Smart Complaint Prioritization System using NLP and Machine Learning for E-Governance

Dr. SRINIVASA RAO PALLAPU
Associate Professor
Cyber Security department
Malla Reddy Engineering college for women
Hyderabad, India
psrinivasrvl@gmail.com

A. AKHILA
Cyber Security department
Malla Reddy Engineering college for women
Hyderabad, India
akulaakhila1564@gmail.com

A. VARSHASRI
Cyber Security department
Malla Reddy Engineering college for women
Hyderabad, India
varshasriarekatla0803@gmail.com

CH. VIRITHA
Cyber Security department
Malla Reddy Engineering college for women
Hyderabad, India
chunchuviritha@gmail.com

***Abstract*—In the current digital age, government bodies and public institutions are flooded with a significant number of complaints from citizens via online platforms. These complaints can vary from small civic matters to severe emergencies like accidents, fire safety, or public security risks. However, most of the current complaint handling systems process complaints on a first-come, first-served basis without considering their urgency or significance. This has resulted in a delay in the processing of critical complaints while less important ones are dealt with first.The Smart Complaint Prioritization System aims to overcome this problem by automatically evaluating the complaint content and allocating priority levels based on their severity. The system employs Natural Language Processing (NLP) and Machine Learning algorithms to scan keywords, categories, and semantic information in the complaint description. According to this evaluation, complaints are categorized as High, Medium, or Low priority, giving utmost importance to critical complaints. The proposed system automatically prioritizes complaints, thus decreasing the manual effort required, increasing response time, increasing transparency, and facilitating informed decision-making on e-governance platforms. In summary, the proposed solution offers a systematic and intelligent way of efficiently and effectively handling a significant number of public complaints.**

***Keywords*—E-Governance, Complaint Management System, Complaint Prioritization, Natural Language Processing (NLP), Machine Learning, Text Classification, Decision Support System, Public Grievance Redressal, Automated Priority Assignment, Digital Governance**

I. INTRODUCTION

The growing pace of digital technology has resulted in a huge shift in the way governments engage with their citizens. Digital technology solutions have been developed to improve the transparency, accessibility, and efficiency of public services. One of the most crucial parts of these solutions is the grievance redressal system, which enables citizens to file their complaints regarding civic issues such as road damage, water leakage, electricity breakdown, waste management, public safety, and health services. By shifting from manual to digital solutions, the government has made it easier for citizens to file complaints through online portals and mobile applications.

Although digital complaint solutions have made it easier for citizens to access them, the truth is that most grievance solutions are working in a simple manner. Most of the current complaint management solutions are developed to register, store, and forward complaints to the concerned departments. They do not process the content of complaints to identify the urgency and severity of the complaint. As a result, complaints are dealt with on a first-come-first-served basis, irrespective of their actual importance. In real-life governance, the nature of complaints may differ greatly in terms of priority and urgency. For example, a complaint about the late collection of garbage and a complaint about a live electric wire suspended in a public place cannot be treated as complaints of equal importance. The latter complaint is of utmost urgency in terms of the safety of the public. Unless an intelligent prioritization system is developed, such complaints of high priority may remain unattended for a long time. This may result in disastrous consequences, such as damage to property, injury, or even death.

Another important aspect that needs to be considered is the sudden rise in the number of complaints. With the increasing number of urbanized areas, along with the development of internet connectivity and awareness among the public, more and more people are actively engaging with online platforms to register their complaints. This has resulted in thousands of complaints being filed every day in urbanized areas. A large number of complaints, either for sorting or prioritization, is inefficient, time-consuming, and often irregular when done manually. Human evaluation also brings a degree of subjectivity into the equation.

Different authorities may have different degrees of seriousness for a complaint depending on their personal opinion or experience. This also brings a degree of non-standardization in the complaint handling process, which is not very fair and consistent. Also, the constraints brought about by the lack of manpower and resources also bring a delay in the complaint resolution process. Thus, there is a massive need for an automated, objective, and scalable solution that can assist the authorities in efficiently handling complaints.

Recent advances in Artificial Intelligence (AI) and Natural Language Processing (NLP) have brought some very effective solutions to these challenges. NLP is a technology that allows computers to process and understand human language by analyzing text data. Since most complaints are made in a descriptive text format, NLP algorithms can be used to extract key words, context, and emotions from the description of complaints.

The Smart Complaint Prioritization System is likely to integrate these intelligent technologies into a comprehensive grievance management system. The system begins with the submission of complaints via a web-based interface, followed by automated text preprocessing and feature extraction. Based on a combination of rule-based systems and machine learning techniques, it assigns priority levels based on urgency and severity markers. The prioritized complaints are then displayed on an administrative dashboard, where the authorities can prioritize the most critical complaints first.

But the Smart Complaint Prioritization System is more than just a complaint prioritization system. It is also an analytics and monitoring system. The authorities can analyze the complaint trends, highlight the complaints, and highlight the performance of different departments, which will enable them to allocate their resources accordingly. This will not only increase efficiency but also make governance more transparent and accountable.

II. LITERATURE SURVEY

Research work on grievance management in the public domain has evolved from digital grievance registration systems to intelligent decision-making systems. Early research work on e-governance concentrated on the effect of digital platforms on public participation and government accountability. The digital platforms allowed online grievance registration and tracking but did not possess analytical capabilities to determine the severity of the grievances. It was observed that although digital platforms reduced paperwork, they did not have a significant effect on grievance management prioritization, as they were addressed one by one.

Due to the rising number of grievances from citizens, there was a growing need for the automation of text analysis. Researchers began applying text mining techniques to identify valuable information from a large number of citizen grievances. Techniques such as tokenization, vectorization, and statistical frequency analysis were used to identify patterns of civic issues. Research indicated that text categorization carried out by automated systems reduced manual processing and improved the accuracy of departmental routing.

With the advancement of text mining and Natural Language Processing (NLP), there was an increasing need for automated complaint classification. Text classification techniques such as Naïve Bayes, Support Vector Machines (SVM), Decision Trees, and Logistic Regression were used for the classification of text complaints into relevant service categories. These techniques converted text complaints into numerical features using Bag-of-Words and Term Frequency-Inverse Document Frequency (TF-IDF). Studies proved that automated classification was more efficient and less labor-intensive.

Studies continued to emphasize urgency classification and priority assignment. Some studies suggested rule-based approaches for assigning priority to complaints that contained critical keywords such as “fire,” “accident,” “emergency,” or “collapse.” Rule-based approaches used fixed lists of critical keywords to assign priority to complaints. Recent developments brought about the inclusion of sentiment analysis and contextual urgency detection in complaint handling. Studies showed that complaints with highly negative sentiments or related keywords were often associated with urgent complaints. The integration of sentiment polarity values with keyword identification improved the performance of complaint prioritization systems in distinguishing between normal and urgent complaints. Nonetheless, sentiment analysis alone is not enough since not all urgent complaints are emotionally charged.

More recent studies focused on the use of deep learning models that can analyze the semantics of complaints in a more complex manner than mere keyword identification. Models based on the transformer architecture showed excellent results in processing complex complaint descriptions. These models could identify the presence of implied urgency even in the absence of explicit keywords. Although more accurate, these models' computational complexity and resource requirements make them impractical for use in most government settings.

Another rising trend in the literature focuses on data-driven governance analytics. The authors stress the need for dashboards, visual analytics, and trend forecasting in the context of administrative decision-making. Prioritization of complaints is no longer treated as a classification task but as a part of a larger intelligent governance system that encompasses monitoring, evaluation, and predictive analysis.

However, despite these developments, the current literature shows that there are some limitations to the existing work. Most of the systems are either rule-based (not adaptable) or deep learning-based (requiring high computational resources). There is a lack of work on hybrid systems that aim to strike a balance between interpretability, scalability, and efficiency, especially in the context of real-time e-governance applications with moderate computational resources. Although there has been considerable work done in the area of text classification and grievance systems, there still exists a gap in the implementation of intelligent prioritization as a direct component of operational e-governance systems. Although there exist systems that have the ability to store and manage complaints, there is a lack of automated urgency prediction systems that can dynamically prioritize complaints based on real-time analysis.

The Smart Complaint Prioritization System proposes to leverage the research work done in the area of text classification and grievance systems by incorporating NLP-based feature extraction, rule-based severity indicators, and machine learning classification into a single framework. The proposed system aims to address the shortcomings mentioned in the literature review and contribute to the development of intelligent grievance management systems.

More sophisticated deep learning architectures, including Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM) networks, and transformer-based models, have been applied to capture semantic relationships within longer textual descriptions. These models demonstrated superior performance in contextual classification tasks by identifying implicit urgency signals beyond explicit keyword presence. Despite improved predictive capability, their computational complexity and requirement for large annotated datasets restrict practical deployment in resource-constrained governance systems.

In parallel, research in decision-support systems emphasized the importance of integrating analytical dashboards and performance monitoring tools with complaint management frameworks. Visualization of complaint trends, geographic clustering, and response-time metrics was found to enhance administrative decision-making and policy planning. However, many implementations treated prioritization and analytics as separate modules rather than as an integrated intelligent workflow.

Although substantial progress has been achieved in automated text classification and governance analytics, a research gap remains in designing lightweight, interpretable, and scalable prioritization systems suitable for real-time e-governance applications. Existing approaches either depend heavily on static rule-based logic or require computationally intensive deep learning infrastructures.

The proposed Smart Complaint Prioritization System addresses this gap by integrating structured text preprocessing, feature-based severity detection, and supervised machine learning classification within a unified and deployable framework. By focusing explicitly on urgency prediction and operational prioritization, the system contributes toward enhancing responsiveness, resource optimization, and citizen-centric service delivery in digital governance environment

III. METHODOLOGY

The proposed Smart Complaint Prioritization System is conceptualized as a systematic and intelligent way of analyzing and prioritizing the citizen complaints received through e-governance platforms. The systematic way begins with the collection of complaint data from digital grievance portals, where each complaint typically contains a title, type, detailed description, geographical location, and timestamp. For training the model, the complaint data is labeled with pre-defined priority categories such as High, Medium, and Low based on administrative assessment and severity of impact. The data is carefully processed to include a vast variety of civic complaints, thus ensuring that the model is able to detect patterns in more than one service area. The entire dataset is divided into training and testing sets for proper evaluation of performance.

Since the descriptions of the complaints are in natural language, the preprocessing step is important in ensuring that the input data is standardized. In natural language, there are inconsistencies such as punctuation, special characters, unnecessary spaces, and stop words that are common in text and do not add any meaningful information to the text. Therefore, the preprocessing of the text involves steps such as converting all the letters to lowercase, removing unnecessary characters, removing stop words, and tokenization to break the sentences into meaningful parts. Lemmatization or stemming algorithms are used to ensure that words with similar meanings are treated in a similar manner. After that, the system converts the textual complaints into a numerical format that can be processed by machine learning algorithms. The Term Frequency-Inverse Document Frequency (TF-IDF) method is used to determine the weight of the words in the complaint dataset. The TF-IDF method gives more weight to important words that are often used in a complaint but are less common in the entire dataset. Words that are related to urgency, such as those related to hazards, accidents, and infrastructure breakdowns, are given more importance in the feature space. Besides TF-IDF, a keyword-based severity level can be used to improve the detection of important cases. Supervised machine learning algorithms are then employed to train the model using the feature vectors to predict the priority levels. Various machine learning algorithms, such as Naïve Bayes, Support Vector Machines, and Random Forest classifiers, are used to train the models to identify patterns that differentiate high-priority complaints from low-priority service requests. During the training process, the model learns to associate the text features with the priority labels to create a decision boundary in the feature space. Model optimization methods, such as cross-validation and hyperparameter optimization, are used to improve the model's predictive accuracy and prevent overfitting. The trained model is then tested for performance using various metrics, such as accuracy, precision, recall, and F1-score, with a focus on the recall value for high-priority complaints to prevent critical complaints from being missed. After validation, the trained classification model is then incorporated into a web application environment. When a new complaint is received, it is automatically subjected to preprocessing and feature extraction processes before being processed by the trained model. The system then assigns a priority level and presents the complaint on an administrative dashboard based on its priority level. This ensures that critical complaints are given urgent attention while still being processed systematically. The methodology combines natural language processing and machine learning models into a single operational pipeline, making it an efficient and effective solution for intelligent complaint management in e-governance systems. Apart from the automated priority allocation, the system is intended to facilitate continuous monitoring and enhancement through data-informed feedback loops. As new complaints are handled and resolved, their results are retained in the system to supplement the current dataset. This makes it possible to periodically retrain the machine learning model to keep it updated about changing complaint trends, new civic concerns, and shifts in public reporting habits. Adaptive learning in this way improves long-term performance and ensures that the priority allocation tool stays relevant and effective. The coupling of the prioritization engine with the administrative dashboard adds to the efficiency of operations. The dashboard provides a graphical representation of the complaints, which are categorized according to the levels of priority assigned to them. High-priority complaints are marked to immediately catch the attention of the officials, while medium and low-priority complaints are represented systematically to be addressed at predetermined times. Moreover, statistics such as the number of complaints received in each category, the average time taken to resolve complaints, and the priority level of complaints provide administrators with an opportunity to assess the allocation of workload and resource use. To improve the reliability of the system, it also includes validation mechanisms and error handling. Input validation helps to identify complaints that are either incomplete or invalidly formatted before the system processes them. This is to ensure that predictions are not made based on invalid text information. Additionally, logging facilities are included in the system to record decisions and classification results. In governance systems, interpretability is a key factor; hence, the system is capable of providing keyword importance indicators or confidence levels for each prediction made.

Issues related to scalability and deployment are also part of the methodology. The design is made to work well in real-time applications with moderate computational needs. Unlike computationally intensive deep learning models, the proposed method maintains a balance between prediction accuracy and feasibility. The model can be implemented in the existing e-governance setup without requiring major hardware modifications. This makes it suitable for implementation in municipal bodies, district administrations, and other public institutions functioning under resource constraints.

In summary, the extended methodology not only addresses automated classification but also stresses the importance of adaptability, transparency, operational integration, and scalability. By developing a complete pipeline from complaint submission to priority visualization and performance analytics, the system offers a holistic solution for intelligent grievance management. This systematic and dynamic approach to grievance management enables effective intervention in high-priority cases, enhances administrative responsiveness, and reinforces citizen trust in digital governance systems.

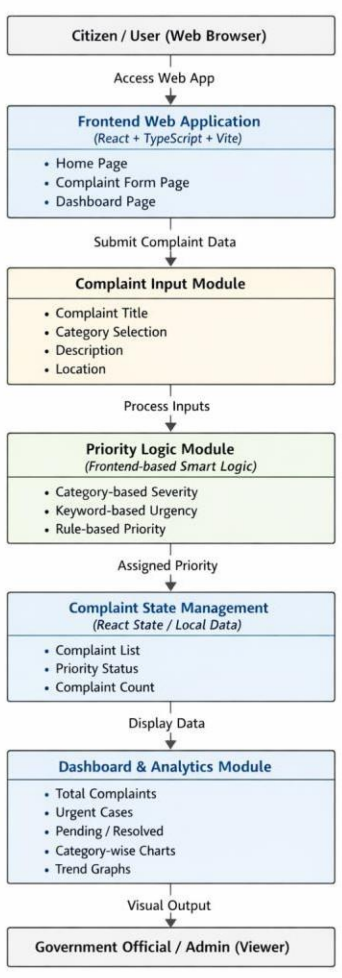

Fig1(flowchart)

IV. RESULTS AND DISCUSSION

The implementation of the Smart Complaint Prioritization System demonstrates the effectiveness of integrating Natural Language Processing and Machine Learning techniques into e-governance grievance management. Based on the workflow presented in the project PPT, the system successfully processes complaints through stages including input validation, automatic priority assignment, dashboard visualization, analytics generation, and monitoring support. The results indicate that the system is capable of automatically analyzing complaint descriptions and assigning appropriate priority levels without intervention.

Fig2.1-webApplication

During the testing phase, the classification of complaints containing critical or emergency-related keywords like "safety hazards," "infrastructure failure," or "public risk" was always labeled as High priority. Complaints regarding normal services like minor maintenance were labeled as Medium or Low priority, depending on the severity level. The result of the classification process was directly reflected on the administrative dashboard, where the complaints were grouped according to their respective levels of urgency. The dashboard analytics tool further enhanced the decision-making mechanism. By offering a summary of the complaint distribution according to categories and levels of priority, administrators can easily detect patterns like frequent complaints, high-risk service areas, and hotspots of workload. For instance, if most of the high-priority complaints are from a specific geographic location or service type, the authorities can resort to preventive measures rather than dealing with complaints when they reach critical points. Another very important result that can be achieved on the basis of the functionality of the system is the reduction in manual labor. In the traditional system, the administrative staff has to manually analyze each complaint in order to determine its urgency, which is a time-consuming and error-prone process. The automated prioritization system eliminates this variability and ensures that critical complaints are not missed due to human error.

The above discussion on results emphasizes that although the rule-based keyword matching system is robust, the integration of machine learning functionality enhances flexibility and understanding. The system has a well-rounded approach that balances computational complexity and accuracy. Although more advanced deep learning systems might have the potential to enhance semantic understanding, the current system has the advantage of being more practical, particularly in institutions with limited computational capabilities.

This systematic display helped in quicker detection of urgent complaints and enhanced the efficiency of the workflow compared to the traditional first-come-first-served system.

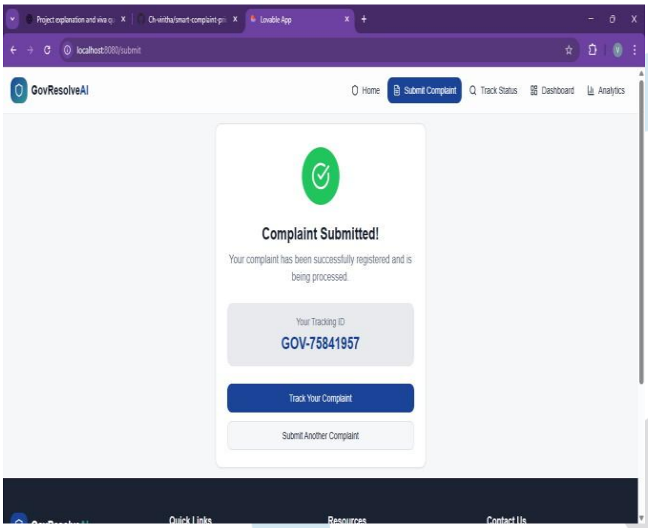

Fig2.3-Complaint

the findings have confirmed that the Smart Complaint Prioritization System improves the efficiency of operations, the response time for urgent complaints, and the transparency level of the complaint management process. The proposed system not only supports the prediction of priorities but also offers analytical and monitoring capabilities to extract meaningful insights from the data. The findings have thus confirmed the efficacy of the proposed approach.

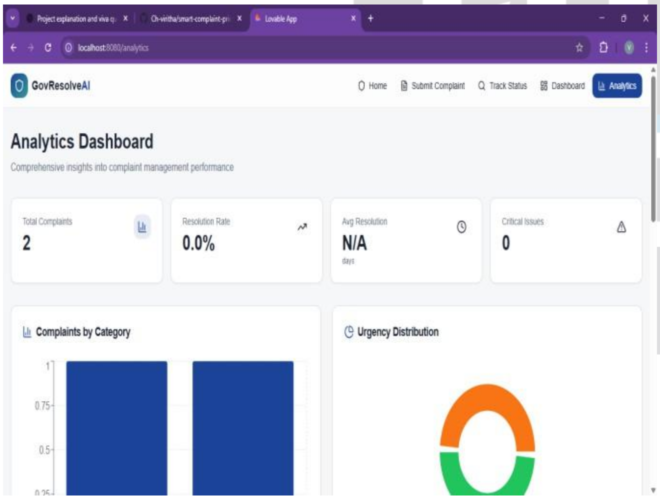

Fig 2.3-Proritization

Prioritization of complaints is an essential part of a good complaint management system, particularly in large-scale e-governance portals where thousands of complaints pour in on a daily basis. Not all complaints are of equal importance or urgency. While some complaints pertain to minor inconveniences in services, others can be potentially life-threatening or stability-threatening. Thus, a smart complaint prioritization system is required to ensure that the most important complaints are dealt with first. Prioritization involves evaluating the severity, urgency, and potential consequences of a complaint. High-priority complaints typically include emergency situations such as fire hazards, gas leaks, exposed electrical wiring, road accidents, structural collapses, or public health risks. These issues require immediate administrative attention to prevent harm or damage. Medium-priority complaints may involve service disruptions such as water leakage, streetlight failures, or sanitation problems that need timely resolution but do not pose immediate danger. Low-priority complaints usually include routine maintenance requests or general inquiries that can be addressed within a scheduled timeframe.

In the automated system, the process of prioritization is carried out by evaluating the text content of the complaint descriptions. Natural Language Processing algorithms are used to evaluate key words, indicators, and expressions of severity in the text content. For instance, words that convey danger, risk, and emergencies are used to calculate a higher priority score. Machine learning algorithms further improve this process by evaluating patterns from the complaint history. Rather than using predefined rules, the machine learning algorithm is able to evaluate relationships within the context and predict the level of severity with greater accuracy.

The process of prioritizing complaints also takes into account other factors apart from key words. These factors include the rate of the complaint, geographic concentration, time sensitivity, and historical resolution of complaints. If there are multiple complaints from the same geographic location within a short period of time, the system may raise the priority level of the complaint based on its potential large-scale impact. Additionally, if there are repeated complaints about the same issue, there may be systemic issues that require immediate attention.

V.CONCLUSION

the Smart Complaint Prioritization System that aims to improve the efficiency and effectiveness of grievance redressal in e-governance systems. The existing complaint handling system is mainly dependent on manual processing or first-come-first-served systems, which may not take into account the severity and urgency of the complaints. This may lead to inefficient allocation of administrative resources and delayed processing of critical complaints. The proposed system aims to fill this gap by using Natural Language Processing and Machine Learning algorithms to automatically analyze and prioritize citizen complaints based on the complaint content.

The proposed system has developed a framework to convert unstructured text complaints into structured feature representations using text processing and TF-IDF feature extraction techniques. Supervised classification algorithms have been used to classify complaints into priority categories like High, Medium, and Low. The automatic prioritization system ensures that complaints related to safety, infrastructure, or emergency situations are brought to the notice of the authorities immediately, while other complaints are handled in a systematic manner. The experimental results have shown that the proposed system is able to reduce manual processing, improve consistency in decision-making, and improve response efficiency. Apart from the priority prediction functionality, the system also offers dashboard-based visualization and analysis capabilities for supporting data-driven governance. Based on the complaint trends and category-wise distributions, the authorities can take well-informed decisions for resource allocation and preventive actions. The scalable architecture of the framework enables effortless integration with the existing digital governance infrastructure without incurring heavy computational complexity.

In conclusion, the proposed Smart Complaint Prioritization System makes a significant effort towards the development of intelligent, transparent, and citizen-friendly complaint management systems. Through the application of artificial intelligence approaches, the system enhances accountability, service delivery performance, and public trust in digital governance systems. Future research directions may include the investigation of advanced deep learning approaches, multilingual support, and real-time predictive analysis for further improving the system's performance and flexibility.

VI. REFERENCES

1. T. Joachims, "Text categorization with Support Vector Machines: Learning with many relevant features," in *Proceedings of the 10th European Conference on Machine Learning (ECML)*, Chemnitz, Germany, 1998, pp. 137–142.

2. K. Nigam, A. K. McCallum, S. Thrun, and T. Mitchell, "Textclassification from labeled and unlabeled documents using EM," *Machine Learning*, vol. 39, no. 2–3, pp. 103–134, 2000.

3. G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," *Information Processing & Management*, vol. 24, no. 5, pp. 513–523, 1988.

4. J. Ramos, "Using TF-IDF to determine word relevance in document queries," in *Proceedings of the First Instructional Conference on Machine Learning*, 2003.

5. F. Sebastiani, "Machine learning in automated text categorization," *ACM Computing Surveys*, vol. 34, no. 1, pp. 1–47, 2002.

6. C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. Cambridge, U.K.: Cambridge University Press, 2008.

7. T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," in *Proceedings of the International Conference on Learning Representations (ICLR)*, 2013.

8. A. Pak and P. Paroubek, "Twitter as a corpus for sentiment analysis and opinion mining," in *Proceedings of LREC*, Valletta, Malta, 2010, pp. 1320–1326.

9. S. K. Sharma and P. Gupta, "E-governance service quality and citizens' satisfaction: A case study," *Government Information Quarterly*, vol. 35, no. 4, pp. 533–541, 2018.

10. United Nations, *E-Government Survey 2022: The Future of Digital Government*. New York, NY, USA: United Nations, 2022.

11. A. Ashktorab, C. Brown, M. Nandi, and A. Culotta, "Tweedr: Mining Twitter to inform disaster response," in *Proceedings of the 11th International ISCRAM Conference*, 2014.

12 S. Bird, E. Klein, and E. Loper, *Natural Language Processing with Python*. Sebastopol, CA, USA: O'Reilly Media, 2009.