Determining of Public Grievances- A Smart Way of City Management

Pratik Jayant Deshpande[1], Prof. Geeta Navalyal[2]

M. Tech Student, Computer Science & Engg, KLE DR M S Sheshgiri College of Engg & Tech, Belagavi, India[1]

Professor, Dept of Computer Science & Engg, KLE DR M S Sheshgiri College of Engg & Tech, Belagavi, India[2]

Abstract: "Data Mining" is a method of exploring huge databases with a view to develop distinct knowledge or information. "Smart city", it's a city that technologically advances and improves the quality of residents daily lives through analyzing real time data. The above mentioned terms describes an efficient process for determining public grievances based on the data set to analyze and predict the similar grievance nature for a city. The grievance consists of grievance category such as garbage, sewage drains, water supply etc., and is also composed of attributes like "latitude and longitude" of the grievance registered. Utilizing the above mentioned attributes and data, analysis is performed on real data collected for various cities using "Bounding Box" concept and clustering algorithm "K-means Algorithm". Results of cluster analysis show the comparison between cities and help the city planners to satisfy and fulfil the needs of citizens and make city sustainable.

Keywords: Data Mining, Smart City project, Google Map API, K-means clustering.

I. INTRODUCTION

Since grievances are a valuable supply of feedback to boost the infrastructure and condition of our town, the citizens of the city would have grievances with relevancy about their surroundings, infrastructure of the city and on the other hand would no more accept the present system within which they need to go through procedure like reaching to the municipal corporation and standing in queue for hours, wasting valuable time and efforts. Cities and municipalities need to really understand how technology and intelligent connectivity can help them solve real world problems like not having enough water, pollution, transportation and traffic.

In today's world, everything is moving towards digital. To bring any information into the internet we need to build application on web. The present work determines the data registered by the citizens of various cities in India. Determination of public grievances helps the municipal authorities and city planners to take required action and resolve the problem to fulfil the requirements of citizens and form a sustainable city. A web application is built presenting another stage for sharing issues regarding electricity, water, garbage, drainage etc., which can be effortlessly utilized by the residents in an ideal way of keeping them ignorant from the procedures. Hence in this regard the approach is about determining the grievances through real world data integrated with "Google Maps" which provides latitude and longitude to perform analysis. The model is built using the concept of "Bounding Box" and "K-means clustering" to find number of similar grievance and also comparison of grievance nature for a city in India.

A. Introduction to Smart City
We all know that cities are not efficient, cities are not sustainable, cities are not resilient and now addressing these challenges become even more important because of the rapid population growth. Infrastructures are the backbone of cities and they do play a crucial role to address these challenges. Over the past years, technologies capable to address some of the challenges are rapidly spreading. This is particularly the instance of data and correspondence advances as of now being developed by universities, technology providers, technology start-ups and many others more. Cameras, sensors and other data collecting devices are producing unprecedented amounts of data for example data about movements of vehicles, mobility behavior or data about water consumption and energy by individuals and buildings. From these data, smart mobility, smart energy, smart buildings and many other smart infrastructures can emerge and ultimately this could lead to what is now called a "Smart City".

B. Introduction to Data Mining and Unsupervised Learning
Nowadays huge amounts of data are available in the field of internet, finance, retail marketing, science, and telecommunication. People have no time to look at this data and hence there exists the powerful technology called Data mining. "Data Mining" is characterized as the system of separating the knowledge from extensive amounts of data. Data mining extracts information from huge amounts of data. It automatically analyzes the data, classifies the data and summarizes the data into useful information. Database technologies, statistics, machine learning visualization,

information science are some of the techniques that go in hand with data mining. "Unsupervised learning" understands unlabeled information without having any predefined dataset for its preparation. Unsupervised learning is capable to access information and search for patterns. It is most regularly utilized for clustering identical input to consistent gatherings. Basic way to deal with unsupervised learning is different types of clustering methods.

In unsupervised learning, given 'X', there is no label to the data and given the different data points it is likely to cluster them or summarize or find patterns in them. There are different 'X's'; x_1, x_2, $x3…x_n$ represented as data and learning algorithm produces clusters which will group this data. Based on the similarity of the data items to each other certain groups among the data can be found out. This is called as unsupervised learning.

II. RELATED WORK

In recent years, the government has acquired more attention towards the needs of citizens and supports the usage of technology named "Information and Communication Technology". It has been proposed in [1] that, in order to make any changes in the orientation of government, it is necessary to figure out the requirement of citizens and fulfil their needs. Its main focus is regarding "Citizen Relationship Management" and hence data mining algorithms are used to determine the complaints of citizens. The analysis is done based on the complaints that frequently occur, the time gap between them and number of days, with the use of "Tehran Municipality" database. Hence, it helps in satisfying the needs of citizens and development of city.

Guanhua Wang presents two information serializing approaches utilized as a part of web applications i.e. XML and JSON [2]. Transmission between these two strategies is an issue in application improvement. Features of the above two approaches were determined and interpreted further using recursive algorithm. The algorithm used in this paper was helpful for the efficiency of web application. Hence web applications may have better compliance and flexibility in data communication. This paper uses the translator structure that ensures less effect on complexity of web applications and makes efficient translation between JSON and XML.

Urban planning is a technical and political process concerned with the design of urban environment. It refers to how the built environment influences human behaviour and interaction. It is a system that shapes our cities and life style of people within them. J. Cranshaw et al. [3] proposed a methodology for studying the structure and social dynamics of cities. The data used is "check-ins" data. The character of the urban area is defined not just by types of places found there, but also by the people who make the area part of their daily routine. Spectral clustering algorithm is used to form clusters of nearby venues, to discover the areas that are visited by the same people.

In Canada, a web application has been in working for Canadian residents, and it is available at "fixmystreet.ca". [4] The web application named as "fixmystreet" Canada is managed by the non-profit visibleGovernment.ca. The site is influenced by the "my society" site, and it has a few benefits. One of the important components of this web application is it guarantees to stay in contact with the city councillor by email or by utilizing "311" hotline number. Another feature of this web application is that the maps are available in other formats, as requested by the users. Another important advantage of this site is the state of "visual complaints". With every complaint, you will see a status bar where the red block indicates the transmission request, the yellow block indicates answered and green block is problem solved.

Another software application named "mark-a-spot" [5] is a completely responsive tool for mobile and desktop, used by the citizens of Germany for reporting civic issues. This software application has used a few new innovations at their specialized side and GUI side. This application includes following features: It gives the opportunity to the citizens to select between open street maps and Google maps. It uses AdHoc's picture, which is a great expansion of this application. "mark-a-spot" is growing progressively turning into a stage for online communication with an attention on geo-referencing. It additionally combines the framework with Facebook and Twitter. The addition of twitter makes the system a new user experience. Citizens can tweet their issues and tweet settled issues specifically to a twitter-account.

Nandine Akkari et al. presented two formulas for finding shortest distance between two locations [6]. The universe is especially important for navigation, but one thing it does not follow from the structure of solid geometry is distance calculation to the surface and is one of the major challenges. "Vincenty formula" and "Haversine formula" are two formulas used for distance calculations. As, the universe is not a perfect sphere, utilizing the above mentioned formulas gives the accurate results. The authors have presented and calculated distances by implementing on "location-based recommender system" and proved that "Haversine formula" provides more accurate results. Consequently, from the above related work, it has been noticed that majority of the methodologies are established on developing web applications. After all, in present work importance is on the cluster analysis for determining public grievances.

III. PROBLEM DEFINITION AND OBJECTIVE

A. Problem Definition

Smart city is a fact that we have data and information that is being shared across one department to another as well as back to the citizens and it's almost like a circular system which makes whole system even more effective and

sustainable. The long procedure in municipal corporation office to file any grievance is a major problem. The main approach is to change the present system and make the procedure smart by use of web technologies. Various concepts of data mining and unsupervised learning are performed on data to analyze grievances.

B. Objective

The main aim of the presented system is to register or file a grievance and notify to the concerned authority which helps them to resolve the issue and determines grievances based on the similar grievance category using concept of "Bounding Box" and "K-means".

C. Architecture Model

Architecture model is the development of decaying a generous complicated structure into substructures. The architectural model contains all the procedure of "Data Mining" process. The structure of the architectural model is depicted in Figure 1

The complete structure is composed of two modules; they are public grievances data, grievance category data group and clustering module. In the primary stage of this structure, population of a city and grievance category data is collected from municipal authorities and grievances data is collected from residents of city by registering a grievance through developed application. The data is composed of population count with respect to "Census 2011" for a city and grievance category like water supply department, electricity department, sewage drains etc. The public grievances data is composed of description of grievance, latitude and longitude fetched from "Google maps", email-id of the resident. In the secondary stage, data mining method is utilized to segregate the grievances.

To determine the grievances registered by the public, the concept of Bounding Box is used and subsequent algorithm is used.

➢ Bounding Box
➢ K-Means Clustering Algorithm

In last step of structure, the outcome is interpreted and distinguished. Hence, the efficiency of a city is recognized as a last step.
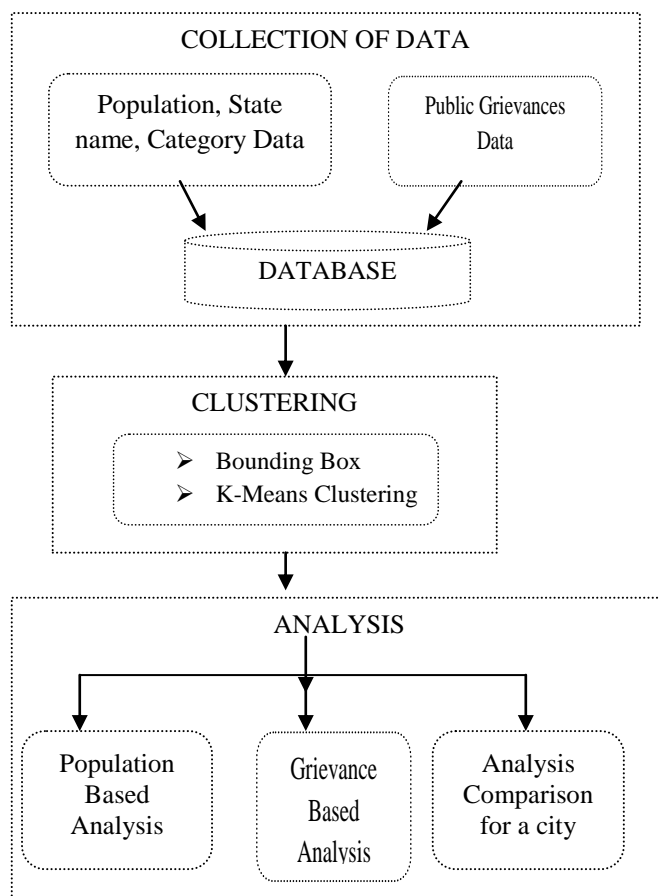

Figure 1. Architecture Model

IV. IMPLEMENTATION

A. Data Collection Procedure

With the progress of data and correspondence innovation, new techniques for collection of data and examination are discovered. This has advanced from phone reviews, paper-and-pencil etc., to the use of Internet. Although, internet is growing fast in contemporary life in many parts of the world, it remains moderately unused for essential collection of data. "Data collection" is essential step used to represent a procedure of assembling and gathering data. It is a precise collection of data being used for appropriate reason from different origin, which has been precisely recognized, registered and classified. For any "decision making" technique, data is the necessary input. The scope of collecting the data is to retrieve information, to preserve data, to form decisions regarding crucial problems and transmission of data to the residuals.

The collected data involves data gathered from residents of various cities. Data is gathered through web page, wherein residents of society file grievances regarding the issues they find around their surroundings. The issues may be water supply problem, garbage, drainage, electricity supply issue etc., the web page is integrated with "Google map" wherein citizens can register their grievance. The "latitude and longitude" are used for forming "bounding box" and the remaining data is used for determination of similar grievances and the number of grievances registered for a city.

After registering the grievances, the data is stored in the server called as "WAMP Server". It stands for "Windows, Apache, MySQL and PHP". It is regularly utilized for testing and developing the web interface, however may likewise be utilized to serve live sites. The data is stored in "MySQL' database and it consists of certain attributes. The attributes considered are title of grievance, category of grievance, grievance description, latitude and longitude fetched from the integrated Google map and email-id of citizens.

B. Google Map Bounding Box

A Google map gives a very responsive, natural mapping interface that provides detailed aerial images of streets, intuitive information and integrated mapping. Additionally, embedding of map controls provides end users to manage navigation of maps. End users can likewise perform outline using the keyboard keys and in addition by dragging the map by means of mouse.

Suppose there are various markers on "Google Map" and it's necessary to be shown such that they appropriately fir within viewport. It is possible to explore analysis with respect to mid point and level of zooming. Thus, consider the possibility that markers are selected dynamically and this can be illustrated by the concept of bounding box.

Bounding box is a field representing a pair of latitudes and longitudes. Latitude and longitude is composed of decimal number ranging from "-90.0 and 90.0" as latitude and longitude ranges from "-180.0 and 180.0".

Bounding box can be represented as

> < "minimum longitude", "minimum latitude",
> "maximum longitude", "maximum latitude" >

To begin with, it is necessary to determine Bounding Box i.e. equivalent "north-east" and "south-west" edges are determined. For this situation, the directions of marker point that is maximum to south and point that is maximum to west gives the directions to "south-west" edges. Similarly, for "north-east" edge. Google Map APT assumes essential part as Bounding Box and Google Map APT requires API key to place the Google Map during the advancement of any web application.

Google Map API proceeds with following URL

> **"http://maps.googleapis.com/maps/api/key=?"**

And outcome possibly be one or other following formats
json: illustrates "JavaScript Object Notation" output.
XML: illustrates XML output format.
Bounding Box uses Google Map API and the parameters required are key i.e. API key to utilize the Google Map service and next parameter is "bounds" which form viewport within Bounding Box and helps to bias the outcome more remarkable.

C. Clustering

After collection of data and obtaining grid boundaries, the subsequent step is clustering. The data is clustered by implementing the clustering algorithm. The algorithm used to perform analysis is "K-means clustering algorithm".
Clustering is utilized for many applications such as, it can be used to organize data automatically, used to understand hidden structure in data and sometimes clustering is used in pre-processing for further analysis of data.

K-Means Clustering Algorithm

The clustering method used is Partitioning method and clustering is based on data partitions. Assume collected data as "n" objects and method of partitioning develops "k" groups of data. Partitioning method uses following "k-means algorithm". In this algorithm 'k' is given and also set of objects or instances are given. Let 'X' be comprised of 'm' objects $<x_1, x_2 \dots x_m>$ and consider that each object is described in terms of 'n' features i.e. x_i can be composed of 'n' features as $<x_{i1}, x_{i2}, \dots\dots, x_{in}>$. Thus, the output is to produce 'k' clusters comprising of $<s_1, s_2 \dots s_k>$. In this particular algorithm, the output produced will form a partition of object i.e. each part cluster is disjoined and together they cover the entire set of objects and each cluster is described by the cluster center.

The following shows the process of K-means Algorithm:

TABLE I   ALGORITHM

1) Select 'k' points as initial centroids.
2) Repeat
   ➢ From 'k' clusters by assigning each point to its closest centroids.
   ➢ Re-compute the centroids of each cluster
3) Output: produce 'k' clusters such that objects within a cluster are similar to each other.
4) Until convergence is satisfied.

V. RESULTS AND ANALYSIS

The data collection is represented in the preceding stage and through examining these collected data; the subsequent results can be procured.
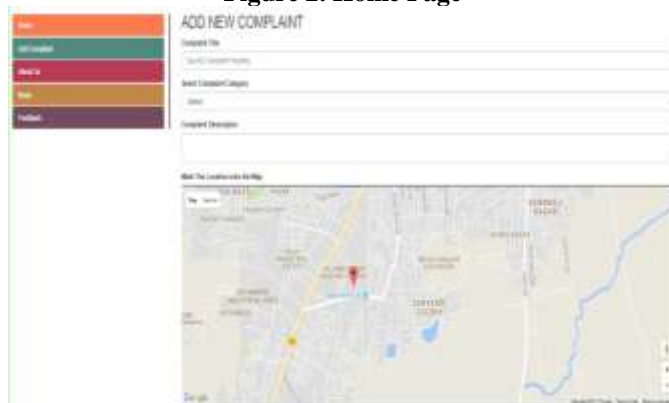

Figure 2. Home Page

Figure 3. Grievance Registration

The above figure shows the home page of web application. It comprises of a form wherein the people of city register their grievances to notify the concerned authorities through this web application. The form is integrated with Google Map to provide exact location of grievance. The latitude and longitude obtained are utilized further for determining similar grievances in comparison with other grievances registered by the people of other city.

Figure 4. Notification message

Once the user registers a grievance and submits, the data is inserted successfully into the database and a notification message is sent on the mobile of the concerned authority.

Figure 5. User Interface

The above figure shows the "JAVA" user interface that is used by the concerned authority for determination of grievances.
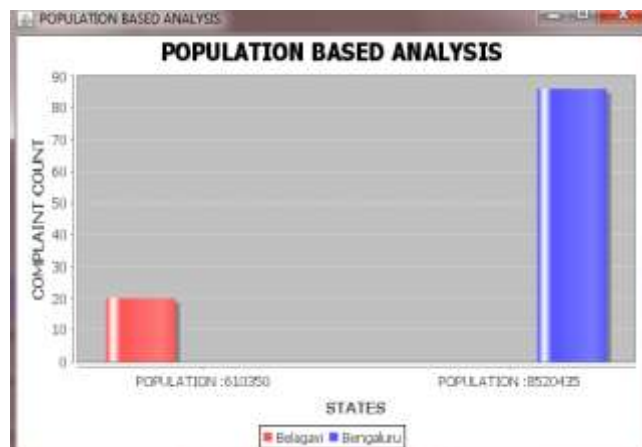

Figure 6. Population based analysis

The above figure shows the bar chart representing the analysis based on the population. The population count is determined with respect to "Census 2011"
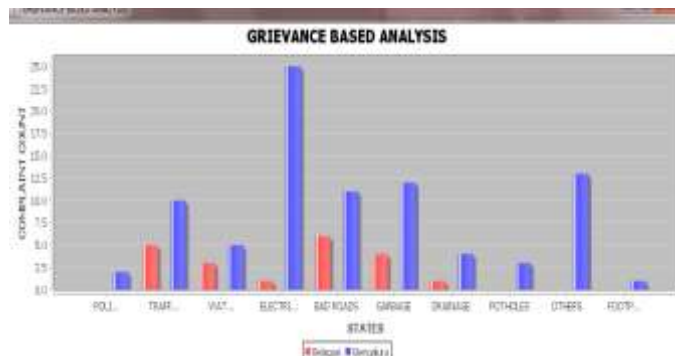
Figure 7 Grievance based analysis

The above figure shows the analysis based on the grievances. It is represented by the bar chart which shows the grievance count for various grievance categories. It also distinguishes between cities with which the grievances are registered. As shown in the figure it displays the bar chart for two cities i.e. "Belagavi" and "Bengaluru" and grievance count can be determined for a grievance category.

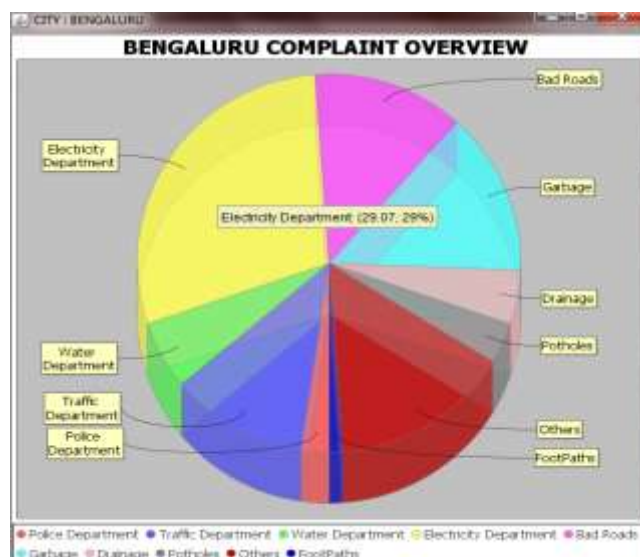

Figure 8. Analysis for Belagavi city

Figure 9. Analysis for Bengaluru city

134

The figure 8 and figure 9 shows the analysis of "Belagavi" and "Bengaluru" region in the form of 3D pie chart and displays the accuracy of grievance nature for a category. The results obtained will help to distinguish the similar category of grievances registered in a city and city planners could compare which city has highest similar type of grievance registered.

VI. CONCLUSION

This preparatory research includes new innovations for improvement of society effortlessly and makes easy to report the issues with an ease of developing web application. With the beginning of new advances and modernization, human endurance has minimized. The proposed "K-means clustering algorithm" forms clusters of grievances with concept of Bounding Box. The algorithm is implemented on the attributes and dataset used to construct the framework. Analysis is performed and behavior of grievances is recognized in selected cities. Hence, developed web application and analysis would help citizens to move towards digital world and save their valuable time and help city authorities to make city smart and sustainable. In the future, analysis can be done to determine grievances for different areas in a city.

REFERENCES

[1] Ahmadvand, B. Bidgoli, and E. Akhondzadeh, "A hybrid data mining model for effective citizen relationship management: A case study on Tehran municipality"

[2] Guanhua Wang, "Improving Data Transmission in Web Applications via the Translation between XML and JSON", in Third International conference on communications and mobile computing, 2011, IEEE

[3] J. Cranshaw, R. Schwartz, J. I. Hong, and N. Sadeh, "The livehoods project: Utilizing social media to understand the dynamics of a city," in International AAAI Conference on Weblogs and Social Media, 2012, p.58

[4] https://www.fixmystreet.ca/

[5] https://www.markaspot.de/en

[6] Hagar Mahmoud and Nandine Akkari, "Shortest Path Calculation: A Comparative Study for Location-Based Recommender System", 978-0-7695-5832-5/16, 2016 IEEE, DOI 10.1109/WSCAR.2016.16

[7] Han, Jiawei, Micheline Kamber and Jian Pei. Data mining: concepts and techniques, Elsevier

[8] http://www.ichangemycity.com/mumbai

[9] http://www.ichangemycity.com/bengaluru

BIOGRAPHIES

Pratik Jayant Deshpande is M.Tech student in the Department of Computer Science & Engineering, KLE Dr. M S Sheshgiri College of Engineering Belagavi. He completed his Bachelor of Engineering in Computer Science & Engineering from KLS Gogte Institute of Technology, Belagavi in the year 2015. His research interests are in Data Mining.

Prof. Geeta Navalyal is Associate Professor in the Department of Computer Science & Engineering, KLE Dr. M S Sheshgiri College of Engineering and Technology, Belagavi. She completed her B.E. in Electronics and Engineering from KLS Gogte Institute of Technology, Belagavi and M.Tech in Computer Network Engineering from VTU.