Proactive Measures to Avoid Illegal Construction

Abstract: Focus is to improve transparency between Government and Citizens and to deliver the government rules, regulations and plans to citizens by using the growth of technology effectively as E-Governance. Especially a proactive measure to avoid illegal construction of buildings on the dried water bodies, river bank areas and agricultural lands as plots and apartments due to urbanization.

To avoid registration of agricultural land for commercial purpose. To block the approval of layout of buildings, new electricity (Electricity Board) connection, new water supply connection and loans for those lands. Brokers can be avoided by this method. We can also avoid corruption. This system will be very helpful on the time of rainfall season, to avoid flood and also save water by using these dried water bodies which will indirectly reduce the water scarcity problem for drinking and also helpful to farmers for agriculture. As all of us know that, “Agriculture is the Backbone of India”. So, we can reduce the price of all the food materials.

Keywords: Data Mining, Social Media Streams, Supervised Learning, Machine Learning, Probability, Sentiment Analysis, Data Management, Web Services, Smart Phone, Android Phone, Android, Application.

I. INTRODUCTION

In Chennai, there was a heavy rain on the November and December month of the year 2015 which caused a flood in many areas of Chennai, Thiruvaloor, Kancheepuram, and Kadloor. More than 280 peoples died in this disaster. In this disaster economic loss was measured as 300 crores and property loss was measured as 90,000 crores. So, overall loss by this flood was measured as 90,300 crores. Major reason for this disaster was identified as the illegal construction of the buildings on the dried water bodies and river bank areas due to the urbanization. In Chennai, the average measurement of rainfall was about 1218.6 mm. On 2nd December 2015, it raised 294.1 mm of rainfall on the November and December months of the year 2015 by Regional Metrological Centre (RMC) and Indian Metrological Department (IMD) of Chennai. In National Geographic Channel, these rainfall and flood are ranked as World's 8th biggest National Disaster on the —Mega Floods‖ show. In Chennai, Thiyagaraaya Nagar (T.Nagar) is formed by occupying the dried water bodies. In Chennai, water bodies are decreased from 650 to 30 water bodies for the last 100 years. It is shown in Fig. 1. In Chennai alone, there are 15 lakhs buildings constructed by violating the law.
A mouli vaakam Apartment building collapsed on 24/06/2014. In this disaster, more than 61 peoples died. Overall loss in this incident was measured as 20.28 cores. Main reason for this incident was the construction of access floors than a number of floors approved by Chennai Metropolitan Development Authority (CMDA).

Due to urbanization, Agriculture lands were decreased by 80 % in Chennai and 35 % to 40 % in other districts of Tamil Nadu for the last 30 years. Agricultural lands were systematically converted into commercial lands. It is estimated that 13.37 lakhs of plots. The areas mostly affected by this conversion are South Chennai, Kancheepuram, Nagercoil, RS Mangalam. The major reason for this reduction is an extension of city and loss in various forms to the farmers in agriculture.

II. RELATED WORKS

Tedeschi and F. Benedetto [1] proposed a cloud-based big data sentiment analysis application for brand monitoring and analysis in social media streams. They have designed and developed a user-friendly application as a cloud-based service derived from the Platform as a Service (Paas) model. They used this application to get the benchmark of the brand based on the Reviews, Blog Posts, Tweets by the Consumers in all over the world. They also extracted the sentiment and polarity from tweets using the SentiWordNet algorithm. They have also used this application to get the customer needs and expectations and satisfy them. Twitter is the only social network they have considered.

João Rosa, Cláudio Teixeira and Joaquim Sousa Pinto [2] proposed an Information and Communication Technologies (ICTs) System in Singapore. This Centralized information system is based on multimedia kiosks, named Automated Traffic Offence Management System (ATOMS) database, which manages the offender's information. This E-Justice system enables citizens to pay their fines without addressing to the court. The initial architecture is poorly planned. So, the entire project may be at risk.

Kalliopi Anastasopoulou and Spyros Kokolakis [3] proposed an E-governance system that initiates the collection and processing of personal data of people’s financial transactions. The intention of this new service offered by the Greek Ministry of Finance called ‘tax card’. Tax card is used to collect information about everyday purchases and aims to diminish tax avoidance. The effect of cultural bias is mostly neglected by policymakers. They fail to address the mindset of specific cultural groups that object these technologies. Their analysis was limited to the Greek region and only to one specific e-government initiative, the tax card.

Spyridoula Lakka and Teta Stamati [4] proposed a model, based on three socio-economic theories namely, institutionalism, endogenous and exogenous growth. Using this framework critical factors are identified, while their impact is evaluated with an econometric analysis on secondary, country level data. Institutionalism is a focus on establishing structures, rules, norms, routines, law and political rights as authoritative guidelines for social behavior. Endogenous Growth outlines the potential of economic growth that is generated within a system, as a result of internal processes, as for example technological advancements and not external, as for instance through trade. Exogenous Growth is the Trade based on Information and Communication Technology (ICT). Use of advanced technologies, education, technological openness and institutional quality in terms of government effectiveness is the driver that led e Gov growth. The increased imports and exports of technologies positively influence countries to create favourable conditions to use new technologies. It is not merely an advanced technological tool.

Muhammad Ovais Ahmad, Jouni Markkula and Markku Oivo [5] proposed a plan raises awareness, attracts more citizens to make use of e-government services and facilitates better understanding and delivery. The goals of the e-government of Pakistan are to increase efficiency, effectiveness, transparency, and accountability in decision making in addition to enhancing delivery of public services to its citizens both efficiently and cost effectively. The success of e-government services depends on government support as well as on citizen’s adoption. It focuses on to fill this gap by exploring the challenges and barriers of e-government services from the user’s perspective. The citizens lack knowledge about the new e-government services. So, The Pakistani government should raise awareness throughout the country regarding their e-services through different advertising channels.

John C. Bertot, Paul T. Jaeger and Justin M. Grimes [6] proposed a method called —Using Information and Communication Technology (ICT) which is cost-effective and convenient means to promote openness and transparency and used to reduce corruption. They developed the measures for transparency. It is for a long term process to obtain the success. It evaluates existing systems for portability and expansion. Reuse rather than reinvent. The social technologies available today are transformative in general and with regard to transparency and anti-corruption in particular.

III. EXISTING SYSTEM

In the existing system, Government of Tamil Nadu has an online portal of registration department of Tamil Nadu and land registration e-service which gives a detail by getting the input from user like Zone Name, District Name, Village Name and Street Name or Survey No. to get the Encumbrance Certification (EC). It displays only informative details but, the administrative details were not available.

In 2009, Section of Land Act 22-A was published on gazette. But, the implementation date was not mentioned on it. In 21/10/2016, Section of Land Act 22-A was implemented by Tamil Nadu State Government.

In Tamil Nadu, Layouts for Home up to 3 floors (including Ground Floor) for lands within 1000 Sq.m to be only approved by Panchayat. In Tamil Nadu, Layouts for Schools, Colleges, Marriage Halls, Hospitals, Clinic, Factory, Industry and other commercial Buildings on lands above 1000 Sq.m. to be only approved by Department of Town and Country Planning (DTCP). Similarly, layouts should be only approved by Chennai by CMDA (formerly called MMDA) for Construction on

Data Extraction will be used to pull the tweets from R code using OAuth facility by Submitting Twitter credentials. Similarly, Facebook Posts will be pulled from R code using FBoauth facility by Submitting Facebook credentials. Corpus Cleaning will be used to clean and present the extracted data to R engine. Lexical Analyzer analyzes the Extracted data as positive or negative data separately by matching it with words on the positive-words.txt and negative-words.txt files. Machine learning will take place to project the result. Bayesian Learning will happen. The end result will be projected Sentiment or Opinion projection, polarity analysis (trend towards positive or negative trait), word cloud most frequent words spoken. The Architecture of the Data Mining Module is as shown in Fig. 2.

IV. PROPOSED APPROACH

The proposed system is derived from the sentiment and polarity analysis of user data retrieved from Facebook posts and tweets from Twitter. The proposed system will incorporate the nature of land details which will be presented as a web application and Android Application for online EC administrative purpose.

In proposed system, an attempt will be taken to block the registration of agricultural land for commercial purpose, approval of plan without following proper rules, getting loans, construction of building on the dried water bodies, river bank area and agricultural lands, getting new electricity (Electricity Board) connection, getting new water supply connection for the buildings on the dried water bodies, river bank area and agricultural lands by using rules and regulations as mentioned on Section of Land Act 22A. Getting loans from the banks for buying and construction of buildings on this kind of agricultural lands, dried water bodies will also be blocked.

Layouts and plans for apartments and plots without following proper rules as mentioned in the Section of Land Act 22A will not be approved by DTCP and CMDA. So, we can also avoid damages caused to people and their belongings by the buildings constructed without proper plan approval at the time of Earthquake.

Brokers can be avoided by using this system. Corruption will also be minimized. Investment of black money on the lands can also be avoided. This system will also avoid the waste of money on the demolition of buildings constructed on those lands.

Nowadays, many peoples are using the internet. So, it is very easy for them to use this E-Governance website for the land registration. Now, most of the people are using the Android Smart Phones. So, the users can be very easy for them to use this E-Governance System for Land Registration through their Android Devices.

4.1 Modules

There are three modules in this Proposed System. There are

1) Module - 1: Data Mining
2) Module - 2: Developing Web Application
3) Module - 3: Developing Android Application

4.2 Module - 1: Data Mining

4.2.1 Architecture

Data Extraction will be used to pull the tweets from R code using Oauth facility by Submitting Twitter credentials. Similarly, Facebook Posts will be pulled from R code using FBoauth facility by Submitting Facebook credentials. Corpus Cleaning will be used to clean and present the extracted data to R engine. Lexical Analyzer analyzes the Extracted data as positive or negative data separately by matching it with words on the positive-words.txt and negative-words.txt files. Machine learning will take place to project the result. Bayesian Learning will happen. The end result will be projected Sentiment or Opinion projection, polarity analysis (trend towards positive or negative trait), word cloud most frequent words spoken. The Architecture of the Data Mining Module is as shown in Fig. 2.

© 2017, www.IJARIIT.com All Rights Reserved

Page 408
4.2.2 Requirements
1) R Version 2.15.3 with packages like ggplot2, WordCloud, SnowballC, tm, rook, Rstem, facebook, twitteR, sentiment, NLP, Topic models, RTTextTools, e1071, bit64 with all of its dependency packages.
2) R Studio Version 0.99.903.

4.2.3 Hash Tags
Facebook HashTags used are as follows
- Chennai rains
- ChennaiRainRelief2015
- SOSShennaiRains

Totally, 3303 Posts retrieved from Facebook. But After Cleaning, 387 Facebook posts are only used for analysis. Because these 387 posts are only posted in the month of November and December 2015.

Twitter Hash Tags used are as follows
- ChennaiRains
- ChennaiRainsHelp
- ChennaiRainsOrg
- ChennaiFloods
- ChennaiUpdates
- ChennaiRain
- ChennaiMicro
- ChennaiFlood
- ChennaiVolunteer
- PrayforChennai

Totally, 1574 Tweets retrieved from Twitter.

4.2.4 Sentiment Analysis
Sentiment analysis is also known as opinion mining. Opinion mining works based on natural language processing and text analysis. Sentiment analysis is widely applied to reviews and social media for a variety of applications, ranging from marketing to customer service. Sentiment analysis aims to determine the attitude of a writer. Sentiment Analysis works based on the identified emotions such as "angry", "sad", and "happy".

Sentiment analysis for the Facebook data retrieved by using the Facebook hashtags mentioned is as shown in Fig. 9. Sentiment analysis for the Twitter data retrieved by using the Twitter hashtags mentioned is as shown in Fig. 10.

4.2.5 Polarity Analysis
Polarity analysis is classifying the polarity of a given text at the document, sentence, or feature/aspect level—whether the expressed opinion in a document, a sentence or an entity feature/aspect is positive, negative, or neutral.

Polarity analysis for the Facebook data retrieved by using the Facebook hashtags mentioned is as shown in Fig. 11. Polarity analysis for the Twitter data retrieved by using the Twitter hashtags mentioned is as shown in Fig. 12.

4.2.6 Word Cloud
Word Cloud is a visual representation of text data, typically used to depict keyword metadata or tags on websites or to visualize free form text. A Word Cloud is also called as Tag Cloud. Word Cloud is derived by using the Frequent Words generated from Document-Term Matrix (DTM). Document-Term Matrix (DTM) is also called Term Document Matrix (TDM).

Word Cloud derived from the Facebook data retrieved by using the Facebook hashtags mentioned is as shown in Fig. 13. Word Cloud derived for the Twitter data retrieved by using the Twitter hashtags mentioned is as shown in Fig. 14.

4.2.7 Machine Learning

Machine learning is a type of artificial intelligence (AI) that provides computers with the ability to learn without being explicitly programmed. Machine learning focuses on the development of computer programs that can teach themselves to grow and change when exposed to new data.

4.2.8 Supervised Learning

Supervised learning is the machine learning task of inferring a function from labeled training data. The training data consist of a set of training examples. In supervised learning, each example is a pair consisting of an input object (typically a vector) and the desired output value (also called the supervisory signal). A supervised learning algorithm analyzes the training data and produces an inferred function, which can be used for mapping new examples. An optimal scenario will allow for the algorithm to correctly determine the class labels for unseen instances.

4.2.9 Bayesian Learning

Bayesian Learning uses Bayesian Theorem. Bayesian means probabilistic. The specific term exists because there are two approaches to probability. Bayes Theorem provides a direct method of calculating the probability of such a hypothesis based on its prior probability, the probabilities of observing various data given the hypothesis, and the observed data itself. Bayesian Learning is used to project the Sentiment and Polarity.

\[
P(\text{A} | \text{B}) = \frac{P(\text{B} | \text{A}) \times P(\text{A})}{P(\text{B})}
\]

Where, 
- \(P(\text{A})\) = prior probability of hypothesis A
- \(P(\text{B})\) = prior probability of training data \(B\)
- \(P(\text{A} | \text{B})\) = probability of \(A\) given \(B\)
- \(P(\text{B} | \text{A})\) = probability of \(B\) given \(A\)
- \(P(\text{B} / \text{A})\) can be represented as \(P(\text{B} / \text{A}) = P(\text{B} \cap \text{A}) / P(\text{A})\)

Maximum Entropy

The Max Entropy classifier is a probabilistic classifier which belongs to the class of exponential models. Max Entropy is shortly called MaxEnt. MaxEnt does not assume that the features are conditionally independent of each other. The MaxEnt is based on the Principle of Maximum Entropy and from all the models that fit our training data, selects the one which has the largest entropy. The Max Entropy classifier can be used to solve a large variety of text classification problems for sentiment analysis. MaxEnt uses the Document-Term Matrix (DTM) to find the Accuracy of this system.

4.2.11 Classifier Accuracy

The accuracy of the predictor refers to how well a given predictor can guess the value of predicted attribute for a new data. Accuracy is the percentage of the testing set examples correctly classified by the classifier.

- Classifier Accuracy of Facebook Data is 1.29%.
- Classifier Accuracy of Twitter Data is 1.42%.

4.2 Module - 2: Developing Web Application

4.3.1 Architecture

The developed Web Application will be used on the Web Browser of the Systems. All the operations performed by this Web Application will send the data to the PHP file on the Server using the POST method. The Architecture of the Web Application Module is as shown in Fig. 3.
4.3.2 Requirements
1) WAMP Server
2) Database: MySQL
3) Scripts Used: HTML, CSS, Java Script, PHP.
4) Software Required: Adobe Reader.

4.4. Module - 3: Developing Android Application
4.4.1 Architecture

The developed Android Application will be used on the Android Devices. This Android Application will add mobility to the system. All the operations performed by this Android Application will send the data to the PHP file on the Server using the POST method as JSON encoded data. The Architecture of the Android Application Module is as shown in Fig. 6.

![Architecture Diagram](image)

**Fig. 6 Module – 3 Architecture**

Android Application's Sample Screenshots are as shown in Fig. 7 and Fig. 8.

![Splash Screen](image)

**Fig. 7 Android Application Splash Screen**

![Services Screen](image)

**Fig. 8 Android Application Services Screen**

**Requirements**

*a. The For Developing PC*

1) Java Development Kit ( jdk1.6 or above).

2) Java Runtime Environment ( jre1.6 or above).

3) Eclipse IDE (Integrated Development Environment) with SDK (Software Development Kit).

*b. For Mobile*

1) Operating System: Android 3.2 (Honeycomb) or above.

2) Apps Required: Browser, PDF Reader.
RESULTS AND DISCUSSIONS

5.1 Sentiment Analysis
Sentiment analysis for the Facebook data retrieved by using the Facebook hashtags mentioned is as shown in Fig. 9. Sentiment analysis for the Twitter data retrieved by using the Twitter hashtags mentioned is as shown in Fig. 10.

Sentiment analysis for the Facebook data shows the Sadness of the people about the flood in Chennai. Sentiment analysis for the Twitter data shows the Fear of people about the flood in Chennai.

5.2 Polarity Analysis
Polarity analysis for the Facebook data retrieved by using the Facebook hashtags mentioned is as shown in Fig. 11. Polarity analysis for the Twitter data retrieved by using the Twitter hashtags mentioned is as shown in Fig. 12.

Polarity analysis for the Facebook data shows the Negative posts and comments of people about the flood in Chennai. Polarity analysis for the Twitter data shows the Negative tweets and comments of people about the flood in Chennai.

5.3 Word Cloud
Word Cloud generated for the Facebook data retrieved by using the Facebook hashtags mentioned is as shown in Fig. 13 and Word Cloud generated for the Twitter data retrieved by using the Twitter hashtags mentioned is as shown in Fig. 14.

Word Cloud for the Facebook data shows that the most frequently used word is ―rain‖ and Word Cloud for the Twitter data show that the most frequently used word is ―bad‖.

CONCLUSION
Data about tweets from Twitter can get only for the 60 days (2 Months). Facebook data is very difficult to mine between ranges of dates. Spelling mistake on Facebook and Twitter by the users may cause errors on the Sentiment and Polarity Analysis. More words to be updated on the positive-words.txt and negative-words.txt frequently to get accurate sentiment and polarity.
Based on the Sentiment Analysis, Polarity Analysis and Word Cloud generated by using the data pulled from Social Medias like Facebook and Twitter the people felt very sad and fear about the flood caused by rain due to the existing system’s drawbacks. I considered each and every data extracted from Facebook and Twitter with the Government rules under the Section of Land Act 22A and developed the best Website and Android Application to solve the found issues.

REFERENCES


[7] Jiawei Han, Micheline Kamber, Jian Pei, —Data Cleaning, Bayes Classification Methods in Data Mining Concepts and Techniques, 2012, pp. 88-92 and 350-354.


© 2017, www.IJARIIT.com All Rights Reserved
BIOGRAPHY

A. Karthikeyan has born in Kalanivasal, Karaikudi, Sivaganga District, Tamil Nadu, India. He has completed his schoolings in Karaikudi. He has received the Bachelor of Engineering Degree on Computer Science and Engineering from the Mount Zion College of Engineering and Technology, Pudukkottai under the affiliation of Anna University, Chennai. He is currently pursuing the Master of Engineering Degree on Computer Science and Engineering from the Mount Zion College of Engineering and Technology, Pudukkottai under the affiliation of Anna University, Chennai.