

AI-Driven Predictive Analytics for Citizen Complaint Trends in Smart Cities

¹Dr.M.Mayuranathan, ²A.Aruneswaran, ³M.Dharshini, ⁴ N.G.Boseraj

¹Associate Professor, ²Student, ³Student, ⁴Student

¹Computer Science and Engineering,

¹SRM Valliammai Engineering College, Chengalpattu, India

Abstract : By creating an AI-Driven Predictive Analytics System for Citizen Complaint Trend Analysis within Smart Cities, this research aims to advance urban governance and service delivery mechanisms. Complaints made by residents regarding sanitation issues, damaged roads, water supply problems, or inadequate street lighting can be analyzed for patterns and future predictions through the application of different machine learning and data analytic techniques. A web-based portal collects citizen complaints, which are organized, documented, and will ultimately be entered into a structured database format for entry into various predictive model applications. These models are intended to identify historic and regionally-based trends, thus enabling city officials to develop their resource allocation strategies and establish preventative measures for future incidents. The Predictive Analytic System can improve decision-making by assisting authorities with determining how best to serve the citizens of their respective communities. Furthermore, Smart City implementation may occur with higher levels of efficiency when coupled with the advantages found within an AI-Driven Predictive Analytics System, thus creating opportunities for city authorities to confront and resolve issues before they escalate or present complications through proactive means.

IndexTerms - Smart Cities, Citizen Complaint Analysis, Predictive Analytics, Machine Learning, Natural Language Processing (NLP), Data Visualization, Urban Service Management, Artificial Intelligence.

I. INTRODUCTION

INTRODUCTION

Smart Cities leverage advanced technology and efficient governance with the goal of enhancing the quality of life for their citizens. One of the challenges for city leadership is the volume of citizen complaints related to public services, such as sanitation, street maintenance, water distribution, street lighting and waste removal. Often, citizens submit complaints through online portals, mobile apps or help centers. There is an urgent need to effectively analyze the complaints so that recurring problems can be identified, and service delivery can be enhanced.

Predictive Analytics involves using statistical algorithms, machine learning techniques and the analysis of historical data to identify trends and predict future events. In the context of Smart Cities, predictive analytics helps government agencies to analyze the historical data on citizen complaints and predict trends for future complaints. As such, city leaders can take proactive measures prior to an issue becoming larger. Technologies that utilize Artificial Intelligence (AI) and Data Analytics are instrumental in converting large amounts of citizen complaint data into meaningful insights.

NEED OF THE STUDY.

Urban areas are experiencing rapid population growth, which creates significant challenges in delivering efficient public services. Citizens frequently report complaints related to infrastructure, sanitation, and public utilities. Traditional complaint management systems mainly focus on recording and resolving individual complaints, but they often lack the ability to analyze complaint trends or predict future issues. As a result, these systems provide limited support for understanding recurring problems across different areas of the city.

Due to the absence of predictive analytics in existing systems, there are often delays in responding to complaints and planning resource allocation. City administrators may struggle to identify areas with a high frequency of issues or anticipate future service failures. An AI-Driven Predictive Analytics System addresses these limitations by analyzing historical complaint data and identifying common patterns. By predicting future complaint trends and highlighting problem-prone areas, the system helps city authorities improve decision-making, enhance service management, and increase overall citizen satisfaction.

3.1 Population and Sample

In a city designed to be smart, all citizens have the ability to file thousands of complaints daily using online portals, mobile apps, and help desk systems relating to public service issues such as sanitation, road damage, water supply, rubbish collection, and street lighting. The population for analysis within this study includes all citizen complaints entered into the smart city complaint management system.

The analysis uses a sample data set of complaints entered into the complaint management system from the various districts of the city for analysis purposes. The sample dataset includes only those citizen complaints entered via the web-based portal during the selected time period. The sample includes complaints by category of complaint type, category of complaint location within the city, category of date filed, and category of complaint status. These complaints are used for analysis purposes to help identify patterns and trends of the citizens' issues in the various areas of the city where the citizens reside.

3.2 Data and Sources of Data

For this research project, we will utilize complaints collected through a digital complaints management platform (the "Platform"). Examples of this would include, but not limited to, using the following data points to capture actual complaint data such as: complaint categories; description/ type of complaint; date/ time of submission; and resolution status, etc.; all of which help provide a clear understanding of all complaints received over a designated period of time.

The database is a structured database and is utilized for analysis / prediction, machine learning models will be used to look for historical patterns/ trends. Additional information such as service categories and/or location-based information assist in understanding where there are multiple complaints that have come from a particular area within Oakland and their relationship to each other, thus providing an accurate picture of frequency and distribution of where these complaints occurred by geographical area.

3.3 Theoretical framework

The theoretical framework of this research includes the dependent & independent variables utilized in predictive analysis. The dependent variable is the amount or frequency of citizen complaints showing the occurrence of a particular issue in a diverse category of public service (see public sectors). The independent variables that are included in the research are; Complaint Type, Location of Complaint, Time of Complaint, and Responsibility of the Service Department for Resolving an Issue. The relationship between these variables is established by using Machine Learning Algorithms to recognize patterns and establish future predictive trends. Using these independent variable & dependent variable combinations can give city decision-makers the necessary information for taking preventive actions and improving service delivery in Smart Cities.

RESEARCH METHODOLOGY

The section on methodology of this research includes a description of the methodological approach and methods employed throughout the course of conducting this research. The specific elements of each methodology include: (1) the manner in which data collected was processed and analyzed, (2) data collection methods, (3) sample used, (4) variables included in the analysis, and (5) framework used to conduct analysis. The following describes each element of the research methodologies.

3.1 Population and Sample

Many complaints are filed each day by citizens about public services. These complaints are made regarding issues such as road maintenance, sanitation, street lights, waste management, and water supply. Most citizens utilize digital complaint portals, mobile applications, or municipal service portals to file these complaints. The total number of citizen-generated complaints entered into the smart city complaint management system provides a complete dataset of complaints for the study.

The entire complaint dataset is sampled and produces a sample dataset of records that have been collected over the time frame of one year. Each complaint within the sample dataset includes information about the category of the complaint, where it originated from (geographic area), the date it was submitted, and its current status. Approximately 5,000 records, representing a year's worth of complaints, are being used as the sample dataset for data analysis.

The selected sample contains records from many different categories of complaint and from multiple geographic locations throughout the city; this will allow the researcher to identify and analyze trends relating to citizen issues within a smart city and as such are the source of data from which all subsequent research will be conducted.

3.2 Data and Sources of Data

The researcher used secondary data for the research conducted in this study. The complaint data consists of public municipalities, municipal complaint management systems, and simulated datasets intended for urban analytics research. The researcher collected the following attribute information about each of the complaints.

1. ComplaintId
2. ComplaintCategory (sanitation, roads damaged by heavy equipment, water, street lights etc.)
3. Complaint description
4. Complaint location (i.e., area, zone)
5. Date and time when complaint was submitted
6. Complaint Status - whether or not a given complaint has been resolved

This dataset consisting of complaints covers the period of January 2022-December 2024. The processed complaint dataset was stored in a MySQL database and analyzed with Python (locally). Data preprocessing methods (cleaning, filtering, and normalization) were applied to the database prior to the analysis.

After preprocessing, the complaint dataset was entered into a machine learning algorithm in order to identify trends, and to be able to predict the future trends that will occur with respect to complaints.

3.3 Theoretical framework

This research study has developed a theoretical framework with dependent and independent variables that will be utilized for analyzing and forecasting patterns of citizen complaints in smart cities. The dependent variable of this study is the amount or frequency of citizen complaints created over time; this measurement will represent the level of problems as reported by citizens to the public sector. The independent variables explored in this research study will include the type of complaint (e.g., sanitation, water supply, road damage, street light), the location/zone where the complaint was reported, when (day, week, or month) the complaint was reported, and the government agency responsible for addressing the complaint.

3.4 Statistical tools and econometric models

A thorough overview of the statistical instrument and method employed to examine the complaint data and make predictions, will be given.

3.4.1 Statistics Description

Statistical descriptions are used to summarize or summarize and analyze the complaint data set. For the questionnaires utilized in this research, the minimum, maximum, average, and/or standard deviation statistics for the complained variable were calculated. Statistical descriptions are helpful in determining where most of the complaints are distributed among the various locations and types. Through the use of statistical descriptions, researchers can see which types of complaints are occurring the greatest frequency in the study area and where the highest number of complaint incidents occurs. Researchers will also perform normal tests on the complaint dataset to see if it meets the criteria for a normal distribution. Normal tests provide researchers with a way to see the degree of variance and the periodic change of complaint patterns for their data set.

3.4.2 Machine Learning Prediction Model

Historical data is used to apply machine learning algorithms to forecast the likelihood of future citizen complaints from historical data on previous complaints. Using the historical complaint data allows the model to find incidents that are consistent with previous patterns of associated variables and can help to develop a predictive model based on the relationship of different types of complaints.

The algorithms used to train the predictive model are the following supervised learning algorithms: Support Vector Machine (SVM), Naïve Bayes and Linear Regression. These algorithms learn by creating trained models that find predictive relationships among the different variables in terms of how they are related to the occurrence of complaints.

The predictive model is expressed in the following form:

$$\text{Compliant Trend} = \beta_0 + \beta_1(\text{Category}) + \beta_2(\text{Location}) + \beta_3(\text{Time}) + \varepsilon$$

Where,

β_0 is the model's y-intercept

β_1 , β_2 , β_3 represent how much influence each variable has on complaint predictions

ε is the error term

Once the models are created using one or more of the above algorithms, the models can predict the likelihood of a complaint being made in a certain geographic location and by service category.

3.4.3 Natural Language Processing for Complaint Classification

Citizen complaints with text descriptions are commonly analyzed and classified through natural language processing (NLP). The text data is converted into numerical features using term frequency-inverse document frequency (TF-IDF), which can then be used to assign complaints to specific categories (such as sanitation, roads, water supply, and street light) and provide city officials effect-feedback on which service category the complaint belongs to. By using automated classification of this type, city officials can quickly and easily categorize a complaint by the appropriate department.

3.4.4 Data Visualization and Trend Analysis

Complaint trends are represented with the help of visualizations such as line graphs and bar/heat charts which allow administrators to determine where high levels of complaints exist and what types of complaints are common. In addition to identifying the types of complaints, trend analysis can also identify recurring themes or patterns in citizen complaints on a seasonal basis to provide the necessary data to allow authorities to take pre-emptive measures.

IV. RESULTS AND DISCUSSION

4.1 Results of Descriptive Statics of Study Variables

Table 4.1: Descriptive Statics

Variable	Minimum	Maximum	Mean	Std. Deviation
Toatl Complaints	5	120	45	20
Sanitation Complaints	3	60	22	12
RoadDamage Complaints	2	55	18	10
Water Supply Complaints	1	40	14	9
Street Light Complaints	1	30	10	7

Table 4.1 Summary of statistics describing the complaints associated with each of the Smart City complaint categories will be reviewed. Exhibiting mean, maximum, minimum, and standard deviation of citizen complaints using the Smart City complaint management system allows for a summary understanding of the overall complaint distribution by complaint type. Mean value of a complaint category within the period of time that the study was conducted, provides the average number of complaints received for each complaint type.

The maximum value in each complaint category corresponds to the number of complaints received for that complaint type during the study but indicates the maximum number of complaints received for each complaint category during the time of the study. Minimum value in each complaint category indicates the amount of complaints received relative to what was received in that complaint category.

Standard deviation of each complaint type indicates that citizen complaints received for that complaint category varied by location and time of occurrence, but also, the standard deviation also indicates that the occurrence of complaints received for each complaint category also varies. Overall, the summary statistics will describe the overall distribution and activity of citizen complaints received in the Smart City, as well as, when and where citizen complaints have been received, thus, allowing for Smart City authorities to have an understanding of citizen complaints and how to address issues associated with individual citizen complaints.

The examination of complaint characteristics shows that citizens require more attention from local authorities in relation to citizen complaints for both road and sanitation service than they do for other types of complaint service according to the results of Table 4.1 in the study documentation. Thus the need for these services to receive greater oversight and maintenance will have greater impact on urban areas compared to other services.

II. ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to the faculty members and project supervisors who provided valuable guidance and support throughout the development of this research work. Their suggestions and technical insights greatly helped in improving the quality of the project. The authors also thank the institution for providing the necessary resources and facilities required to complete this study successfully. Appreciation is extended to all individuals who contributed directly or indirectly to the completion of this research on AI-driven predictive analytics for smart city complaint management.

REFERENCES

- [1] Batty, M., Axhausen, K., Giannotti, F., Pozdnoukhov, A., Bazzani, A., Wachowicz, M., Ouzounis, G., & Portugali, Y. (2012). Smart cities of the future. *European Physical Journal Special Topics*, 214(1), 481–518.
- [2] Hashem, I., Chang, V., Anuar, N., Adewole, K., Yaqoob, I., Gani, A., Ahmed, E., & Chiroma, H. (2016). The role of big data in smart city development. *International Journal of Information Management*, 36(5), 748–758.
- [3] Albino, V., Berardi, U., & Dangelico, R. (2015). Smart cities: Definitions, dimensions, performance, and initiatives. *Journal of Urban Technology*, 22(1), 3–21.