governance. Consequently, addition of one more s-factor (sentiment based) in s-governance reforms the existing model into a *sentic-social* governance (S-governance) with a view to make it more sustainable. Opinion mining facilitates this implementation and formalization of the S-governance model. It is referred to as the computational study [16,17] to extract and analyze public opinion/sentiment about any topic, subject, event or entity for better decision making process by applying various intelligent techniques over a large volume of user generated data (social big data) over web. The use of different intelligent learning techniques such as machine learning, lexicon-based, hybrid and concept based has been reported across pertinent literature studies [18] within the domain. As an attempt to comprehend public opinion on the induction and adoption of cloud-based governance model, this research put forwards an optimized predictive learning model based on real-time data.

The vital sub-task of the polarity classification (opinion polarity: positive, negative, neutral) process is feature extraction, which converts the input data (unstructured textual data indicative of opinion), into an array of representative features. Commonly, the feature extraction task is done using intrinsic filtering methods which are fast, classifier-independent methods that rank features according to predetermined numerical functions based on the measure of the importance of the terms. A variety of scoring functions such as, tf-idf, chi-square, mutual information, information gain, cross-entropy etc., have been used as statistical measures to pick features with the highest scores [19].The accuracy of the classifier strongly depends on the selection of high quality

data features that is the training dataset. Moreover, the training sets are typically prepared manually. Past literature conforms that an optimal feature selection [20] improves the classifier performance (in terms of speed, predictive power and simplicity of the model), reduces dimensionality, removes noise and helps visualizing the data for model selection. In feature selection the features are kept intact and n best features are chosen among them, removing the redundant and co-linear features. This sub-task of selecting the relevant subset of features and discarding the non-essential ones is computationally challenging and expensive task. Motivated by these issues, in this research we propose a semantic knowledge based polarity classification process.

An ontology [21,22] which is specification of conceptualization is utilized as a filtering method for finding important as well as hidden features. Ontologies are primarily tools which can drive the feature engineering process by:

- Structuring semantic information as concepts, properties, instances and hierarchies for feature identification.
- Extracting explicit features to build the feature space using relationship between concepts.
- Uncover important features using concept hierarchy defined by the ontology.

Hence, to optimize the feature space without sacrificing remarkable classification accuracy in this work, we put forward an intelligent ontology based data analytics solution for opinion prediction in social big data concerning fog enabled governance. The proffered solution is put to test for the sentiment classification tasks on tweets pertaining to *"Meghraj: The National Cloud"*, a government intelligence cloud initiative launched by Government of India. The conventional classification process is done using TF-IDF (Term Frequency-Inverse Document Frequency) feature extraction method on the cleaned dataset. Five supervised machine learning classifiers namely Nave Bayesian (NB), Support Vector Machines (SVM), Multilayer Perceptron (MLP), k-Nearest Neighbour (k-NN), and Decision Tree (DT) are empirically compared. Ontology based feature optimization is then performed to semantically analyze the concept and make rise in reusability, interoperability, knowledge acquisition and its verification. Thus, the contribution of this research is to build an optimal opinion mining model as follows:

- To implement five supervised learning algorithms to classify opinion polarity using tf-idf feature extraction: NB, DT, SVM, kNN & MLP
- To build a domain ontology for optimal feature extraction: Domain Ontology for Meghraj (DOM)
- To implement five supervised learning algorithms using ontology guided feature extraction method: TF-IDF on ontological features
- Performance analysis on the basis of efficacy measures

The objective of this paper is to implement and evaluate an opinion mining model for analysing public opinion on government cloud initiative. It is characterized as a semantic knowledge (ontology) based model of fog enabled services offered by government and consequently comprehends the significance of two s-components (sentic and social) within the Fog enabled Sentic-Social Governance.

The subsequent sections are lined up as follows: Section 2 discusses the background work related to fog enabled services in governance and ontology driven opinion mining. Section 3 abstracts the information about Meghraj as a techno typhoon for government prosperity and its scope in the landscape of Indian governance. Section 4 explicates the proffered model for ontology driven opinion prediction of fog enabled governance It also illustrates the dataset details and implementation process. Section 5 determines the opinion classification of the chosen concept as per tweets polarity. It substantiates the ontological model by comparing classifier performance for optimal feature extraction with baseline supervised learning techniques to analyse public opinion about the program. At last, section 6 sums up the inferences drawn from the results and discusses the future work as an open scope of this research work.

2. Related Literature Review of Cloud/Fog enabled Services for Digital Governance. Conventional e-governance faces many challenges in terms of cost, software and hardware requirements, network, security, business & policy adoption and implementation etc. Recent years have shown cloud computing as a technology to solve these problems. Pokharel et al. [23] proposed cloud computing as the future solution for e-governance. Mukherjee et al. [24] put forward a future framework for e-governance based on cloud computing consisting of three layers. Sharma et al. [25] enlisted the applications of cloud computing in e-governance. Work done by authors Cellary et al. [26] discussed about cloud computing and service-oriented architecture for

e-governance. In the research done by Yeh et al. [27], cloud computing has been used in e-governance to change its function towards service, push forward the green technology and promote industrial upgrading. Author Rastogi [28] proposed a model-based framework to implement cloud computing in e-governance. Tripathi & Parihar in 2011, Alshomrani & Qamar in 2013 [29,30] analysed cloud computing and its applications in e-governance and explained how cloud may lead to cost savings.

The term, *Opinion Mining*, also known as sentiment analysis or sentiment mining, was initially witnessed in the published work by Dave et al. [31] in 2003 and since then both primary and secondary studies have been reported across pertinent literature [18,32,33]. As discussed, various techniques have been used to perform the task of opinion mining namely: machine learning, concept based, lexicon based, and hybrid. Pertinent literature reveals [18] various application areas of opinion mining namely, business intelligence, information and security analysis, market intelligence, sub component technology, government intelligence, smart society services etc. Government intelligence (GI) has been the least explored application area with only few relevant studies done within the domain, which include GI using concept based opinion mining, lexicon-based opinion mining , hybrid approach , and opinion mining using machine learning. Recently, Kumar & Sharma [34, 35] proposed few models/frameworks for opinion mining in the different application area of digital governance. To the best of our knowledge, no related work on mining public opinion to understand adoption and acceptance process of a cloud based e-governance initiative has been done.

Feature extraction primarily transforms raw data into features suitable for modelling. For example textual features include n-grams, word2vec and TF-IDF etc [36]. The training set is high-dimensional and the classifiers accuracy strongly depends on the quality of hand-crafted features. To deal with the dimensionality of training feature sets for improved classifier performance, various techniques have been successfully applied. Feature selection [37] and dimensionality reduction [38] approaches have been used to select features with the highest "importance"/influence on the target variable, from a set of existing features. The higher the number of features, more challenging and computationally expensive it gets to visualize the training set. Ontology guided feature extraction can thus be used as a feasible and to optimize the feature space representative of the training dataset. Domain ontologies can be used in selection of features for improved feature-based opinion mining. Pealver-Martnez et al. [39] have discussed the concept of ontological feature based opinion mining based on recent studies. The two major dimensions were given attention namely, opinion classification and feature based opinion mining. They reviewed the improvement process of feature level opinion mining by incorporating the concept of ontology. Also, a proposed approach for ontology-supported polarity mining (OSPM) has been examined for the sake of enhancing the process of opinion classification.

3. *Meghraj: The National Cloud* - An Initiative by Indian Government for Proliferation of Fog/Cloud in Governance. In May 2016, National e-Governance Plan (NeGP) [40] was initiated by Government of India for better accessibility of all government services to citizens. The vision is to access these services in the citizens own locality in a cost-effective manner via common service delivery outlets. As a step towards this different states have placed and setup their respective ICT infrastructure. The deployment and installation of application software has been done through assistance and funding provided by central government. Although, the whole process has been outsourced but still the entire endeavour was strapped for time due to inadequate application development initiatives, detailed and exhaustive procurement process, shortage of in-house experts for managing large procurements etc. The government was looking forward for a common shared platform to gain momentum in the implementation process. This infused the need for the adoption of a government private cloud environment that would expedite the ICT enabled service enrichment process with an affordable cost in and among various government departments at state or central level. Figure 3.1 represents the hierarchy of adoption of cloud [41] in Indian governmental structure.

Meghraj, the GI cloud initiative [15,42] was launched by Government of India under the Department of Electronics and Information Technology (DeitY), Ministry of Communications and Information Technology. The major objective(s) is to (a) accelerate the development and deployment of eGov applications (b) ensure optimal use of resources and infrastructure (c) make standardized and certified applications available (d) replicate successful applications in order to reduce effort, time and cost across states and much more. The functional flow of GI cloud system [42] is illustrated in figure 5. The architecture of GI cloud follows specific standards, guidelines and a set of protocols issued by Government of India. The task of GI cloud services directory is

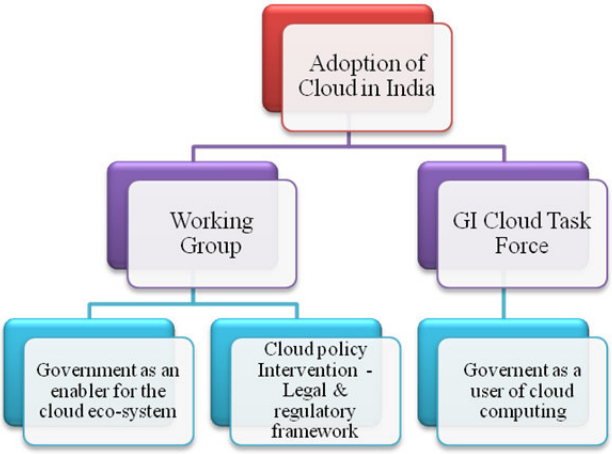

FIG. 3.1. *The National Cloud Initiative [41]*

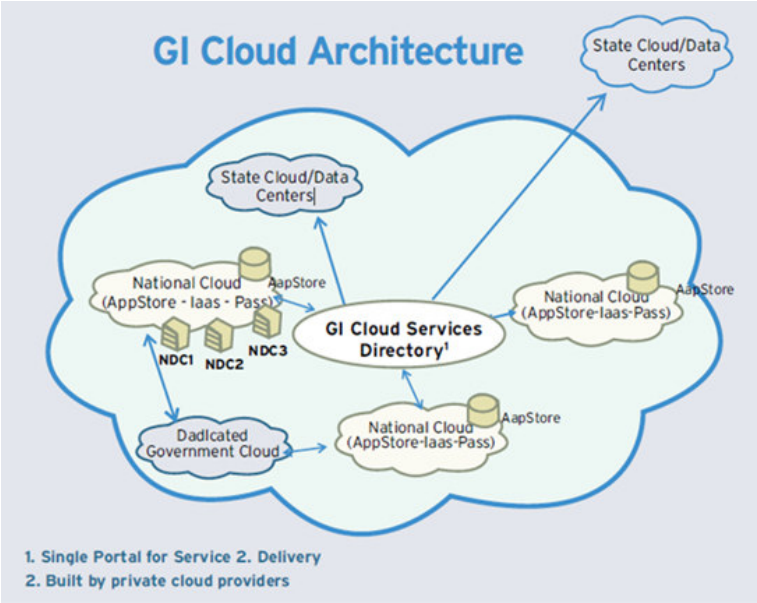

FIG. 3.2. *GI Cloud System Flow [42]*

to publish the services proffered by GI cloud. It comprises of discrete and multiple private cloud computing environments at national and state levels with a view to provide high level blueprint of ICT enabled government functions. The architecture allows the state data centres (SDCs) or state clouds of different states to associate with GI cloud either by acting as independent cloud environments or by lending their IT Infrastructure as part of GI Cloud but these states can have their own data centres as well that exists outside the GI cloud environment. All such states are motivated to jump on GI cloud resources as and when they exhausted their own. SDCs of 21 states have been made operational and running numerous applications such as e-Procurement, Bhoomi, commercial tax, mandi board etc. SDCs of 4 states are in advanced stage of implementation and cloud adoption.

National Cloud was launched in February 2014 under Meghraj initiative and implemented by National Informatics Centre (NIC) [43] offers various services: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), Software as a Service (SaaS) and Storage as a Service (STaaS). Table 3.1 enlists various potential

TABLE 3.1
Key drivers and issues of GI cloud

Benefit(s)	Hazard(s)
Optimal usage of available resources and infrastructure	Application design approaches and cloud standards
Rapid deployment and Reusability	Intensive change management and lack of skilled resources
Security, Scalability, Manageability and Maintainability	Vendor lock-in and data location
Reduced time, effort and cost in managing technology	Loss of portability and control
Increased standardisation and user mobility	Licensing and funding model

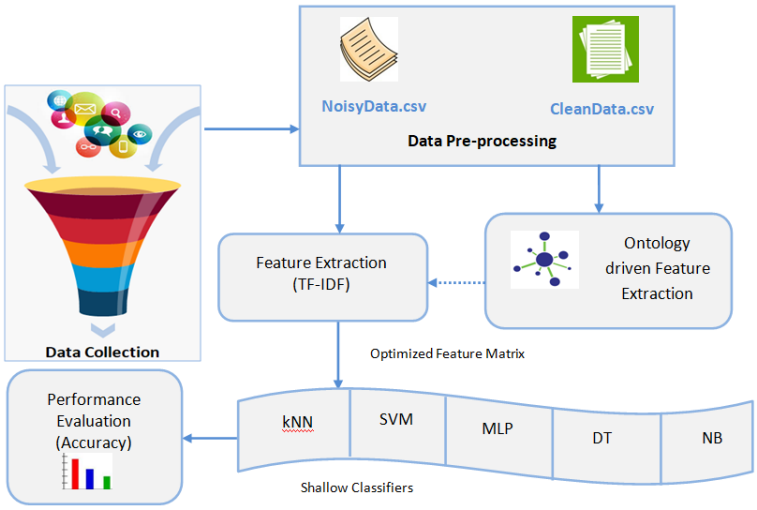

FIG. 4.1. *System Architecture of Ontology based Opinion Prediction Model*

benefits and risks of GI cloud.

There are three adoption phases for the establishment of GI cloud [44] namely, (i) Strategy, policy and guidelines establishment (ii) Implementation and (iii) Monitoring, management and ongoing improvement. The first phase focuses over the formation of practices, strategies, rules and guidelines in order to prepare a development plan. Implementation phase executes the prepared program in a well planned manner and the last phase monitors the present workings, evaluates the shortfalls and works towards continuous improvements. In this paper, the evaluation of GI cloud initiative has been done by capturing opinionated response of citizens for predictive analytics. Domain Ontology has been built to model the feature space and it is optimized using the conventional statistical feature extraction TF-IDF filter. The proposed framework is proposes in the following section.

4. Ontology based Opinion Predictive Analytics. The proposed framework comprises of four phases namely, data collection, data pre-processing, ontology driven optimal feature extraction and classification. Figure 4.1 depicts the system architecture of ontology based opinion prediction model.

4.1. Data Acquisition. Collection of data is the foremost step in the process of opinion mining. Various social media portals are available for capturing real-time data of public perception about any topic, subject, event or area of concern. Analysis of real-time data sets relevant to any phenomena provides more accurate results or evaluation of the system and hence improves the overall performance. Twitter, being the most popular micro blogging site [45] has been used in this research work to capture the public sentiments about Meghraj. It is one of the most prominent social media platforms exercised by government for facilitating

TABLE 4.1
Weekly record of tweets collected

Duration	Tweet Count	Duration	Tweet Count
Week 1	192	Week 4	328
Week 2	194	Week 5	160
Week 3	402	Week 6	140

TABLE 4.2
Tweets Distribution based on Polarity

Week	N	Nu	P	Total
Week 1	29	14	149	192
Week 2	23	16	155	194
Week 3	19	76	307	402
Week 4	11	65	252	328
Week 5	4	79	77	160
Week 6	1	85	54	140
Total	87	335	994	1416

direct government citizen interaction. A huge volume of users share their views, ideas, suggestions and update themselves with the information posted on twitter. The diversity of opinion in various public communities attributed to their regional, cultural, social, economical and educational backgrounds is considered in order to obtain a realistic view of the system. Provision of various APIs (Application Program Interfaces) accelerate the process of tweets extraction over a specific topic using #(hash tag) leading with the topic name(#topicname). The process of tweet extraction has been done by running scripts using an application developed in python. The search query comprises and executes with various hash tag words such as #meghraj, #gicloud, #meghrajcloud, #nationalcloud. A .csv file has been created with the tweets relevant to #topicname as a result of search query. Tweets associated with hashtags for national cloud have been collected near the launch date of program for a duration 6 weeks. A count of 1416 tweets have been gathered in this duration in order to figure out the public inclination about the program. Table 4.1 enlists the weekly status of tweets collected.

The tweets are classified into three polarity categories namely, positive, negative and neutral. Table 4.2 reflects the tweets distribution based on their opinion polarity.

For the selected duration, MeghRaj able to capture a count of 1416 tweets that consists of 950 positive tweets, 335 neutral tweets and 126 negative tweets. A graphical depiction of the aforesaid stats of weekly polarity distribution of tweets is represented in figure 4.2. Approximately 70% of tweets were positive, whereas the percentage of neutral and negative tweets was 23.6% and 6.1% respectively. Deprecation in the count of positive tweets has been reported after week 3 whereas an average rise of tweet count can be observed in neutral tweets after week 2. However, there is a continuous deterioration in the frequency of negative tweets. Week 3 was reported as a critical week being the maximum number of positive tweets whereas week 6 reflects the maximum number of neutral tweets. The percentage of negative tweets was high in the initial phases due to certain technical limitations in implementing procedure of this scheme. Also, the gradual increase in the neutral tweets by the time is due to the rapid occurrence of informational tweets posted by government, allied agencies, and media or civic. This information was all about the campaign features, functionality, latest updates and future processing's and hence making the polarity of tweet neutral. Python packages like Scipy, nltk, Numpy, Scikit-learn etc., machine learning and python scripts (version 3.6) have been used for implementation.

4.2. Pre-processing of data. The outcome of data collection is a .csv file that contains data with a lot of noise. The second phase, data pre-processing, is another essential step performed in order to transform the existing file into a new one for feature selection by cleansing the collected data. This conversion of NoisyData.csv file to CleanData.csv file and the pre-processing procedure includes following steps:

- Removing tweets replication, stop words, number in tweets with placeholders, mentions etc.
- Replacement of URLs.
- Eliminating figures, special characters such as , #.
- Natural Language tool-Kit (NLTK) [46,47] for tokenization.

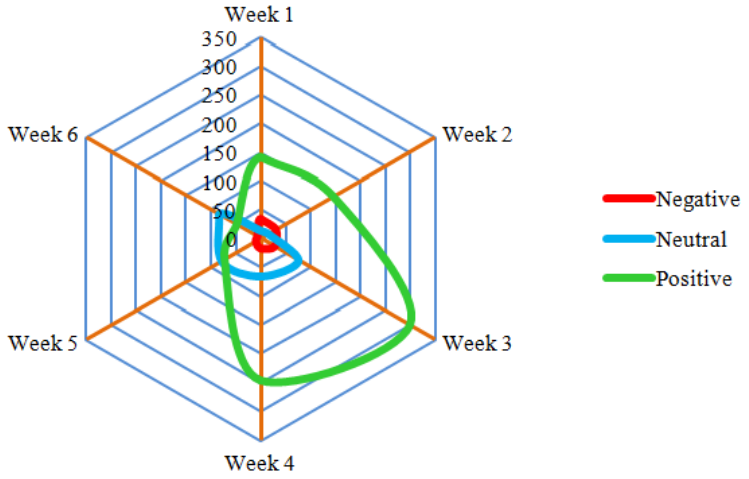

FIG. 4.2. *Polarity distribution of tweets on weekly basis for Meghraj*

- Porter's stemmer [48,49] for stemming to the root word.
- Removal of non-ASCII English character.

This cleansed data is then utilised for feature extraction. Some keywords for the selected attributes are: projects, e-services, IT giants, operational, future computing, cloud solution, departments, deployment, infrastructure, server, administration etc.

4.3. Feature Extraction. Feature extraction is one of the critical and complex tasks in opinion mining. The objective is to recognize the entity (person, service or an object) that is being referred in opinion. The automation of the process of feature identification in opinion analysis with the use of NLP (Natural language processing) techniques makes it harder to comprehend. In this paper, the two methods used for feature extraction are as follows:

- **Conventional Feature Extraction (TF-IDF based)**
  TF-IDF stands [50,51] for Term Frequency - Inverse Document Frequency. It is a weight statistically measured to evaluate the importance of a word to a document in a corpus. The importance of a word increases as its frequency increases in a document but is offset by the frequency of word in corpus. The Term Frequency, TF(t, d) simply counts the frequency of a term in a document as follows:

$$TF(t,d) = \left(\frac{No.\ of\ times\ term\ t\ appears\ in\ a\ document\ d}{Total\ no.\ of\ terms\ in\ the\ document} \right)$$

  The Inverse Document Frequency, IDF (t, D) checks whether the word is rare or common in the corpus so as to measure how much information is provided by a specific word. It is calculated as follows:

$$IDF(t,D) = log_e \left(\frac{Total\ no.\ of\ documents}{No.\ of\ documents\ with\ term\ t\ in\ it} \right).$$

  Thus, TF-IDF is calculated as:

$$TF - IDF(t,d,D) = TF(t,d) * IDF(w,D)$$

  where t denotes the terms; d denotes each document and D denotes the collection of documents.

- **Optimal Feature Extraction (Ontology driven TF-IDF)**
  Ontology is specifically defined as [22,52] a conceptual reference model that describes the semantics of a system or domain. It represents the relationship between concepts; both in human comprehensible

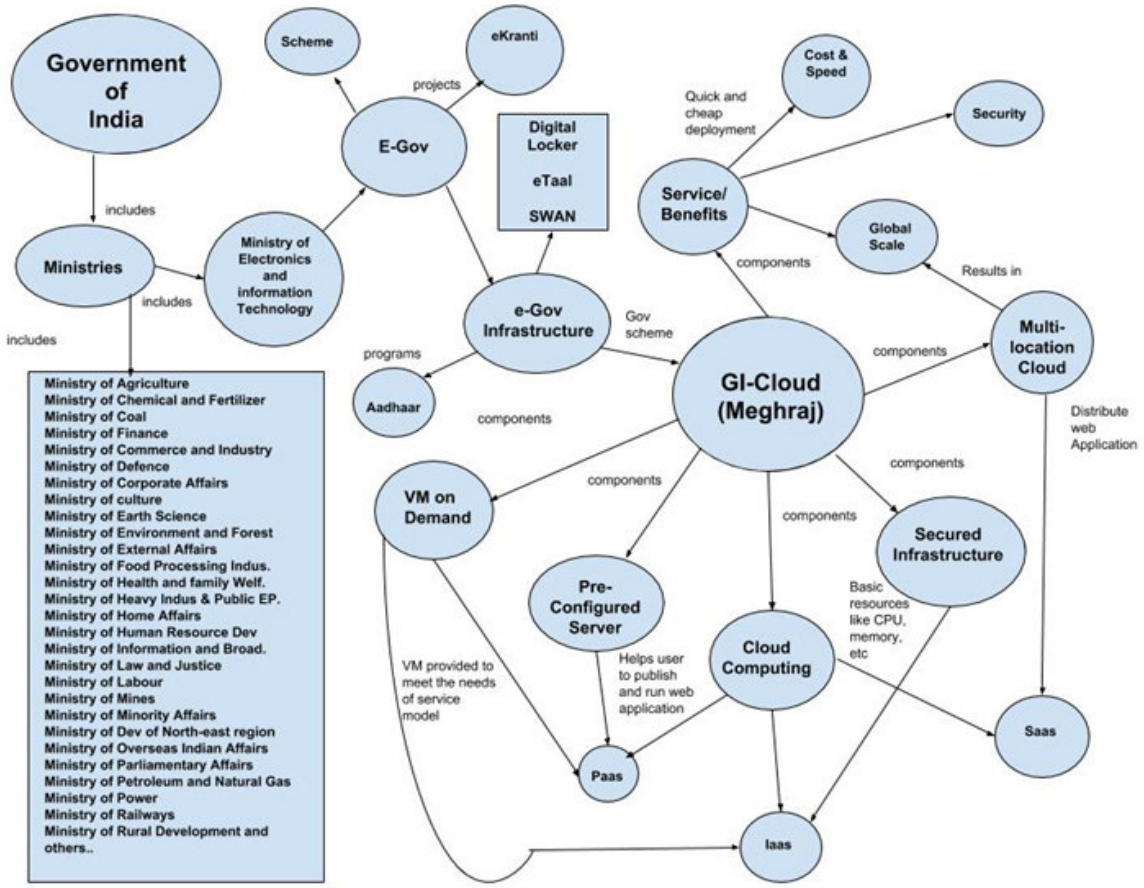

FIG. 4.3. *Domain Ontology of Meghraj (DOM): An initiative towards fog enabled governance*

and machine processable manner. It represents a concept or categories of a particular subject area that exhibits the characteristics and relationship between them. The ontological representation of the concept of Meghraj (DOM - an initiative taken by government of India to provide fog enabled services for a sentic-social facet of governance) is illustrated in figure 4.3 that represents the existing entities, how they are grouped and related in a hierarchy and sub classified based on their similitude & dissimilarity. The ontology represents various ministries that come under the umbrella of government of India. Among those, Ministry of Electronics and Information Technology provides the facility of electronic delivery of services i.e. e-Governance to facilitate citizens. Various components that comprise e-Governance are policies; projects (such as e-Kranti) and infrastructure (include programs such as aadhar, services such as digital locker, eTaal, SWAN etc., schemes). Meghraj, the first Indian GI cloud is one of the prime governmental schemes enhancing the e-Governance infrastructure. Different sub components that combines to form Meghraj are cloud computing, secured infrastructure, multi-location cloud, pre-configured server and VM (virtual machine) on demand. The figure depicts the sub components of cloud computing, elements of services offered by MeghRaj and the inter-relationship between these components along with remaining elements of ontology.

4.4. Opinion Polarity Classification. In this phase, polarity of opinion is classified into three pre-defined categories namely, positive, negative and neutral. The optimal feature set generated in the earlier step are used to build the training and testing sets of the classifier. In this paper, five supervised learning based classifiers namely, Support Vector Machine (SVM), Decision Trees (DT), Nave Bayesian (NB), k-Nearest Neighbour (k-NN) and Multi-Layer Perceptron (MLP) have been implemented. All the machine learning techniques are

TABLE 4.3
Accuracy obtained using conventional and optimal feature extraction

Technique Used	Conventional Approach (TF-IDF) Accuracy (%)	Optimal Approach (Ontology + TF-IDF) Accuracy (%)	Increase In Accuracy (%)
NB	82.2	91.9	9.7
DT	89.8	96.6	6.8
SVM	92.6	98.5	5.9
k-NN	91.3	97.7	6.4
MLP	90.2	98.3	8.1

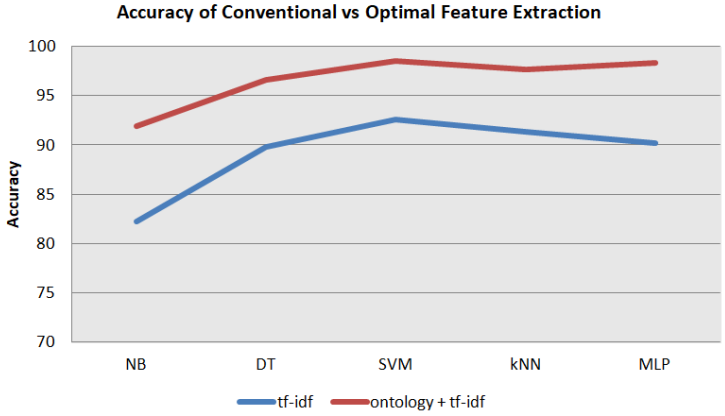

FIG. 4.4. *Accuracy obtained using conventional and optimal approach*

described in detail [34,53] across relevant literature.

4.5. Results and Discussions. This section discusses about the performance of classifiers for selected machine learning techniques (NB, DT, SVM, kNN & MLP) based on accuracy. It also compares and contrasts the difference between conventional feature extraction (using TF-IDF) and optimal feature extraction (ontological TF-IDF) techniques. The results along with percentage increase in accuracy have been listed in Table 4.3.

Result states that the best accuracy with conventional feature extraction over collected data set has been obtained by Support Vector Machine (SVM) algorithm with 92.6%. Next is k-Nearest Neighbour (kNN) with a classification accuracy of 91.3% followed by Multi-Layer Perceptron (MLP) with 90.2% and Decision Trees (DT) with 89.8. Amongst all, Naive Bayesian (NB) attained the lowest accuracy of 82.2%. The best accuracy using ontology driven feature extraction is achieved by SVM, i.e. 98.5% followed by MLP with 98.3%. k-NN and DT have occupied the next level with 97.7% and 96.6% respectively. Amongst all, NB showed the lowest accuracy of around 91.9%. The graphical comparison is represented in figure 4.4.

The maximum accuracy gain was obtained by NB (9.7%) followed by MLP (8.1%) while SVM, k-NN showed an appreciable gain in accuracy (approximately 6%). The average accuracy gain is of 7.38%.

The count of features selected in both the approaches is listed in Table 4.4. In conventional approach all the classification algorithms used the same number of features (866). After applying ontology for feature extraction the minimum number of features selected was 507 (SVM) which is 58.5% selection and maximum was 574 (NB) which is 66.2% selection. The Table 4.4 reflects an average of 62.02% features was selected.

5. Conclusion. User-generated big data from online social portals is a goldmine for extracting and analyzing stance and opinion. This knowledge discovery framework within the unstructured web data setting defines a novel socially aware and sentiment driven governance. Based on this, the work proffered in this research made two primary contributions to evaluate the response of citizens towards government initiatives, schemes or policies. Firstly, the role of social media analytics in government intelligence was investigated. Public opinion in tweets for the national cloud initiative of Government of India was examined. This opinionated information

TABLE 4.4
Number of features obtained using conventional and optimal approach

Technique	Conventional Approach (TF-IDF) #Features (%)	Optimal Approach (Ontology + TF-IDF) #Features (%)	Features Selected (%)
NB	866	574	66.2
DT	866	526	60.7
SVM	866	507	58.5
k-NN	866	547	63.1
MLP	866	534	61.6

can be an imperative phase in a government stratagem for public policy evaluation. Secondly, the learning for the predictive model of opinion mining was driven using optimal semantics-driven feature space generation. Domain ontology for Meghraj (DOM) was built and used for feature extraction with the intrinsic TF-IDF filtering method. The results demonstrated an average accuracy gain of 7.3%. SVM outperformed all the other classifiers (NB, DT, k-NN, MLP) for both conventional and ontology-drive model training. The use of ontology also built an optimal feature space automatically with only 62% of the features selected.

REFERENCES

[1] E-GOVERNANCE, https://en.wikipedia.org/wiki/E-governance
[2] WHAT IS GOVERNMENT-TO-CITIZEN (G2C), https://www.igi-global.com/dictionary/government-to-citizen-g2c/12392
[3] INTRODUCTION OF DIGITAL INDIA, https://digitalindia.gov.in/content/introduction
[4] VISION OF DIGITAL INDIA, https://digitalindia.gov.in/content/vision-and-vision-areas
[5] DIGITAL INDIA, https://www.mygov.in/group/digital-india/
[6] HOW DIGITAL INDIA WILL BE REALIZED: PILLARS OF DIGITAL INDIA., https://digitalindia.gov.in/content/programme-pillars
[7] DAWES, S. S, *The evolution and continuing challenges of egovernance*, Public Administration Review., 68 (2008), S86-S102.
[8] LEE-GEILLER, S. AND LEE, T. D., *Using government websites to enhance democratic E-governance: A conceptual model for evaluation.*, Government Information Quarterly., (2019)
[9] SMAC - THE PARADIGM SHIFT - CREATING FUTURE OF THE ENTERPRISE., https://home.kpmg/in/en/home/insights/2014/09/smac-theparadigmshift.html
[10] KUMAR, A. AND SHARMA, A., *Paradigm shifts from e-governance to s-governance*, The Human Element of Big Data: Issues, Analytics, and Performance., 213(2016)
[11] ARMBRUST, M., FOX, A., GRIFFITH, R., JOSEPH, A. D., KATZ, R., KONWINSKI, A., ET AL., *A view of cloud computing.*, Communications of the ACM., 53(4) (2010), 50-58
[12] VAQUERO, L. M. AND RODERO-MERINO, L., *Finding your way in the fog: Towards a comprehensive definition of fog computing.*, ACM SIGCOMM Computer Communication Review, 44(5)(2014), 27-32.
[13] FOG COMPUTING VS EDGE COMPUTING., https://erpinnews.com/fog-computing-vs-edge-computing
[14] GUPTA, H., VAHID DASTJERDI, A., GHOSH, S. K. AND BUYYA, R, *iFogSim: A toolkit for modeling and simulation of resource management techniques in the Internet of Things, Edge and Fog computing environments.*, Software: Practice and Experience, 47(9) (2017), 1275-1296.
[15] ABOUT MEGHRAJ., https://cloud.gov.in/about.php
[16] LIU, B., *Sentiment analysis and opinion mining.*, Synthesis lectures on human language technologies, 5(1) (2012), 1-167.
[17] PANG, B. AND LEE, L., *Opinion mining and sentiment analysis.*, Foundations and Trends in Information Retrieval, 2(12) (2008), 1-135.
[18] KUMAR, A. AND SHARMA, A., *Systematic Literature Review on Opinion Mining of Big Data for Government Intelligence.*, Webology, 14(2) (2017).
[19] CHANDRASHEKAR, G. AND SAHIN, F., *A survey on feature selection methods.*, Computers & Electrical Engineering, 40(1) (2014), 16-28.
[20] KUMAR, A., KHORWAL, R. AND CHAUDHARY, S, *A survey on sentiment analysis using swarm intelligence.*, Indian Journal of Science and Technology, 9(39) (2016).
[21] GRUBER, T., *Ontology.*, Encyclopedia of database systems.1963-1965.
[22] MAEDCHE, A. AND STAAB, S., *Ontology learning for the semantic web.*, IEEE Intelligent systems, 16(2) (2001), 72-79.
[23] POKHAREL, M. AND PARK, J.S., *Cloud computing: future solution for e-governance.*, In Proceedings of the 3rd international conference on Theory and practice of electronic governance (pp. 409-410) (2009). ACM.
[24] MUKHERJEE, K. AND SAHOO, G., *Cloud computing: future solution for e-governance.*, nternational Journal of Computer Applications, 7(7) (2010), pp.31-34.
[25] SHARMA, M.K. AND THAPLIYAL, M.P., *G-Cloud(e-Governance in Cloud).*, International Journal Engg. TechSci, 2(2) (2011), pp.134-137.

[26] Cellary, W. and Strykowski, S., *E-government based on cloud computing and service-oriented architecture.*, In Proceedings of the 3rd international conference on Theory and practice of electronic governance (2009)(pp. 5-10). ACM.
[27] Yeh, C., Zhou, Y., Yu, H. and Wang, H., *Analysis of E-government service platform based on cloud computing.*, In Information Science and Engineering (ICISE), 2010 2nd International Conference (2010) (pp. 997-1000). IEEE.
[28] Rastogi, A., *A model based approach to implement cloud computing in e-Governance.*, International Journal of Computer Applications, 9(7) (2010), pp.15-18.
[29] Tripathi, A. and Parihar, B., *E-governance challenges and cloud benefits.*, In Computer Science and Automation Engineering (CSAE), 2011 IEEE International Conference on (Vol. 1, pp. 351-354). IEEE. 2011, June
[30] Alshomrani, S. and Qamar, S., *Cloud based e-government: benefits and challenges.*, International Journal of Multidisciplinary Sciences and Engineering, 4(6) (2013), pp.1-7.
[31] Dave K, Lawrence S, Pennock DM, *Mining the peanut gallery: Opinion extraction and semantic classification of product reviews.*, Proceedings of the 12th international conference on World Wide Web. ACM. 2003; 519-528.
[32] Kumar, A. and Jaiswal, A, *Systematic literature review of sentiment analysis on Twitter using soft computing techniques.*, Concurrency and Computation: Practice and Experience, e5107.
[33] Gamal, D., Alfonse, M., M El-Horbaty, E. S. and M Salem, A. B., *Analysis of Machine Learning Algorithms for Opinion Mining in Different Domains.*, Machine Learning and Knowledge Extraction, 1(1) (2019), 224-234.
[34] Kumar, A. and Sharma, A., *Socio-Sentic framework for sustainable agricultural governance.*, ustainable Computing: Informatics and Systems.(2018)
[35] Kumar, A. and Sharma, A., *Opinion Mining of Saubhagya Yojna for Digital India.*, In International Conference on Innovative Computing and Communications (2019) (pp. 375-386). Springer, Singapore.
[36] Asghar, M. Z., Khan, A., Ahmad, S. and Kundi, F. M., *A review of feature extraction in sentiment analysis..*, Journal of Basic and Applied Scientific Research, 4(3) (2014), 181-186.
[37] Chandrashekar, G. and Sahin, F., *A survey on feature selection methods.*, Computers & Electrical Engineering, 40(1) (2014), 16-28.
[38] Sorzano, C. O. S., Vargas, J. and Montano, A. P., *A survey of dimensionality reduction techniques.*, arXiv preprint arXiv:1403.2877. (2014)
[39] Pealver-Martnez, I., Valencia-Garca, R. and Garca-Snchez, F., *Ontology-guided approach to feature-based opinion mining.*, In International Conference on Application of Natural Language to Information Systems (2011, June) (pp. 193-200). Springer, Berlin, Heidelberg.
[40] National e-Governance Plan, `https://meity.gov.in/divisions/national-e-governance-plan`
[41] Indian Government Cloud Initiative (GI Cloud - MeghRaj), `https://www.slideshare.net/nasscom/meghraj-nasscom`
[42] GI Cloud (MeghRaj), `https://meity.gov.in/content/gi-cloud-meghraj`
[43] GI Cloud Initiative - Meghraj, `http://vikaspedia.in/e-governance/national-e-governance-plan/gi-cloud-initiative-meghraj`
[44] Government of India's GI Cloud (Meghraj) strategic Direction Paper, `https://meity.gov.in/writereaddata/files/GI-Cloud\%20Strategic\%20Direction\%20Report\%281\%29_0.pdf`
[45] A. Kumar and T.M. Sebastian, *Sentiment analysis on twitter.*, IJCSI International Journal of Computer Science, vol. 9, no. 4, pp. 372-378, Jul. 2012.
[46] Natural Language Toolkit, `https://www.nltk.org/`
[47] Natural Language Toolkit, `https://en.wikipedia.org/wiki/Natural_Language_Toolkit`
[48] Porter Stemmer, `http://people.scs.carleton.ca/~armyunis/projects/KAPI/porter.pdf`
[49] Porter's Algorithm., `http://people.ischool.berkeley.edu/~hearst/irbook/porter.html`
[50] Aizawa, A., *An information-theoretic perspective of tfidf measures.*, Information Processing & Management, 39(1) (2003), 45-65.
[51] Jing, L. P., Huang, H. K. and Shi, H. B., *A Improved feature selection approach TFIDF in text mining.*, In Machine Learning and Cybernetics, 2002. Proceedings. 2002 International Conference on (Vol. 2, pp. 944-946). IEEE.
[52] Cristani, M. and Cuel, R., *A survey on ontology creation methodologies.*, International Journal on Semantic Web and Information Systems (IJSWIS), 1(2) (2005), 49-69.
[53] Kumar, A. and Jaiswal, A., *Empirical Study of twitter and tumblr for sentiment analysis using soft computing techniques.*, In Proceedings of the World Congress on Engineering and Computer Science (2017) (Vol. 1).

Edited by: Anand Nayyar
Received: Feb 18, 2019
Accepted: Mar 20, 2019