SWARM-BASED OPINION MINING MODEL FOR DIGITAL GOVERNANCE

A DISSERTATION

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE
OF

MASTER OF TECHNOLOGY
IN
SOFTWARE ENGINEERING

Submitted by:

HIMANSHU SHEKHAR

2K18/SWE/07

Under the supervision of

DR. ABHILASHA SHARMA

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

JUNE, 2020

M.Tech (Software Engineering) HIMANSHU SHEKHAR 2020

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi-110042

<u>CANDIDATE'S DECLARATION</u>

I, Himanshu Shekhar, Roll No. 2K18/SWE/07 student of M.Tech (Software Engineering), hereby declare that the Project Dissertation titled **"Swarm-based Opinion Mining Model for Digital Governance"** which is being submitted by me to the Department of Computer Science and Engineering, Delhi Technological University, Delhi in partial fulfilment of requirements for the award of degree of Master of Technology is original and not copied from any source without proper citation. This work has not previously formed the basis for the award of any Degree, Diploma, Associateship, Fellowship or other similar title or recognition.

Himanshu Shekhar

Place: Delhi **HIMANSHU SHEKHAR**

Date: 30-06-2020

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi-110042

CERTIFICATE

I hereby certify that the Project Dissertation titled **“Swarm-based Opinion Mining Model for Digital Governance”** which is submitted by **Himanshu Shekhar**, Roll Number 2K18/SWE/07, Department of Computer Science and Engineering, Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of degree Master of Technology is a record of the project work carried out by the student under my supervision. To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Abhilasha Sharma

Place: Delhi

Date: 30-06-2020

Dr. ABHILASHA SHARMA

SUPERVISOR

Assistant Professor

Department of Computer Science & Engineering

Delhi Technological University

ACKNOWLEDGEMENT

It gives me immense pleasure to express my deepest sense of gratitude and sincere thanks to my highly respected and esteemed guide Dr. Abhilasha Sharma, Assistant Professor, SWE, DTU, Delhi, for their valuable guidance, encouragement and help for completing this work. Their useful suggestions for this whole work and co-operative behaviour are sincerely acknowledged.

I also wish to express my indebtedness to my parents as well as my family members whose blessings and support always helped me to face the challenges ahead.

Himanshu Shekhar

HIMANSHU SHEKHAR

ABSTRACT

Government in a country is the foremost legislative body responsible for taking decisive steps, planning schemes and implementing them with zero margins of error. These schemes and policies directly or indirectly affect the population of the country and direct the rate of social and economic growth. Effective policy framing and implementations have been the primary aim of all governments. But for good governance with long term sustainability taking opinions of the general public becomes indispensable. Twitter is one such open platform for a new type of social interaction where people come forward and express their views not only on products, movies and celebrities but also those critical policies and schemes designed by the government with aim of the overall development. These opinions have a lot more weight and convey a major message to the policymakers if evaluated correctly. This paper elucidates one such framework which mines opinion of general users tweeting on twitter about government policies and classifies them into three different polarities i.e. positive, negative and neutral. Machine Learning and Deep Learning method along with Natural Language Processing techniques has been utilized to extract the sentiments of the tweet and perform analysis on its polarity. The results of this detailed analysis can act as feedback to the governing bodies which can give them a better idea of the demography of the public's opinion in an effective manner. Thus, this research works presents a technology-based solution for smart governance and interactive policy framing.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

OM	Opinion Mining
NLP	Natural Language Processing
ML	Machine Learning
DL	Deep Learning
LR	Logistic Regression
SVM	Support Vector Machine
DT	Decision Tree
RF	Random Forest
NB	Naive Bayesian
XGB	XG Boost
KNN	K-Nearest Neighbours
MLP	Multilayer Perceptron
PSO	Particle Swarm Optimization
ACO	Ant Colony Optimization
BI	Business Intelligence
GI	Government Intelligence
ISA	Information Security & Analysis
MI	Market Intelligence
SCT	Sub Component Technology
SSS	Smart Society Services
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
LSTM	Long Short Term Memory
GRU	Gated Recurrent Unit
NMCG	National Mission for Clean Ganga
GPIs	Grossly Polluting Industries
GoI	Government of India
API	Application Programming Interface

CHAPTER 1

INTRODUCTION

This chapter briefly introduces the research work proposed in the thesis. Section 1.1 gives an overview of the research undertaken. Section 1.2 sets out the research objectives. Section 1.3 illustrates the proposed framework and the main contributions arising from the work undertaken. Finally, section 1.4 presents an outline of the thesis describing the organisation of the chapters.

1.1 Introduction

Today, a large amount of information is available on the internet where people constantly share and exchange their thoughts and ideas. Various social networking sites such as facebook, twitter etc has become a popular platform for expressing opinions and views about numerous things happening around the globe. These expressions and opinions carry a major part of the user emotion associated with the matter of concern. Tetsuya Nasukawa and Jeonghee Yi first established the term Sentiment Analysis and Opinion Mining in the year 2003 [1]. Sentiment Analysis is characterised as the evaluation of the poster's thoughts by determining the polarities of positive, negative and neutral for using a vast number of techniques which are an essential part of natural language processing. Textual data is often graded into two primary categories: opinions and facts. Facts can be described as an objective assertion relating to any subject, entity, events and their attributes. Opinions, on the contrary, are the subjective assertion that portrays an individual's attitude and sentiments towards people, groups, organizations, events and their attributes [2,3,4]. On the internet, people not only convey their views about services and products but also discuss diverse social and economic topics. These opinions can be used as feedback for government schemes and can help improve the framework of good governance. Endorsement of good governance becomes indispensable especially when the administrative bodies are becoming attentive towards planning and instrumenting its strategy for the overall benefit for the citizens. The United Nations Development Programme interpret the attributes of good governance as Responsiveness, Involvement, Transparency, Ordinance rule, Consensus oriented decision making, Neutrality, Efficiency, Effectiveness, Accountability, and Strategic insight [5]. Thus, to harness the ethical governance goal which is important for a country's development, this paper attempts to find a tech-based solution of providing a kind of feedback which effectively conveys the general sentiments of the nation about a scheme [6]. This is an attempt to answer the important question which is that, how a specific proposed policy impacts the life of a common individual or a social class or a nation as a whole? Thus, while framing and implementing a policy it is important to keep people's opinion in mind. Social networking has evolved as an e-participation and e-communication synchronic mediating technology that support the public-government association. Its increased accessibility and visibility, comprehensive outreach and economical exposure make it a substantial government assessment tool. Twitter emerged as one of the most popular microblogging platforms [7,8]. It has always been the home to people's opinion where people tweet about a wide variety of topics. These tweets contain general sentiments of the user which can be mined to find out the polarity of the user regarding the topic of concern [7,9]. Reportedly, about 275 million people use twitter actively within a month [8]. Due to a large amount of data generated on twitter, using twitter datasets for mining public sentiments is a research practice used a lot in these digital times. There are myriads of astounding evidence on social media which proves the usability of twitter data for improvement in the functioning of civic bodies. Encouraged by this, we develop a socio-affective framework for smart policy framing process using supervised learning and deep learning based on natural language processing and opinion mining on the social web [10,4]. Tweets corresponding to government proposed schemes and policies have been extracted to implement and corroborate the framework which explores to ascertain public opinion concerning the policies.

The outcome enumerates the performance of the classifier for gauging public opinion on government policies. As a new measure with an eye to promote good governance for economic advancement, the proposed framework illustrates the implementation of machine learning model as a part of government policy evaluation phase using the concept of opinion mining on the social web. The phrase 'Socio-Affective' or 'Socio-Sentic' used here to represents the two major components of the model. The word 'Socio' relates to the 'sociological' or 'society' part of the model which implies social surroundings or public network. The word ‘Affective or Sentic’ here pertains to implicit features and emotions connected with vernaculars, languages, and culture exploited for application such as sentiment identification out of speech and text [11,10,12]. Hence, by using the two terms together, we tend to convey that the proposed framework uses social media and opinions of people socially expressed to find out the sentiments behind their views and use them as a feedback to the governing bodies which can later use them to make amends to the framed policies.

1.2 Research Objectives

Statement of Research Question

"Can the extensive and massive user-generated text on social media platform be mined to yield perception into public opinion for understanding the orientation change from traditional governance to digital governance?"

In response to the identified need to better exploit the knowledge capital in the form of opinions accumulated on social web, this unifying research question can be broken down into the following three questions, each of which will be addressed by this research:

- How opinion mining can exploit web 2.0?
- How can the opinion polarity of user generated big data be determined?
- How opinion mining framework can assist and foster digital governance?

Consequently, the three main research objectives of the work undertaken are:

1. **Research Objective I** – To seek the convergence of Web 2.0 technologies and opinion mining.
2. **Research Objectives II** – To propose a novel framework for determining opinion polarity of user generated big data.
3. **Research Objectives III** – To find out use case of opinion mining in digital governance by forming a political and social decision support framework for governance.

1.3 Proposed Framework

The proposed framework propounds a technological solution which adds two s-elements to governance giving it a social and sentiment based dimension. Thus, the aim of this work is bifold. Firstly, we strive to outline a unified conceptual model illustrating the relationship of governance with different aspects of government policies and secondly, we attempt to establish the socio-sentic facet to the unified model by introducing an optimal predictive analytics framework for assessing public sentiment over a social media platform.

1.4 Organization of Thesis

The paper is organized as follows:

Chapter 1 describes the overview of the research work undertaken. It also addresses the research questions arisen from the research objective. Further, it establishes a brief overview of the proposed framework as a technological solution.

Chapter 2 describes the background concept related to the research work domain. It also highlights the literature survey of the related works in the area of opinion mining. summarizes a government schemes as a case study.

Chapter 3 presents a general architecture of opinion mining model. Two case study is also described which is used as a way to validate the proposed framework. It further illustrate a conceptual framework for each case study equipped with opinion mining which facilitate an insight on how government programmes and initiatives is apprehended by its relevant stakeholder groups.

Chapter 4 elucidates a machine learning based evaluation model for each case study. It highlights the observations and findings.

Chapter 5 analyses the findings and observations based on statistical measures.

Chapter 6 present the concluding remarks and the future scope.

CHAPTER 2

SYSTEMATIC LITERATURE REVIEW

This chapter discusses the background concept in the domain of digital governance and opinion mining. Further, it discusses the related works in the domain of opinion mining.

2.1 Background Concept

Digital Governance

Digital-governance is a framework for establishing accountability, roles, and decision making authority for an organization's digital presence. The intent of digital governance is to ensure that common citizens should be a part of decision-making processes as it affects them directly or indirectly, which in turn improves their conditions and the quality of lives. This new facet of governance will assure that citizens are active contributors in deciding the kind of services they want.

Table 2.1: People participation in Digital vs. Conventional Governance Models

People Participation	Conventional Governance Models	Digital Governance Models
Mode of Participation	REPRESENTATIVE	INDIVIDUAL/COLLECTIVE
Domain of Participation	IN-SITU	EX-SITU
Approach to Participation	PASSIVE/REACTIVE	PRO-ACTIVE/INTERACTIVE
Impact of Participation	INDIRECT/DELAYED	DIRECT/IMMEDIATE

How the perspective and participation of people changes with the transformation of governance model from conventional to digital is listed in table 2.1 [45].

From the matrix above, it is evident that the use of digital governance leads to closer contact of individuals with decision-maker in the government & the impact is immediate. On the whole, it puts greater access and control over governance mechanism in the hands of individuals, and in process leads to a more transparent and efficient governance. This shift from passive to active to the current need of interactive governance can thus be conceptualized, giving an insight to the social model of governance.

The emergence of social web and the consequential abundant data can be mobilized to define a S-Governance model (where S – stands for Social), a model of government-citizen engagement that complements the web-based e-government services. As a step towards intelligent governance, we expound a new perspective of "Sentiment" in S-governance.

S-governance comprises of two S-factors in the domain of governance: Social & Sentiment (Sentic). The social factor refers to societal interaction of person/entity for their collective co-existence and sentiment describes an expression of strong influence of people/society. The goal of social and sentiment intelligence based governance is to look forward towards concerned audience and give credence to their views/thoughts for the purpose of information broadcasting, looking for civic inputs in policy making, employment, granting access to services, to uplift and foster stakeholders, etc. The electronic journey of digital governance as the web evolved is represented in fig. 2.1.

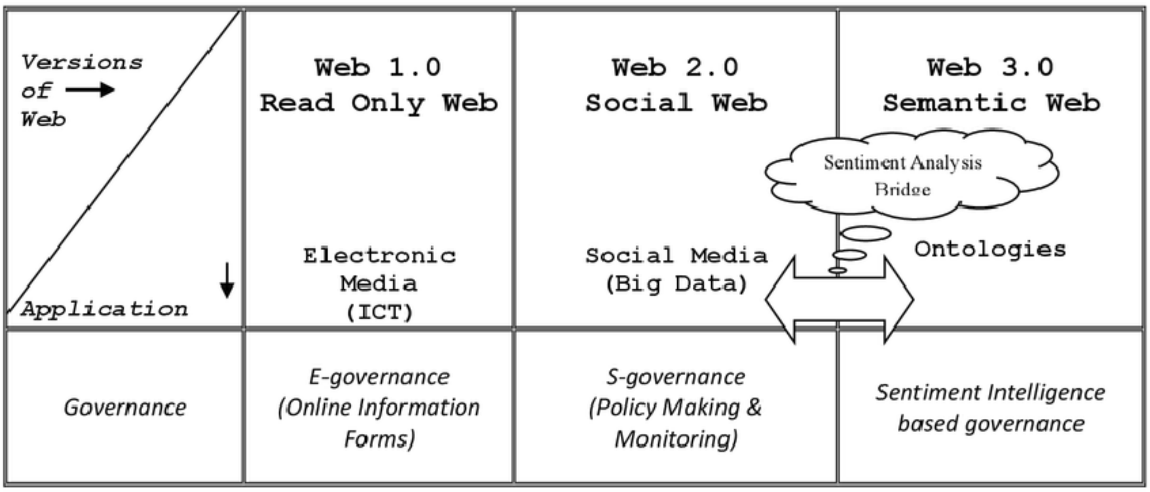

Fig. 2.1. Evolution of Digital Web Governance

Opinion Mining

Opinion mining is a field of study that tends to use the natural language processing techniques to extract, capture or identify the viewpoint of a person with respect to a particular subjects. It is the automated mining of attitudes, opinions, and emotions from text, speech, and database sources through NLP. The primary task is to opinionate, i.e. to sort and categorize one's perspective into positive, negative or neutral views. Once it is done, it can be further sub categorized into two parts, one of them focusing on the information that is factual, more likely to be an objective description of a unit, while the other emphasizes on sentiments, that are subjective in the expression of the opinion holder. Both of them hold equal importance in deducing conclusions.

Approaches of Opinion Mining

Analysing the content of social media for opinion mining is a tedious task as it requires a thorough and extensive knowledge of the rules associated with the NLP. Three main techniques used for opinion polarity classification are as follows:

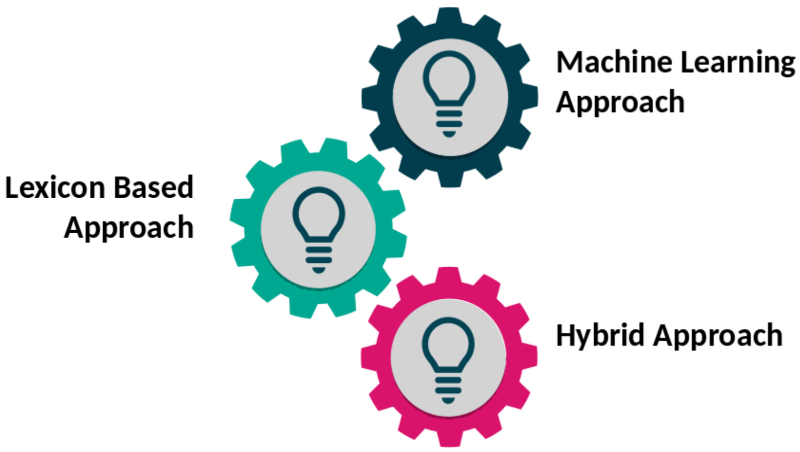

Fig. 2.2. Approaches of Opinion Mining

- **Machine Learning Approach** – The machine learning approaches can be grouped into: supervised and unsupervised learning methods. In supervised method, learning done from training data is applied to a new test data, whereas, in unsupervised method there is no prior learning (i.e. no training data), and the task is to find hidden structure in unlabelled data.

- **Lexicon Based Approach** – The lexicon-based approaches tend to be dependent on the sentiment vocabulary, that provides a collection of known and precompiled sentiment terms.
- **Hybrid Approach** – The hybrid approach is the combination of both the above mentioned methods and plays an important role in decision making as the techniques of both the approaches are collaborated for a better result.

Table 2.2: Coherence of three dimensions with their specialization

<table>
<tr><th colspan="2">OM Techniques</th><th>OM Tasks</th><th>OM Application</th></tr>
<tr><td>ML based</td><td>Hybrid Technique</td><td rowspan="2">Feature Extraction & Selection</td><td rowspan="2">Government Policy Evaluation</td></tr>
<tr><td>Supervised Machine Learning</td><td>ML + Swarm-based</td></tr>
</table>

Table 2.2 represents the relational matrix of opinion mining three dimensions along with their specialization that has been used in this research work. The selected application area is government where opinion mining has been incorporated in policy evaluation phase of policy life cycle and the proposed opinion mining models have been validated using machine learning techniques (ML + Swarm-based) by performing tasks such as feature extraction and feature selection. Fig. 2.3 [6,46] represents the three dimensions of opinion mining.

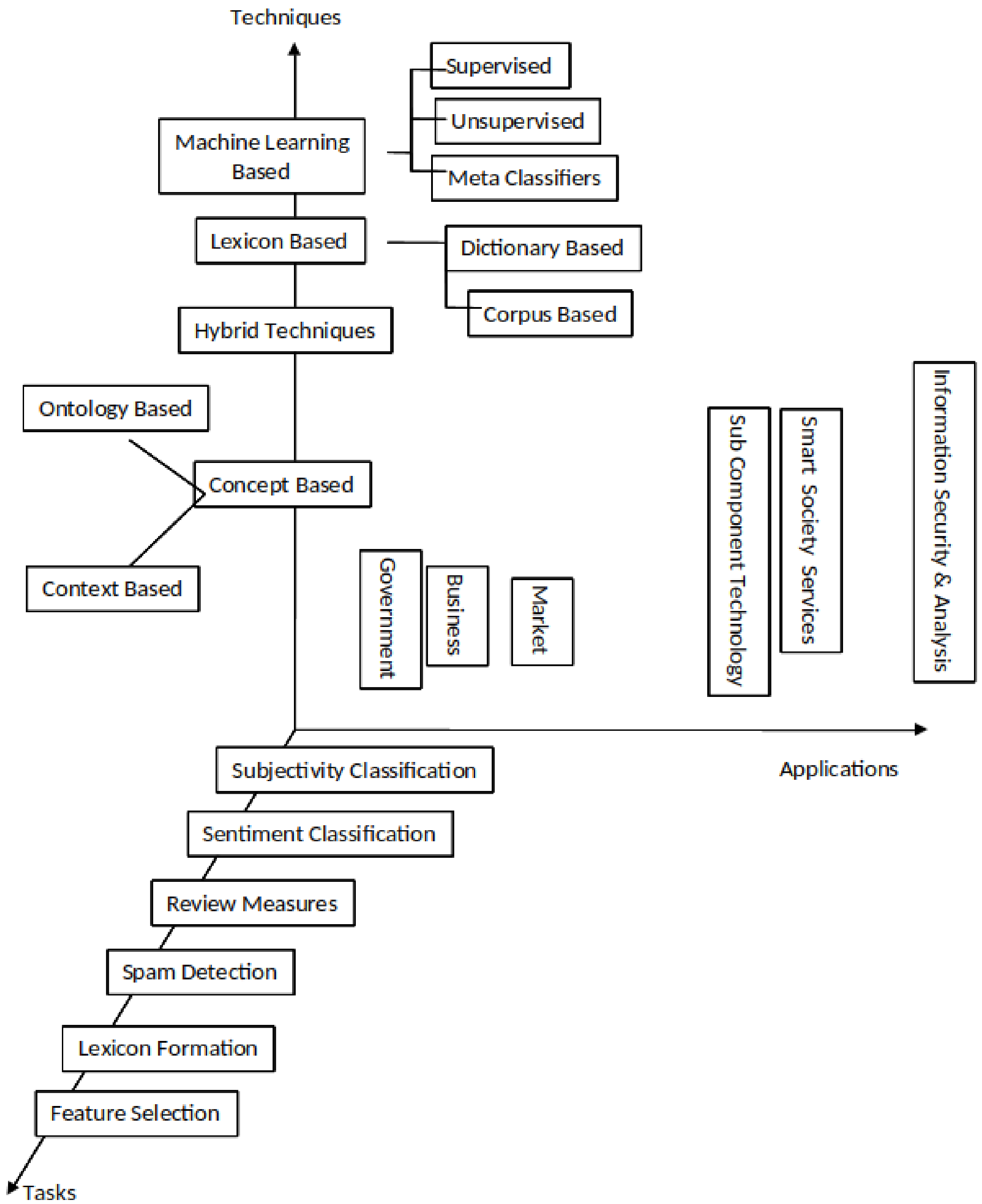

Fig. 2.3. Three dimensional model of opinion mining

Swarm Optimization

This work implements swarm optimized approach that uses the combination of nature inspired algorithms along with machine learning techniques. Swarm Intelligence (SI) [13,14,15] introduced in 1989 by Gerado Beni and Jing Wang. "In particular, the discipline focuses on the collective behaviours that result from the local interactions of the individuals with each other and with their environment" [16]. It is a contemporary computational and behavioural metaphor for solving optimization problems that take collective biological patterns provided by social insects (ants, termites, bees, wasps, moths etc.) and other animal societies (fish, birds, grey wolfs etc.) as stimulus to model solutions. Nature inspired computing is seeking attention due to (a) capability to cater complex problems. (b) the behaviour of naturally occurring phenomena which is used to solve complex problems. A standard and generic swarm intelligence system consist of many individuals that are homogeneous in nature and the interaction between them is based on simple behavioural rules. Individuals exchange the information either directly or through their environment and this communication results in the overall behaviour of the system. Altogether, a swarm represent an intelligent behaviour. Moreover, the integration of artificial intelligent techniques with swarm in the field of opinion mining is creating a next generation prediction analysis model which provide researchers a new and innovative way to propose biological/swarm based algorithms. This research endorse to solve real world problems with the use of swarm intelligence techniques.

2.2 Related Works

A literature survey is carried to review the state-of-the-art research in the area of opinion mining. The studies have been evaluated based on techniques and application areas of opinion mining. As per literature so far, application areas of opinion mining can be broadly classified into six categories namely, Business Intelligence (BI), Government Intelligence (GI), Information Security & Analysis (ISA), Market Intelligence (MI), Sub Component Technology (SCT) and Smart Society Services (SSS) whereas four classes of techniques have been used so far which include Machine Learning, Lexicon.

The following table 2.3 represents the summary of opinion mining techniques and the respective application areas [3] in which these have been used as observed from the selected final studies.

Table 2.3: Usage of OM techniques in its application areas

Application Area → / Techniques ↓	BI	GI	ISA	MI	SCT	SSS
Machine Learning	√	√	√	√	√	√
Lexicon Based		√	√	√	√	√
Hybrid		√		√	√	√
Concept Level		√		√		√

Fig. 2.4 and 2.5 illustrate the percentage of work done in various application areas of opinion mining and the percentage usage of its techniques/approaches in these application areas.

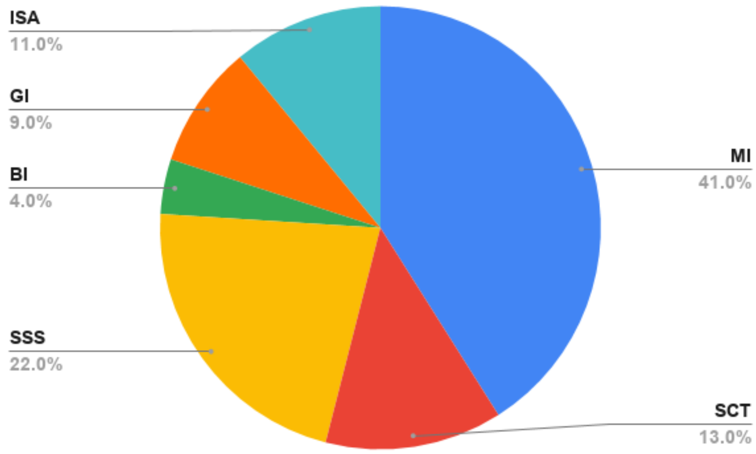

Fig. 2.4. Work done in OM applications (%)

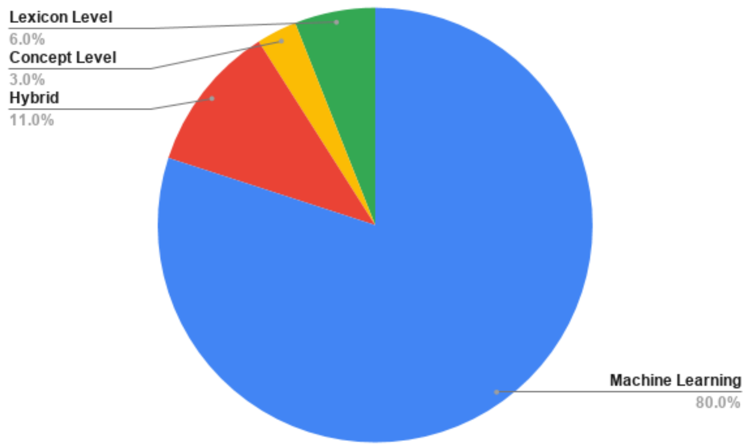

Fig. 2.5. Work done in OM techniques (%)

The following charts in fig. 2.6 and 2.7 represents the percentage of work done in various field actions of government intelligence and the percentage use of opinion mining techniques in these action areas.

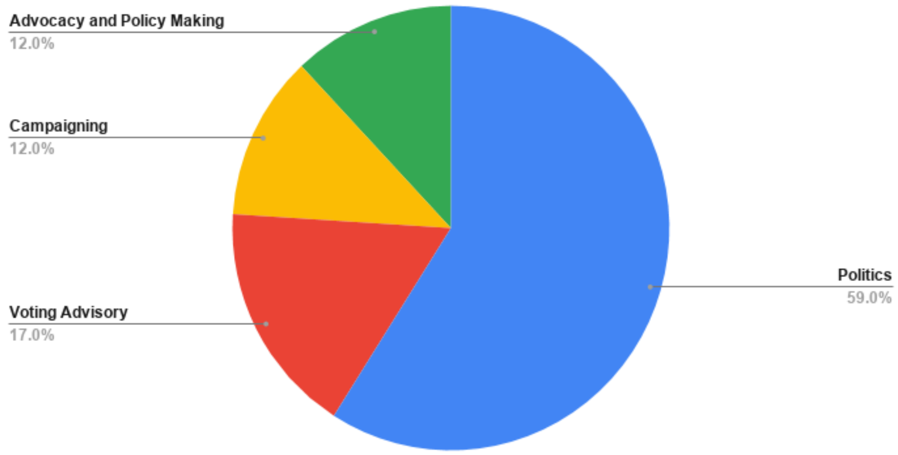

Fig. 2.6. Work done in GI action areas (%)

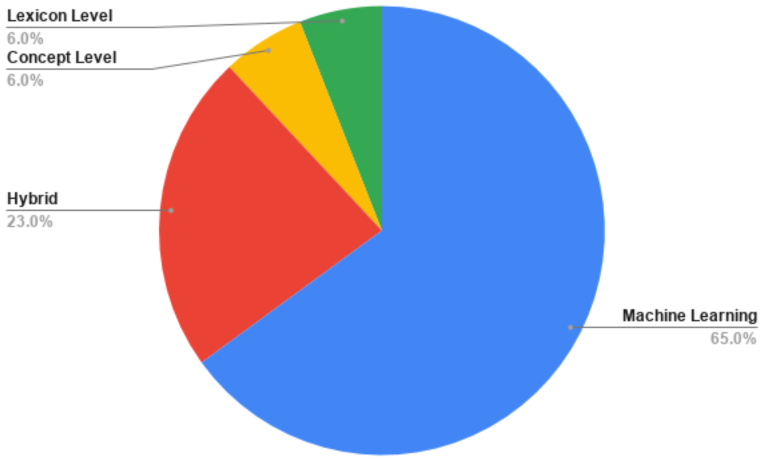

Fig. 2.7. Work done in GI with OM techniques (%)

CHAPTER 3

OPINION MINING MODEL FOR GOVERNMENTAL POLICY EVALUATION

This chapter elaborates a general architecture of opinion mining model and the steps involved in it. It also consider two case study which acts as a validation tools for the proposed framework. It further illustrate a conceptual framework for each case study equipped with opinion mining which facilitate an insight on how government programmes and initiatives is apprehended by its relevant stakeholder groups.

3.1 Opinion Mining Model

The general architecture of opinion mining model consist of various steps which helps to classify the opinion polarity. The steps involved are discussed as follows:

- Data Gathering and Pre-processing
- Feature Extraction Approach
- Optimized Feature Selection Approach
- Classification Model
- Opinion Polarity Prediction

3.1.1 Data Gathering & Pre-processing

Data gathering is the primary and essential measure in the context of opinion prediction. Numerous social media platform is accessible which convey the public notion concerning several domains. Twitter is a platform [17] where public express their opinion, belief, recommendations and sentiments by tweeting tweets over any subjects or events. Twitter provides standard search APIs (Application Program Interfaces) which returns a collection of relevant tweets matching a specified query by passing #(hashtag) with the topic name as parameter. The python script employing query searching option for the hashtag #NamamiGange #Budget2019 has been executed for tweets collection. The search query fetches tweets on the #topic and is stored as a .csv file for further processing. The collected data is pre-processed which results in clean and transformed data for next step i.e., feature extraction. The steps involved in pre-processing are as follows:

- Remove twitter handles (@user)
- Remove unwanted text patterns from the tweets
- Remove punctuations, numbers and special characters
- Remove short words
- Remove non-ASCII characters to restrict the range of data for English language
- Tokenization of strings using NLTK [18,19]
- Text Normalization using NLTK's Porter Stemmer [20,21]

3.1.2 Feature Extraction Approach

"Term Frequency – Inverse Document Frequency (TF-IDF) [22] is a standard statistical weighing factor which measures how important a word is to a document by considering the frequency of any word occurring in a corpus". TF indicates the raw count of a term in a document. Raw count implies the number of times any term t occurs in document d. IDF is the inverse frequency of the terms in the text which simply display the importance of each terms.

$$tf(t,d) = \left(\frac{No.of\ times\ term\ t\ appears\ in\ a\ document\ d}{Total\ no.of\ terms\ in\ the\ document}\right) \tag{3.1}$$

$$idf(t,D) = \log_e\left(\frac{Total\ no.of\ documents}{No.of\ documents\ with\ term\ t\ in\ it}\right) \tag{3.2}$$

Thence, TF-IDF is computed as:

$$tf - idf(t,d,D) = tf(t,d) * idf(t,D) \tag{3.3}$$

where t indicates the terms; d indicate specific document and D indicate cluster of documents.

3.1.3 Optimized Feature Selection Approach

Particle Swarm Optimization

Particle Swarm Optimization (PSO) [23] is a nature-inspired evolutionary and stochastic optimization technique to solve computationally hard optimization problems. It is a robust metaheuristic approach predicated on the movement and intelligence of swarms. It mimics the navigation and foraging of a flock of bird or school of fishes. In PSO, a swarm of n particles interact either directly or indirectly with one another using search directions. The preferred algorithm utilizes a set of particles flying over a search space to identify a global optimum. During an iteration, each particle updates its position in accordance with its experience and experience of its neighbours [24]. The pseudo code of the PSO algorithm is described in fig. 3.1.

```
For each particle
    Initialize particle
end for
do
    for each particle
        calculate fitness value
        if the fitness value is better than the pBest in history
            Set current value as the new pBest
    end for
Choose the particle with the best fitness value of all particles as the gBest
for each particle
    Calculate particle velocity
    Update particle position
end for
while max iteration or min error criteria is not attained
return gBest as the best estimation of the global optimum
```

Fig. 3.1. Pseudo code of PSO algorithm

In the PSO algorithm every solution of a given problem is considered as a particle, which is able to advance in a search landscape. As a means to update the position of each particle two vectors are taken into account, velocity vector and position vector. The velocity vector reflects the orientation of movement and the position vector reflects the positioning of the particle in the landscape. These two vectors are updated in each iteration using equations (3.4) and (3.5) [23, 25].

$$Vi(t+1) = w.Vi(t) + c1.rand1(t).[Pi(t) - Xi(t)] + c2.rand2(t).[Pg(t) - Xi(t)] \quad (3.4)$$

$$Xi(t+1) = Xi(t) + Vi(t+1) \quad (3.5)$$

where, t stand for iteration number; w is inertia weight; c1 and c2 are positive constant, called learning factor; rand1 and rand2 indicate a random function whose value lies in the range [0, 1]; Pi = (p_{i1}, p_{i2}, ...p_{iD}) indicate the best previous position of swarm and P_g indicate global best among all the particles in the population; Xi = (x_{i1}, x_{i2}, ..., x_{iD}) display the position of the i^{th} particle in a search space with D dimensions; Vi = (v_{i1}, v_{i2}, ..., v_{iD}) indicate the variation rate of position. Each components of the velocity vector are defined as a separate term. The first component is termed as inertia since it retains the current velocity. The second component is termed as cognitive or individual component. This is because each particle considered the distance between its personal best and the current location. The last component however is termed as social component because the particle calculates the distance between its current and the best position found by the entire swarm. The effect of cognitive and social components on the motion of particles can be altered by tuning the coefficient c1 and c2. The values of these parameters are randomly generated from a number drawn from a uniform distribution and is finally chosen as 2.1 and 1.9 respectively. The initial weight tunes exploration and exploitation and is standardly reduced linearly from 0.9 to 0.2.The proposed feature selection technique is based on the binary version of PSO algorithm [26]. Each particle's position is acknowledged as binary bit strings where every bit constitutes a feature. The bit value 1 designates that the feature is selected whereas, the bit value 0 designates that the feature is not selected. The process starts by generating a number of particles which are then placed randomly on the search space. The choice of number of particles to place is equal to the number of features within the data. Each particle proceed towards the solution by changing its initial position and velocity in each iteration. Each particle calculates the fitness value and update the personal best and global best if required. In each iteration the selected features are evaluated and is chosen as the best subset features. The algorithm stops once it reaches its max iteration or minimum error criteria is not attained. The best feature subset is returned as the best estimation of the global optimum.

Ant Colony Optimization

This sub-task introduces swarm-based feature selection approach. The optimal feature subset is identified with Ant Colony Optimization (ACO) algorithm. The block diagram of ACO algorithm is illustrated in fig. 3.2.

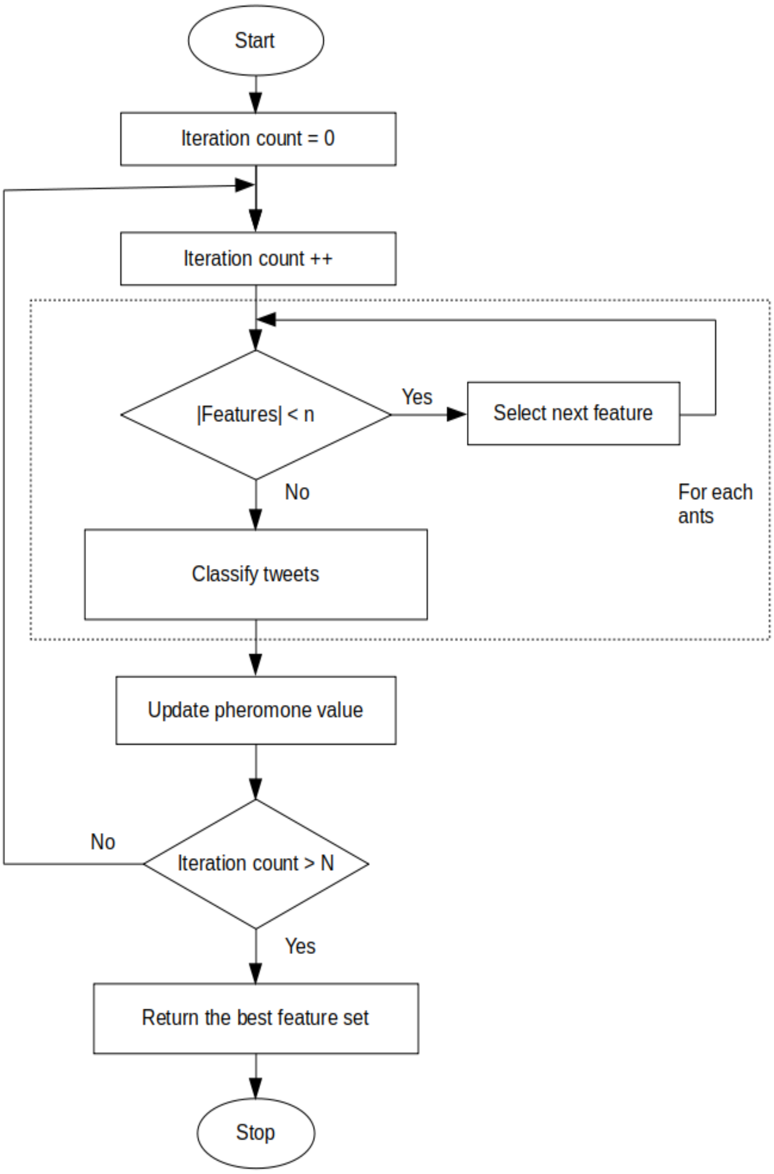

Fig. 3.2. Block diagram of ACO algorithm

In our suggested approach, each feature constitutes a node, and entire nodes are independent among themselves. Nodes are chosen in accordance with their selection probability $P_{k(r,s)}$ which is expressed in equation (3.6). Primarily, all nodes possess uniform selection probability:

$$P_{k(r,s)} = \frac{[\tau(r,s)]^{\alpha}[\eta(r,s)]^{\beta}}{\sum_{u\epsilon N_k}[\tau(r,u)]^{\alpha}[\eta(r,u)]^{\beta}} \tag{3.6}$$

where τ is the pheromone level or pheromone trail value and $\eta_{r,s}$ indicates the quality of the edge r-s on the graph. α indicates the degree significance of pheromone, β indicate degree visibility and $u\epsilon N_k$ is a decision that pertains ant k (neighbourhood) when it was at node r. With the parameter α and β we can increase or decrease the impact of τ or η in the process of making decisions. The denominator of this equation as the pheromone and quality of all edges that can be considered from the node r. This probability is calculated for all the edges connected to the current node and is a number in the interval of 0 and 1. Since, we are interested in the shortest path, $\eta_{r,s} = 1/L_{r,s}$. That means the length of an edge or the cost of an edge indicate how good it is in the process of calculating the probability of choosing that edge. The pheromone trail value is initialized to 10, and parameters α and β are initialized to 1 as a suitable initial value according to [27]. Once all the ants have established the entire circuit, the pheromone trail is updated in accordance with the equation (3.7) as

$$\tau_{r,s} = (1-\rho)\tau_{r,s} + \sum_{k=1}^{n} \Delta\tau_{r,s}^{k} \tag{3.7}$$

where, $\Delta\tau^{k}_{r,s}$ shows the amount of pheromone deposited by k^{th} ant on the edge connecting node r to the node s is given by (3.8) as

$$\Delta\tau_{r,s}^{k} = \begin{cases} \frac{1}{L_k} & kth\ ant\ travels\ on\ the\ edge\ r,s \\ 0 & otherwise \end{cases} \tag{3.8}$$

L_k is the length of the path found by the kth ant and since we are trying to find the shortest path we divide it by 1. So, the shorter the path the higher pheromone should be deposited by the ant. Now to calculate the amount of pheromones on each edge we need to add a summation, where m is equal to the total number of ants. Since we want to simulate the evaporation we have to add $(1-\rho)\tau_{r,s}$ to the equation. ρ is a constant that allows us to define the evaporation rate. When ρ is equal to 0 there is no evaporation and when ρ is equal to 1 then the evaporation is at the maximum level. According to [27], ρ value is selected as 0.2. In (3.8), L_k implies F-measure value of ant k's feature subset. This signifies that greater the F-measure value of the ant's selected subset, the more the pheromone deposited on the features used in the subset, and these features have greater likelihood of selection in the next loop. The classification efficiency is evaluated in terms of F-measure [28] value which is described in (3.9) as

$$F - measure = \frac{2 * recall * precision}{recall + precision} \tag{3.9}$$

Eventually, the feature subset holding the highest F-measure value is elected as the best feature set.

3.1.4 Classification Models

Machine Learning based Model

Table 3.1: Elaboration of machine learning algorithm

Machine Learning Algorithm(s)	**Description**
Logistic Regression	Uses logistic function to model a binary dependent variable.
Support Vector Machine	belongs to the class of discriminative supervised learning based classifier for identification of classification pattern. It classify the data by a hyperplane.
Random Forest	ensemble of decision trees; combine together to get more specific results.
XG Boost	provides gradient boosting framework.

Multilayer Perceptron	represents a network of neurons named perceptron belongs to the class of feed forward artificial neural network.
k-Nearest Neighbors	fundamental machine learning algorithm that does not make any underlying assumptions about data distribution. Here, classification of objects is based on the voting of neighbours and the class assigned to the object is usually among its k nearest neighbours
Decision Tree	tree model of decisions and their results used as a decision support tool. developed from top to bottom with a single root node at the top and branching of several leaf nodes with probable outcomes.
Naive Bayesian	part of a class of simple probabilistic classifiers used for the estimation of classification parameters.

The optimal features subset obtained in the last stage is exploited to train and test the classifier into three predefined class of opinion polarity, precisely positive, negative and neutral for the process of prediction analysis. In this work, different machine learning algorithms have been used namely, Support Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF), XG Boost (XGB) and Multilayer Perceptron (MLP), k-Nearest Neighbors (KNN), Decision Tree (DT), Naive Bayesian (NB) to build statistics-based opinion prediction model. An elaboration of these algorithms is elaborated in table 3.1[29,30]. An extremely important parameter, accuracy is used as standard evaluation metrics.

Deep Learning based Model

The extracted feature is also implemented over custom deep learning-based model. The sentiment classification model consists of the following components connected in sequence (fig. 3.3). Word Embedding is a representation of a word as a sequence of tokens which may or may not be syntactic where similar meaning words possess similar representation. In the deep learning frameworks, the concept of word embedding is managed by embedding layer which maintain a lookup table to map the words to their dense vector representations. The sequence of embedding vectors is passed as input to deep network and is converted to a compressed representation. The deep network part used here is CNN i.e. Convolutional Neural Network and RNN i.e. Recurrent Neural Network and some forms of it like LSTM/GRU along with dropout to handle overfittings. Multilayer perceptron along with batch normalization further takes the compressed representation as input and convert it into the final output class. As a final step softmax activation layer is added for multi-classification output.

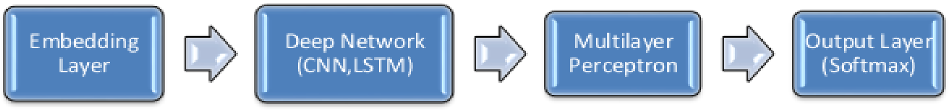

Fig. 3.3. Deep learning classification architecture

This deep learning model involves LSTM network layer builds upon a series of convolutional and max pool layer which in turn is built upon the embedding layer. The convolutional layer captures the features and contextual meaning of the words used in the tweets. The dataset extracted after pre-processing the tweets were divided in the ratio of 4:1 where 80% of the dataset was used to train the model and the remaining 20% was used to cross-validate. The model was trained on Google Cloud Platform for efficient and faster learning. The metric used for training here was accuracy. The attempt was to accurately classify the tweets while minimizing the cross-entropy loss function. The model has been trained for 10 epochs with batch size on each epoch as 128. The 'adam' optimizer was used to optimize the stochastic gradient descent process.

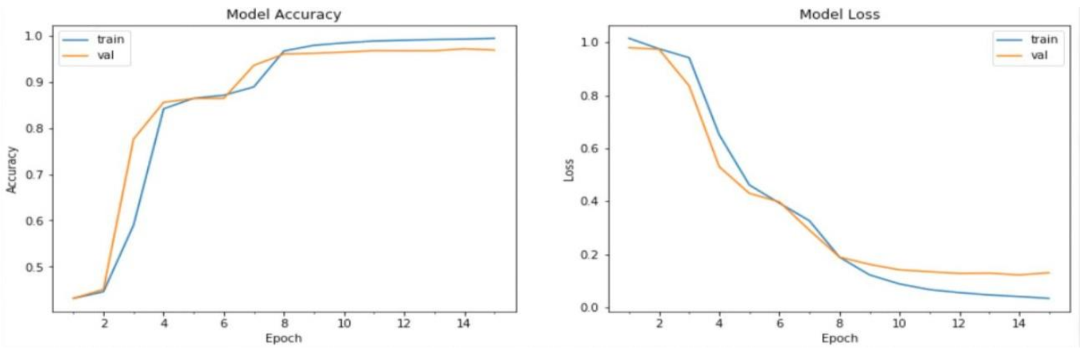

Fig. 3.4. (a) Accuracy vs. epoch plot (b) loss vs. epoch plot

On the training set the model was able to score an accuracy of 99.4% whereas the accuracy on the validation set was observed as 96.9%. The plot of model accuracy and loss function with respect to number of epochs is shown in fig. 3.4.

3.1.5 Model Evaluation

As a benchmark of policy assessment machine learning techniques has been implemented to perceive opinion polarity of tweets related to the chosen policy. Opinion mining being a significant part of policy evaluation process in a policy life cycle model can assist the governing institution to refine the action of decision-making process and resource escalation. The basic concept of opinion mining can be deployed to comprehend the public inclination towards any scheme or programme.

3.2 Case Study: An Overview

3.2.1 Namami Gange

Statistics reveal that nearly 30% of citizen in India dwell in cities that are likely to double in population by 2050 [31]. With economic expansion and change in style of living, the stress on already constrained water resources is escalating. The growing demand of water for domestic, industrial and agricultural use leads most of the river basins to water dearth. This phenomenon is further accentuated by the fact that water demand is unequally distributed throughout the country. Growing demand from a rising population combined with economic activity amplifies the tension over already stressed water resources. The escalation and heterogeneity of harmful human activities [32] such as industrialization, urbanization, deforestation, agronomics and other urban activities resulted in rapid deterioration of national river Ganga. The key factors underlying degradation includes exploitation of natural resources, dumping of noxious substances, fall in water retention capacities and restoration of waterways, mutilation of rivers by gradually engineering operations to topographical processes in the basin [33]. Consequently, the imperative demand is to reinforce the significance of governance in water management industry by formulating and designing policy frameworks in line with their relentless progression with an eye to make them a success. Considering India's vast challenges and needs, a new programme is required for efficient water management. "Namami Gange is an integrated conservation mission by the union government under National Mission for Clean Ganga (NMCG), initiated in

June 2014 with a financial investment of Rs. 20,000 crores to achieve the dual objectives of effectual reduction of pollution, conservation and rejuvenation of national river Ganga" [34]. The ultimate objective is to design a blueprint for restoring the wholesomeness of the Ganga river.

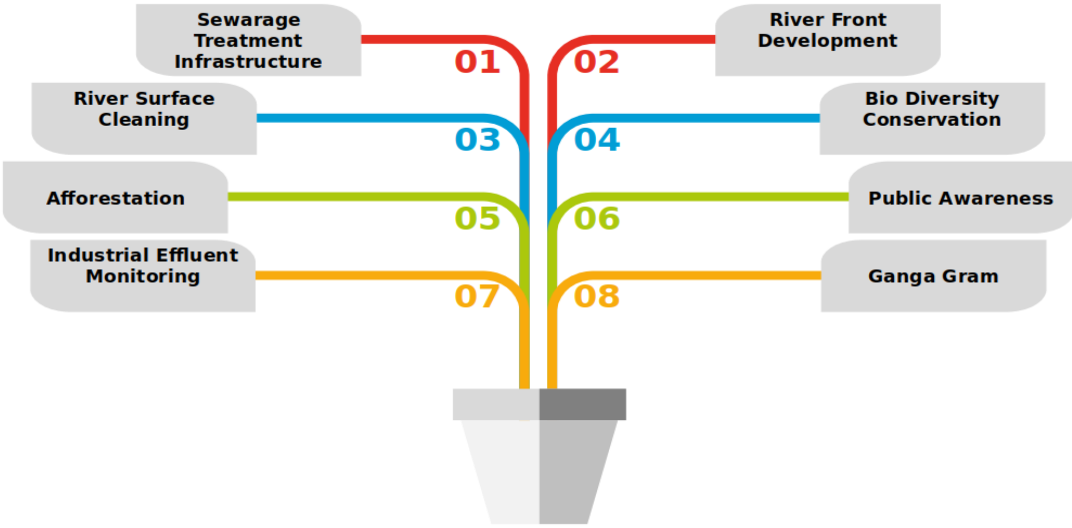

Fig. 3.5 Objectives of Namami Gange

The entire programme has been launched to address the eight major objectives [34] represented in fig. 3.5. and listed as follows:

- *Sewerage Treatment Infrastructure* aims to remove physical, biological and chemical contaminants from waste water and domestic sewage.
- *River-Front Development* project has been initiated for upgradation and restoration of river Ghats.
- *Watercourse Cleaning* for segregation of debris material from river surface. Its disposal is underway and there is an ongoing effort to push this service in many regions.
- *Bio Diversity Conservation* measures variation at various ecosystem to preserve the continuity of food chains. The conservation project includes

Ganga rejuvenation, fishery conservation, dolphin conservation and restoration of identified priority species.

- *Afforestation* aims to restore forests and also prevent flooding and soil erosion. Forest involvement for Ganga has been initiated through wildlife institute of India in order to achieve this objective.

- *Public Awareness* programme and activities such as workshops, seminars, and events are organized for wider public outreach. Various awareness activities such as campaigns, cleanliness drive, rallies are also organized at different levels to create mass awareness.

- *Industrial Effluent Monitoring* stations has been installed in various Grossly Polluting Industries (GPIs) and has been given deadlines to conform to prescribed norms.
- *Ganga Gram* a sanitation-based project for integrated development of model villages in five state for identified Gram Panchayats.

Several measures [35] have been undertaken to clean up the River Ganges project represented as timeline in fig. 3.6.

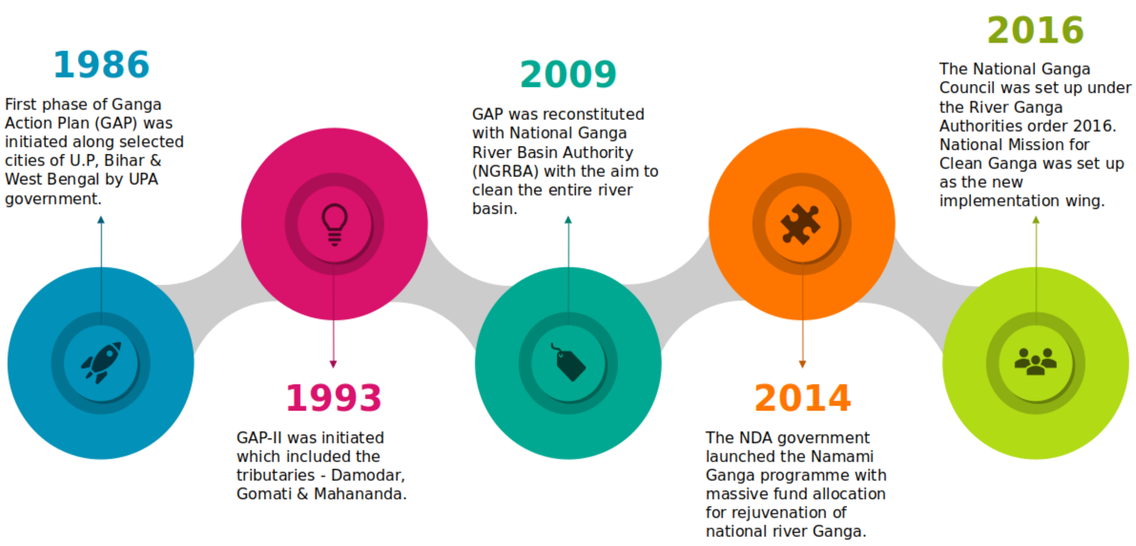

Fig. 3.6. Timeline status of Ganga cleaning plan

The progress status of the varying modules and subtask with reference to the data from NMCG's July 2018 project status report [36] describing the funds allocated and percentage of funds utilized years and activity wise is illustrated in fig. 3.7.

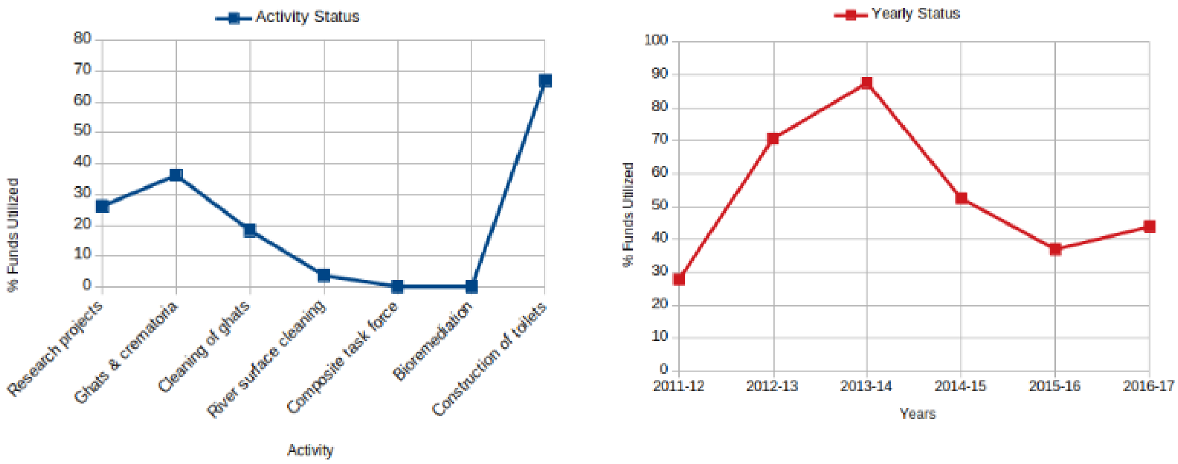

Fig. 3.7. (a) Activity wise funds utilization (b) Year wise funds utilization

The trend in year wise fund utilization shows 2013-14 was the year of maximum funds utilization while 2011-12 shows the least percentage of funds utilization followed by 2015-16. Considering the trend in activity wise funds utilization zero per cent of funds was utilized for the composite task force and bioremediation while maximum funds were utilized for construction of toilets.

3.2.2 Budget 2019

Acting Finance Minister Piyush Goyal introduced the Interim Union Budget of India for the year 2019 on 1 February 2019. The financial outlay was amended for six prime social schemes emphasizing on strengthening the well-being of farmers and the poor, further stating a new scheme aimed at direct cash transfers to farmers. Finance Minister Arjun Jaitley presented Interim Budget 2019 on 1 February 2019 with main features listed below [37]:

1. Tax proposals for Individuals
2. Tax proposals for Businesses
3. Measures for the poor and backward class

4. Women Empowerment
5. Banking Reforms & Insolvency and Bankruptcy Code (IBC)
6. Positive disruptions in Pension Sector
7. Agriculture Reforms

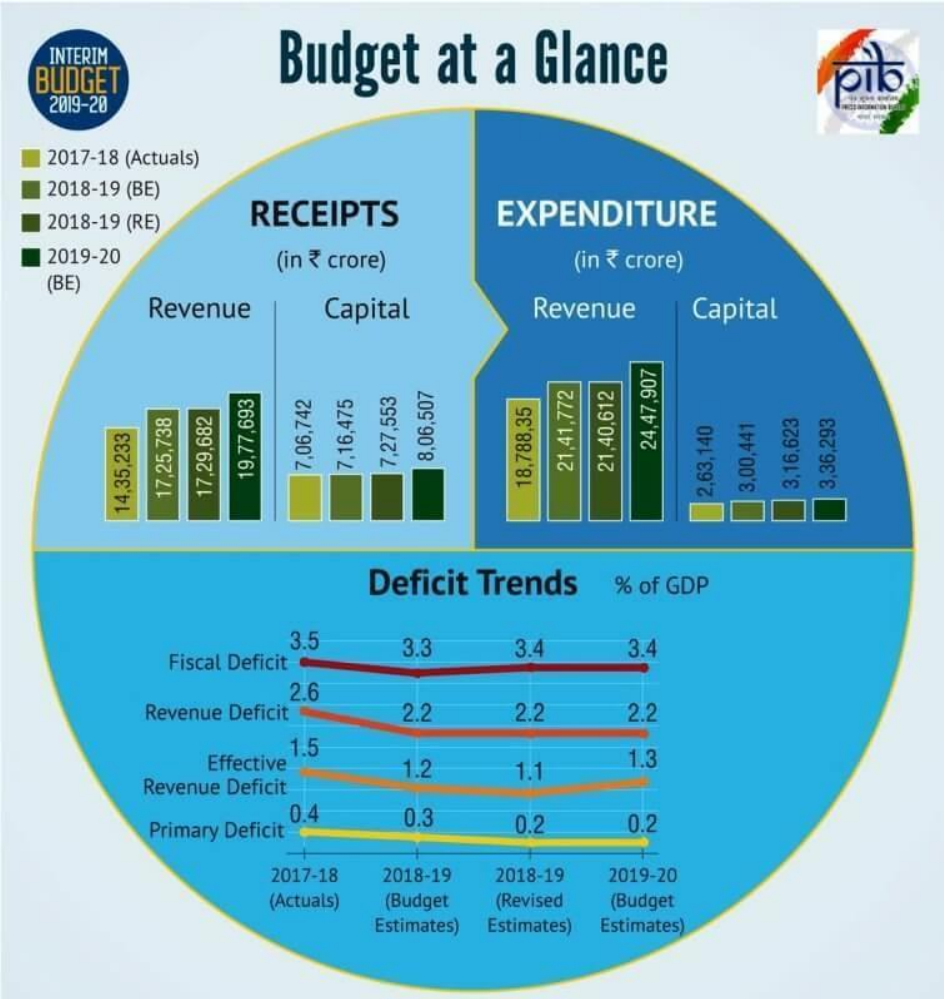

Fig. 3.8. Budget at a glance

Direct Tax Proposals made in the Budget 2019 (FY mainly includes 2019-20/ AY 2020-21)

1. Full tax rebate for individual taxpayers having total income up to Rs. 5 lacs.
2. Enhanced standard deduction.
3. Allowing self-occupation of two residential house properties without tax liability for national rental income of second property.

4. Increase in threshold for TDS on interest from bank/post office deposits.
5. Increase in threshold for TDS rent.
6. Allowing investment in two residential houses for claiming capital gains exemption under section 54, etc.

3.3 Proposed Frameworks for each Case Study

3.3.1 Cohesive Framework for Sustainable Water Governance

Governing water sector incorporate the enactment, initiation and implementation of water policies, legal institutions, simplification of the tasks and duties of government in relation to water management. The aftermath relies on how the participants behave in relation to the regulations and responsibilities that have assigned to them. The water management domain is a component of extensive social, political, administrative and economic advancement [38] and therefore, it is influenced by the decision of collaborator outside the water sector. Literature exhibits a lot of theoretical work [39,40,41] that has been done in order to understand a closer affinity between water and governance system. But no conventional and conceptual schema or tech-driven solutions has proven to be proposed until now. This research attempts to perceive the cooperative unification of water and governance system by propounding a unified framework of water governance. Fig. 3.9. illustrates the association capturing the two spheres i.e. governance and water management.

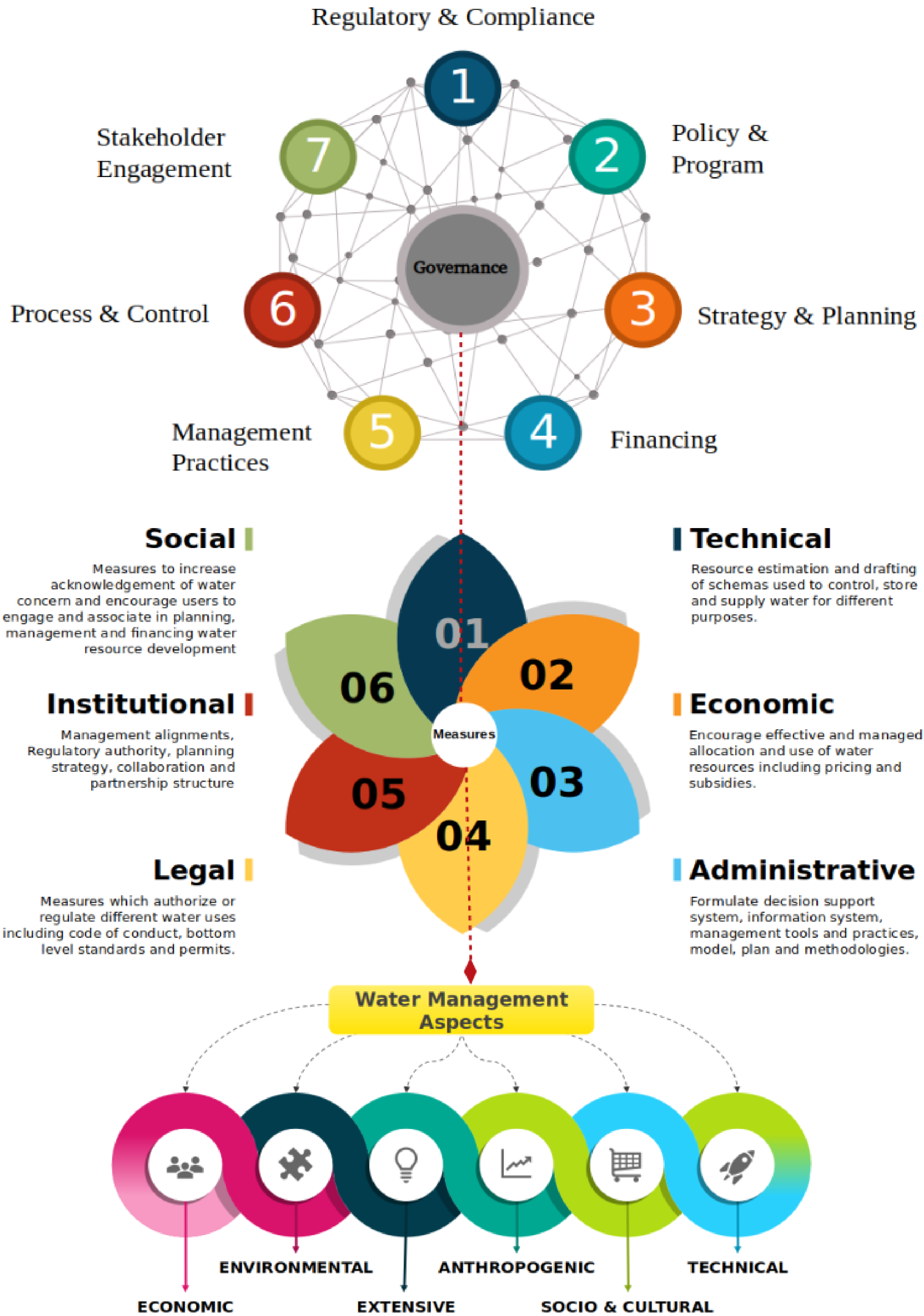

Fig. 3.9. Unified framework of water governance

Regulatory & Compliance, Policy & Program, Strategy & Planning, Financing, Management Practices, Process & Control and Stakeholder Engagements [38]. All the modules are interlinked and interdependent for the effective functioning of the system. Simultaneously, several aspects of water management and water conservation are elaborated under E^3AST analysis and are broadly classified into six divisions: Economic, Environmental, Extensive, Anthropogenic, Socio & Cultural, and Technical. Table 3.2 exemplify E^3AST analysis of water management aspects. Each aspect comprises of various factors that influence water management in our day to day life. Each factor plays a vital role and contributes to planning, developing and distributing the water resources throughout the entire geographical region.

Table 3.2: E^3AST analysis of water management aspects

E^3AST	**WATER MANAGEMENT ASPECTS**	**FACTORS INFLUENCING WATER MANAGEMENT**
E^3	Economic Aspects	• Change in world energy prices; • Changes in economic conditions; • Intensity of local and regional economies
	Environmental Aspects	• Change in precipitation or evapotranspiration; • Change in temperature or biological diversity; • Water quality deterioration; • Global climate change

	Extensive Aspects	• Municipal and Industrial sector; • Agricultural and rural activities; • Domestic household water use
A	Anthropogenic Aspects	• Resource overuse; • Pollution overload; • Water under-replenishment; • River mutilation; • Geologic Disruption
S	Socio-cultural Aspects	• Spatial distribution of people in the region; • Trends in population growth and distribution; • Patterns of water use; Industrialization; • Urbanization; • Lifestyle changes; • Deforestation
T	Technical Aspects	• Computerized irrigation management techniques; • Pumping technology; • Construction of dams; • Water related ecology

All the aspects are crucial and demand deliberation for the nourishment of sustainable water governance. The main cause for concern from a governance frame of reference is that they are generally treated in isolation and the connection between them are often overlooked. Several measures [42] were also integrated into the framework, viz. Social, Institutional, Legal, Technical, Economic and Administrative. These measures act as a linkage between governance and water management in order to ensure better policy formulation across the water sector. Considering the digital reformation, the traditional water governance prototype has been redesigned to intuitive and agile water governance model. In this series, we propose a pioneering sustainable water governance framework which consolidates water management and governance for public and community welfare. Two key elements, social and sentic are incorporated into a unified framework in order to achieve sustainable water governance.

Fig. 3.10. delineates the essential s-elements of the sustainable water governance structure. These elements illustrate water governance as follows:

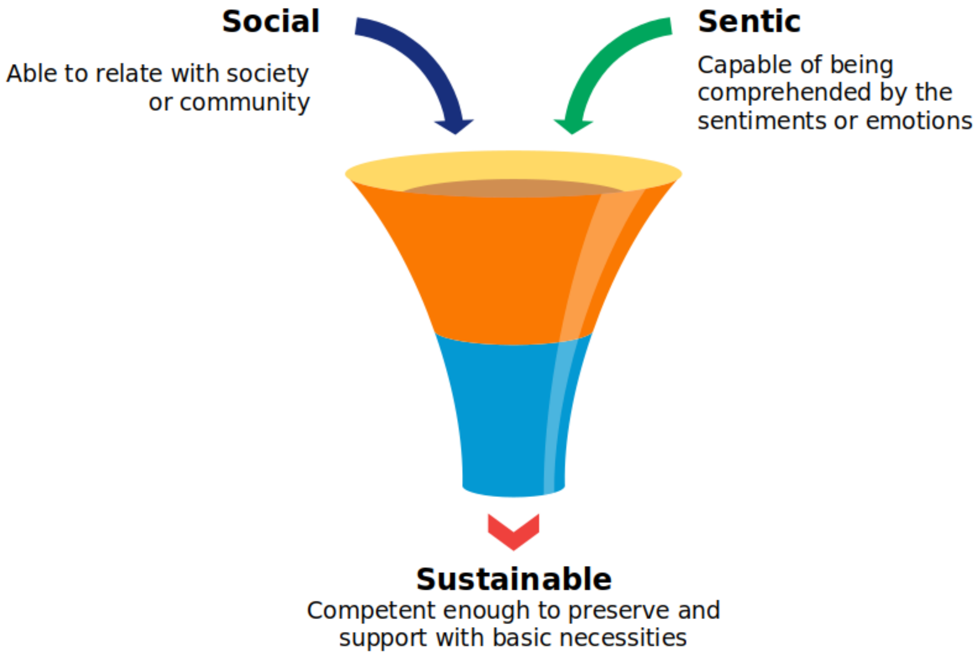

Fig. 3.10. Essential S-elements of sustainable water governance

- *Sustainable:* competent enough to preserve and support with core requirements.
- *Social:* relating to society or community.
- *Sentic:* capable of being comprehended by the sentiments or emotions.

Fig. 3.11. illustrates the correlation between governance and water resulting in a cohesive framework of Sustainable WAter Governance.

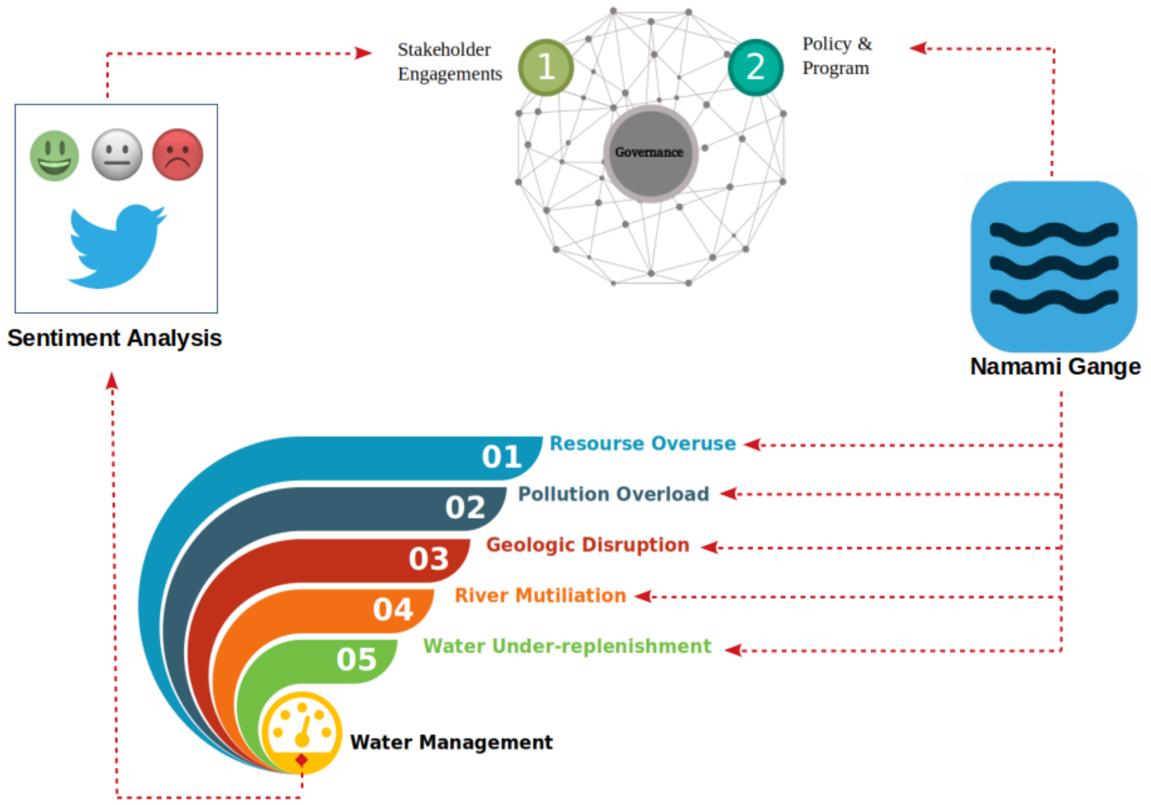

Fig. 3.11. Cohesive framework of sustainable water governance

The framework explains the interrelation and interdependence between governance and water management. In this paper, we elucidate the correlation among two components of governance i.e. (i) Stakeholder Engagement and (ii) Policy & Program through various aspects of water management. Namami Gange (meaning 'Obeisance to the Ganges river') initiated by the GoI [43] has been selected as a case study for this conduct. The related policy has been launched keeping anthropogenic factor as one of the major areas of concern and thus, it is selected as a part of policy evaluation in the cohesive framework of Sustainable WAter Governance. The programme was launched with the aim to rejuvenate river Ganga [34] by collaborating the current ongoing efforts and planning under it to a concrete future plan of action. Consequently, reflecting the public perception and opinion concerning the programme is a key aspect for the purpose of deciding subsequent measures and margin for refinement. In similar fashion, several dimension of water management & water uses are connected with other basic modules of governance to perform opinion mining for the valuation of any programme and to assess folk's reaction (positive or negative, praise or criticism). This collaboration of water and governance thus accelerate public and environmental welfare.

3.3.2 Socio-Affective Framework for Budget 2019

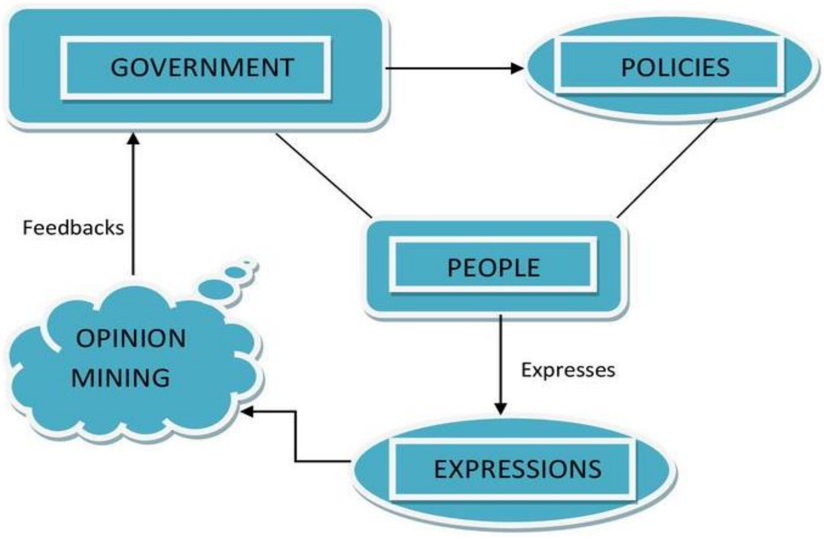

Fig. 3.12. Socio-Affective framework for smart governance

Twitter is a very convenient application for data acquisition as there are significant numbers of micro messages, also called as tweets, and retrieving them is technically naive relative to web scraping. Tweets are extracted using an API offered by twitter. These tweets can be used to extract opinions using corpus-based methods, machine learning methods, swarm computing methods [44]. In this framework, we have used supervised learning and deep learning methods to perform the required task for us. As constituted in the statistics and facts exhibited in the last section, the social media platform represent itself as a 'big data' source of people's voice. If Government organization can consistently keep a check on the pulse of its inhabitant, it can lead the way for more effective governance. Social sentiment analysis can be a very effective instrument to attain the same. It can address the following questions which Government organization would be very keen to get an answer:

1. How does citizens perceive new initiative, policies and programme?
2. What are the benefits and pitfalls of the programmes?
3. Is it possible for the government agency to replicate the positive attributes of a distinct programme to other programmes as well?
4. Is there any negative gabble that the organization should reciprocate?
5. Does the agency pay attention to the concerns of general mass when formulating administrative decisions?

Answers to such questions would assist agencies to refine their strategies to redress and tackle specific concerns; remodel their communication and community outreach programs to manifest any fallacies; facilitate an insight on how its programmes and initiatives is apprehended by its relevant stakeholder groups; identify best practices from positively perceived programmes and replicate it in others; formulate an effective performance model; and devise exhaustive social business strategy.

CHAPTER 4

CASE STUDY EVALUATION

This Chapter elucidates a machine learning based evaluation model for each case study. It further describes the datasets used and highlights the observations and findings.

4.1 Namami Gange

4.1.1 Datasets

The python script employing query searching option for the hashtag #NamamiGange has been executed for tweets collection. The search query fetches tweets on the #topic and is stored as a .csv file for further processing. A count of 1497 tweets for a duration of three months have been gathered. The collected data is pre-processed which results in clean and transformed data.

4.1.2 Opinion Mining Model

The proposed predictive analytics model comprises of four fundamental statistical steps, namely, data gathering and pre-processing, feature extraction and feature selection, prediction analysis and classifier performance prediction. Fig. 4.1 represents the systematic functional flow of the proposed model.

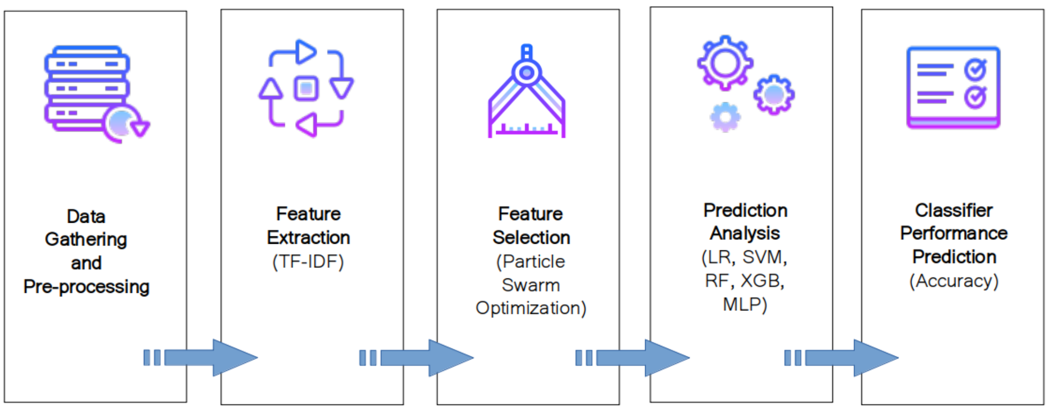

Fig. 4.1. Systematic flow of the predictive model

4.1.3 Observations and Findings

This section emphasizes the outcome and observations related to the performance of the proposed model. Opinion mining has been performed over a dataset of 1497 tweets of Namami Gange by implementing supervised learning techniques. Table 4.1 illustrates the opinion polarity of tweets collected for Namami Gange under three categories i.e. Positive, Neutral and Negative. Fig. 4.2. illustrates the opinion polarity distribution of tweets. Out of total 1497 tweets, 29.9% are positive, 20.2% are negative and 49.9% are neutral. Result reveals that approximately 30% of the citizen are in support of the programme whereas approximately 20% of citizen does not agree with the programme. The remaining 50% of opinion are neutral as they are informational in nature (updates of the latest issues of this mission).

Table 4.1: Opinion polarity of tweets

OPINION POLARITY	TWEET COUNT
Positive	448
Negative	302
Neutral	747
Total	1497

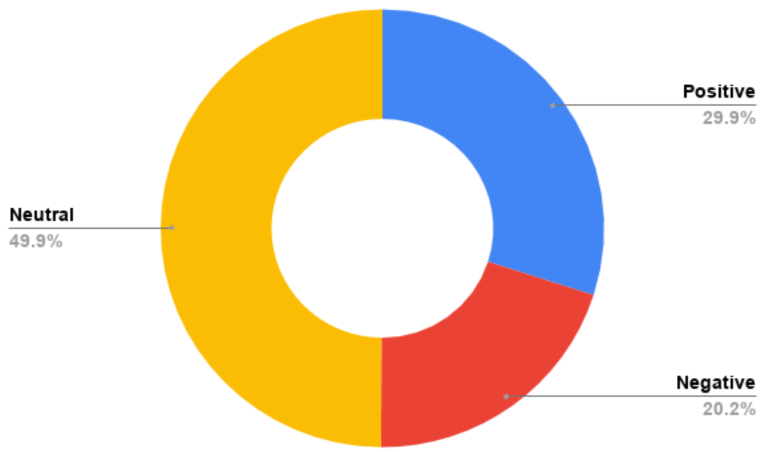

Fig. 4.2. Opinion polarity distribution of tweets

The results also capture the effect of optimized technique using TF-IDF with PSO over conventional technique using only TF-IDF. The standard measures of evaluation (accuracy) have been used to analyse the performance of the classifier.

Table 4.2: Evaluation metrics for classifiers

MACHINE LEARNING CLASSIFIER	ACCURACY (%)		CHANGE IN ACCURACY (%)
	CONVENTIONAL TECHNIQUE (TF-IDF)	OPTIMAL TECHNIQUE (TF-IDF + PSO)	
LR	88.23	98.51	10.28
SVM	91.57	99.49	7.92
RF	91.65	98.79	7.14
XG Boost	91.88	98.89	7.01
MLP	86.07	92.43	6.36

The result illustrated in Table 4.2 determines that XG boost with conventional approach performed with a maximum accuracy of 91.88%. Next was random forest with an accuracy of 91.65% followed by SVM of 91.57%. Considering the optimal approach SVM classifier achieved the best performance among all with an accuracy of 99.49% followed by XG boost with 98.89%. Random forest scored adjacent to XG boost with an accuracy of 98.79% whereas multilayer perceptron has achieved least accuracy of 92.43%. Consequently, logistic regression has shown the maximum accuracy increase of 10.28% and multi-layer perceptron has shown the least accuracy increase of 6.36%. Fig. 4.3. represents a comparative analysis of the result for each classifier.

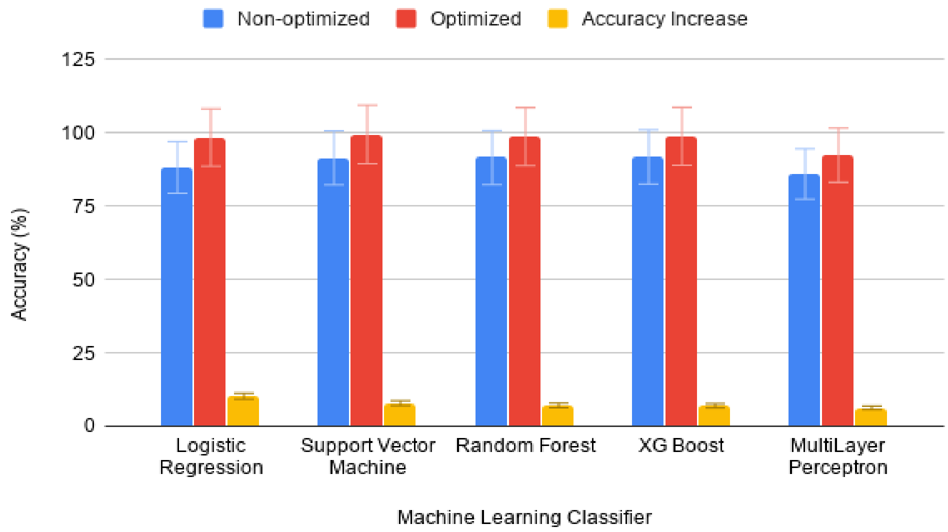

Fig. 4.3. Comparative analysis of classifier performance

Table 4.3 illustrate the contrast between the two approach of feature selection. Non-optimized approach uses the same number of features (864) for all classifier. The least number of feature (486) was selected using SVM whereas multi-layer perceptron chooses maximum number of features of 578. The above results by the use of PSO has helped to achieve a significant advantage: high classification performance and small number of feature selection with lower computing time.

Table 4.3: Contrast between feature selection approach

MACHINE LEARNING CLASSIFIER	NON-OPTIMIZED APPROACH #FEATURES	OPTIMIZED APPROACH #FEATURES	SELECTED FEATURES (%)
LR	864	551	66.77
SVM	864	486	56.25
RF	864	524	60.64
XG Boost	864	503	58.21
MLP	864	578	66.89

4.2 Budget 2019

4.2.1 Datasets

The script was executed by passing keyword search parameter for the hashtag #Budget2019 to extract data related to the schemes considered in this work using twitter API resulting in numerous tweets comprising of the mentioned hashtag. A total of 50959 tweets were compiled for pre-processing.

4.2.2 Opinion Mining Model

The model used for the purpose of opinion mining is outlined in the fig. 4.4.

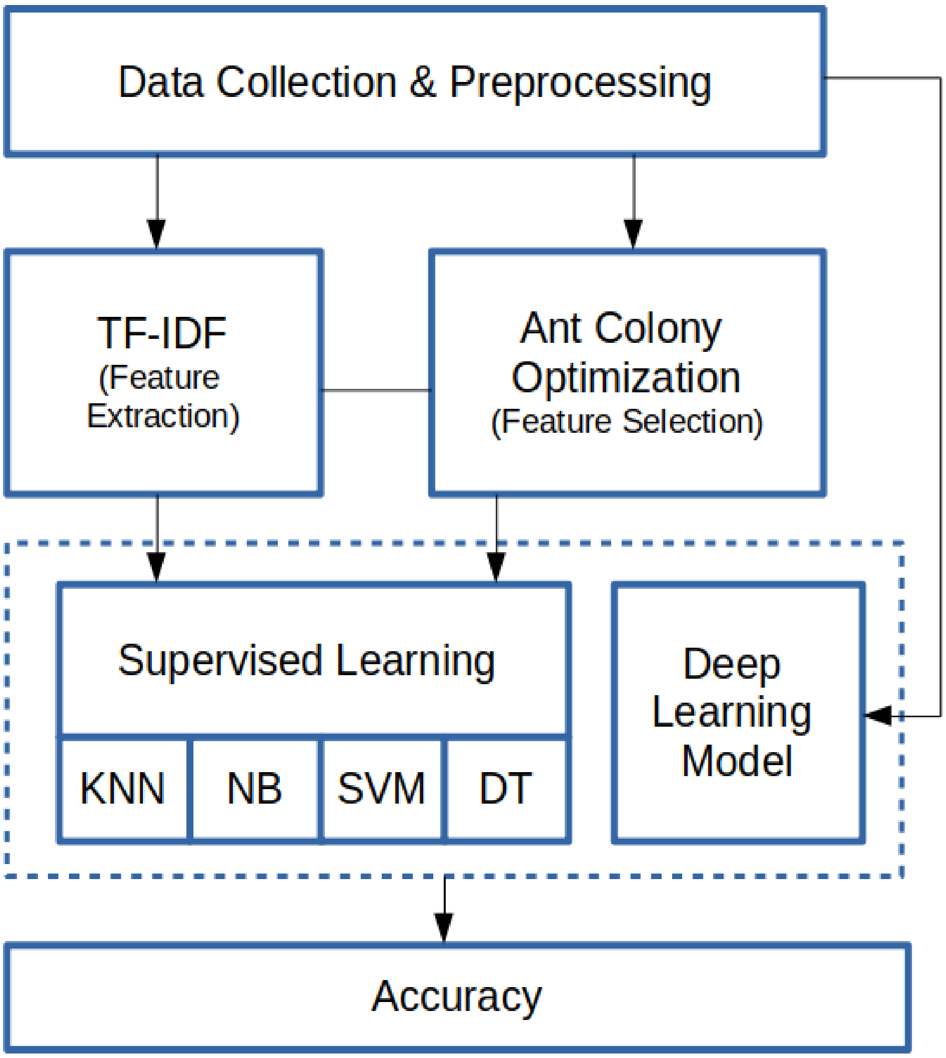

Fig. 4.4. Opinion mining model

4.2.3 Observations and Findings

This section summarizes the effect of optimized approach (using TF-IDF with ACO) over non-optimized approach (using only TF-IDF) on the classification performance employing accuracy as the statistical measure. It can be observed from table 4.4 that classification accuracy with non-optimized approach is observed by the DT, i.e. 91.92%. Next was SVM with an accuracy of 91.31% succeeded by kNN of an accuracy of 90.41%. Across all, NB achieved the least accuracy of approximately 89.71%. The best classification accuracy using optimized approach is observed by SVM, i.e. 99.47% succeeded by DT and NB. Across all, kNN achieved the least accuracy of approximately 97.72%.

Table 4.4: Classification accuracy of non-optimized vs. optimized approach

CLASSIFICATION ALGORITHM	NON-OPTIMIZED APPROACH (TF-IDF) ACCURACY (%)	OPTIMIZED APPROACH (TF-IDF + ACO) ACCURACY (%)	ACCURACY INCREASE (%)
NB	89.71	98.82	9.11
DT	91.92	99.31	7.39
SVM	91.31	99.47	8.16
kNN	90.41	97.72	7.31

Table 4.5 illustrate comparison of selected features by different approach. Non-optimised approach uses same number of features (847) for all the classification algorithms. The best classification accuracy of SVM was observed by considering 493 features based on optimized approach of ACO which was 58.20% and maximum was 549 (kNN) which is 64.81% selection.

Table 4.5: Selected Features of non-optimized vs. optimized approach

CLASSIFICATION ALGORITHM	NON-OPTIMIZED APPROACH #FEATURES	OPTIMIZED APPROACH #FEATURES	SELECTED FEATURES (%)
NB	847	529	62.46
DT	847	505	59.62
SVM	847	493	58.21
kNN	847	549	64.82

CHAPTER 5

RESULTS AND DISCUSSION

This chapter summarizes the comparative analysis of results achieved using different machine learning approaches for different case study in order to provide a holistic view of entire system. It provide the highlights of using optimized approach over non-optimized approach.

5.1 Results and Discussion

Table 5.1 shows contrast between optimized over non-optimized approach for percentage increase in accuracy.

Table 5.1: Contrast between accuracy of optimized over non-optimized approach for each government policy

GOVERNMENT POLICY	ML TECHNIQUES	NON-OPTIMIZED APPROACH ACCURACY (%)	OPTIMIZED APPROACH ACCURACY (%)	INCREASE IN ACCURACY
Namami Gange (Particle Swarm Optimization)	LR	88.23	98.51	**10.28**
	SVM	91.57	**99.49**	7.92
	RF	91.65	98.79	7.14
	XG Boost	**91.88**	98.89	7.01
	MLP	86.07	92.43	6.36
Budget 2019 (Ant Colony Optimization)	NB	89.71	98.82	**9.11**
	DT	**91.92**	99.31	7.39
	SVM	91.31	**99.47**	8.16

	kNN	90.41	97.72	7.31

SVM provides the best accuracy in both optimized approach, PSO as well as ACO. However, LR shows the maximum accuracy increase using PSO while NB shows the maximum accuracy increase using ACO. Table 5.2 shows the comparative analysis of feature reduction. SVM shows the least feature selection in both optimized approach.

Table 5.2: Contrast between features selection of optimized over non-optimized approach for each government policy

GOVERNMENT POLICY	ML TECHNIQUES	NON-OPTIMIZED APPROACH #FEATURES	OPTIMIZED APPROACH #FEATURES	FEATURE SELECTED (%)
Namami Gange (Particle Swarm Optimization)	LR	864	551	66.77
	SVM	864	**486**	56.25
	RF	864	524	60.64
	XG Boost	864	503	58.21
	MLP	864	578	66.89
Budget 2019 (Ant Colony Optimization)	NB	847	529	62.46
	DT	847	505	59.62
	SVM	847	**493**	58.21
	kNN	847	549	64.82

CHAPTER 6
CONCLUSION AND FUTURE SCOPE

This chapter elaborates the concluding remarks and the future scope that can extend the applicability of the work.

6.1 Conclusion of Research

In this paper, we attempted to know the polarity of the general user opinion extracted from twitter data. The result achieved is by far encouraging and promotes the scope of further work in this section. The overall analysis result demonstrates that the socio-affective framework has proven quite effective in determining the overall orientation of citizens towards government policies. It has effectively captured people's feedback which can be forwarded to the policy setter to make amends in the policies thus aiding the overall process. The feature selection approach based on ant colony optimization (ACO) and particle swarm optimization (PSO) is demonstrated that helps to removes irrelevant and redundant features while capturing more essential features from large collection of datasets.

The applicability of the case study incorporated in the conceptual framework served as a good practise to validate the proposed framework. It has effectively captured people's feedback which can be forwarded to the policy setter to make amends in the policies thus aiding the overall process. In conjunction with the main objectives, adverse shortcoming and challenges are also connected that either enhance the individual socio-economic condition or may influence the living standards of any community. Henceforth this work intends to assess the possibilities of using cutting edge technologies for social welfare which leads to strengthening the government citizen relationship. Feature selection using particle swarm optimization is demonstrated that helps to eliminate irrelevant and redundant features while preferring more essential features.

6.2 Future Scope

The future scope of this work seeks to enhance towards optimization by using other bio-inspired techniques. More quantifying approach for instance, hybrid or ontology-based may be incorporated to enhance the inclusive performance of classifiers. The feature selection can be further extended by examining other nature inspired algorithms.

References

[1] Tetsuya Nasukawa and Jeonghee Yi, Sentiment analysis: Capturing favorability using natural language processing. In Proceedings of the 2nd international conference on Knowledge capture, ACM, pp. 70-77, October (2003).

[2] A. Kumar, P. Dogra, V. Dabas: Emotion Study Of Twitter using Opinion Mining, Contemporary Computing (IC3), 2015, 285-290, IEEE.

[3] Akshi Kumar, Abhilasha Sharma: Systematic Literature review on Opinion Mining of Big data for Government Intelligence, Webology, 2012, 14(2).

[4] Akshi Kumar and Abhilasha Sharma: Socio-sentic framework for Sustainable Agriculture Governance, IEEE 2018.

[5] United Nations Developement Programme (UNDP) policy document. Governance for sustainable human development, New York: UNDP, 1997. Disponivel em: <http://magnet.undp.org/policy/default.htm>.

[6] Akshi Kumar and Abhilasha Sharma: Paradigm Shifts from E-Governance to S-Governance, The Human Element of Big Data: Issues, Analytics, and Performance, 2016, 213.

[7] Vu Dung Nguyen and Blesson Vaghese: Royal Birth of 2013: Analyzing and Visualizing Public Sentiment in the UK using Twitter, Research Gate, (2013).

[8] Social Media Statistics. http://www.statista.com

[9] Kun-Lin Liu and Wu-Jun Li: Emoticon Smoothed Language Models for Twitter Sentiment Analysis, AAAI, 2012.

[10] A. Kumar and T.M. Sebastian: Sentiment analysis: A perspective on its past, present and future, International Journal of Intelligent Systems and Applications, 4(10), p.1, 2012.

[11] A. Kumar and A. Jaiswal: Empirical Study of Twitter and Tumblr for Sentiment Analysis using Soft Computing Techniques, In Proceedings of the World Congress on Engineering and Computer Science, vol 1, 2017.

[12] Anant Arora, Chinmay Patil, Stevina Correia: Opinion Mining: An Overview, International Journal of Advance Research in Computer and Communication Engineering, 2015.

[13] Zhang, Y., Agarwal, P., Bhatnagar, V., Balochian, S., Yan J. (2013). Swarm intelligence and its applications. Sci. World J., vol. 2013, Art. no. 528069.

[14] Lim, S. M., & Leong, K. Y. (2018). A Brief Survey on Intelligent Swarm-Based Algorithms for Solving Optimization Problems. In Nature-inspired Methods for Stochastic, Robust and Dynamic Optimization. IntechOpen.

[15] Kumar, A., Khorwal R., Chaudhary, S. (2016). A Survey on Sentiment Analysis using Swarm Intelligence, Indian Journal of Science & Technology, vol. 9, no. 39, pp. 1–7.

[16] Swarm Intelligence. Retrieved from http://www.scholarpedia.org/article/Swarm_ intelligence

[17] Twitter. https://twitter.com/, 2019 (accessed 8 August 2019).

[18] Natural Language Toolkit. https://www.nltk.org/, 2019 (accessed 8 August 2019).

[19] Python Package Index, Natural Language Toolkit. https://pypi.org/project/nltk/, 2019. (accessed 8 August 2019).

[20] Natural Language Toolkit, Stemmers. http://www.nltk.org/howto/stem.html/, 2019 (accessed 8 August 2019).

[21] OpenNLP,PorterStemmer. https://opennlp.apache.org/docs/1.7.2/apidocs/opennlp-tools/opennlp/tools/stemmer/PorterStemmer.html/, 2019 (accessed 9 August 2019).

[22] Robertson, Stephen. Understanding inverse document frequency: on theoretical arguments for IDF. Journal of documentation 60.5 (2004): 503-520.

[23] Peter Wilson, H. Alan Mantooth, Chapter 10 - Model-Based Optimi-zation Techniques, Model-Based Engineering for Complex Electronic Systems, Newnes, 2013, Pages 347-367, ISBN 9780123850850, https://doi.org/10.1016/B978-0-12-385085-0.00010-5.

[24] Xiangping Meng, Zhaoyu Pian, Chapter 2 - Theoretical Basis for In-telligent Coordinated Control, Intelligent Coordinated Control of Complex Uncertain Systems for Power Distribution Network Relia-bility, Elsevier, 2016, Pages 15-50, ISBN 9780128498965, https://doi.org/10.1016/B978-0-12-849896-5.00002-7.

[25] Rahime Ceylan, Hasan Koyuncu, Chapter 7 - ScPSO-Based Multi-thresholding Modalities for Suspicious Region Detection on Mammo-grams, Soft Computing Based Medical Image Analysis, Academic Press, 2018, Pages 109-135, ISBN 9780128130872, https://doi.org/10.1016/B978-0-12-813087-2.00006-3.

[26] Menhas M.I., Fei M., Wang L., Fu X. (2011) A Novel Hybrid Bi-nary PSO Algorithm. In: Tan Y., Shi Y., Chai Y., Wang G. (eds) Advances in Swarm Intelligence. ICSI 2011. Lecture Notes in Com-puter Science, vol 6728. Springer, Berlin, Heidelberg.

[27] Dorigo M., Stutzle T.: Ant Colony Optimization. The MIT Press; 2004.

[28] C.J vanRijsbergen: Information Retrieval. 2nd edition. Londo, UK: Butter-worth-Heinemann; 1979. – 3.1.3

[29] Kotsiantis, Sotiris B., I. Zaharakis, and P. Pintelas. Supervised ma-chine learn-ing: A review of classification techniques. Emerging artifi-cial intelligence ap-plications in computer engineering160 (2007): 3-24.

[30] Kotsiantis, Sotiris B., Ioannis D. Zaharakis, and Panayiotis E. Pin-telas. Ma-chine learning: a review of classification and combining techniques. Artificial Intelligence Review 26.3 (2006): 159-190

[31] Water World, Urban Water Management in India. https://www.water-world.com/international/wastewater/article/16201696/urban-water-manage-ment-in-india/, 2019 (accessed 5 June 2019)

[32] Goudie, Andrew. The human impact on the natural environment. No. Ed. 3. Basil Blackwell Ltd., 1990.

[33] Van de Meene, S. J., Rebekka R. Brown, and Megan A. Farrelly. To-wards understanding governance for sustainable urban water manage-ment. Global environmental change 21.3 (2011): 1117-1127.

[34] National Mission for Clean Ganga, Namami Gange Programme. https://nmcg.nic.in/NamamiGanga.aspx/, 2019 (accessed 19 March 2019).

[35] Money Control, Clean Ganga is still a dream. https://www.moneycon-trol.com/news/business/economy/clean-ganga-is-still-a-dream-data-shows-no-real-work-done-2325841.html/, 2019 (accessed 10 July 2019)

[36] Press Information Bureau, Government of India, Ministry of Water Resources, Progress of Clean ganga Mission. http://pib.nic.in/newsite/PrintRe-lease.aspx?relid=137894/, 2019 (accessed 19 March 2019).

[37] Wikipedia, 2019 Interim union budget of India. https://en.wikipe-dia.org/wiki/2019_Interim-Union_budget_of _India

[38] Akhmouch, Aziza, and Francisco Nunes Correia. The 12 OECD prin-ciples on water governance–When science meets policy. Utilities pol-icy 43 (2016): 14-20.

[39] Brooks, David B., Oliver M. Brandes, and Stephen Gurman, eds. Making the most of the water we have: The soft path approach to water management. Earthscan, 2009.

[40] Araral, Eduardo, and Yahua Wang. Water governance 2.0: a review and second generation research agenda. Water Resources Manage-ment 27.11 (2013): 3945-3957.

[41] Cook, Christina, and Karen Bakker. Water security: Debating an emerging paradigm. Global environmental change 22.1 (2012): 94-102.

[42] Harris, Daniel, Michelle Kooy, and Lindsey Jones. Analysing the governance and political economy of water and sanitation service de-livery. London: Over sea development institute (ODI). Accessed on 27.01 (2011): 2013.

[43] Planning Commision India, Ganga Action Plan. http://planningcom-mission.nic.in/reports/E_F/Gangaactionplan.pdf/, 2019 (accessed 4 April 2019).

[44] R. Shah and R. Zimmermann: Literature Review In: Multimodal Analysis of User-Generated Multimedia Content, Socio-Affective Computing, vol 6. Springer, Cham, 2017

[45] The Digital Governance Initiative. (n.d.). Digital Governance Concept Retrieved from http://www.digitalgovernance.org/index.php/concept

[46] Ravi, K., & Ravi, V. (2015). A survey on opinion mining and sentiment analysis: tasks, approaches and applications. Knowledge-Based Systems, 89, 14-46.

APPENDIX

LIST OF PUBLICATIONS

Accepted Paper

1. H. Shekhar, and A. Sharma, Intelligent Learning based Opinion Mining Model for Governmental Decision Making. In *International Conference on Smart Sustainable Intelligent Computing and Applications,2020. Procedia Computer Science.* **(SCOPUS)**

Communicated Paper

1. H. Shekhar, and A. Sharma, SWAG - A predictive analytics framework for Sustainable Water Governance. In *International Conference on Sustainable Technologies for Environmental Management, 2019.* Sustainable Computing. **(SCI)**