Strategies for Crisis-Responsive Governance: Automated Anomaly Identification in Public Services

Hakan Unveren
University of Vermont
hunveren@uvm.edu

Ayesh Abu Lehyeh
University of Vermont
ayesh.abulehyeh@uvm.edu

Duygu Pamukcu
HEC Montréal
duygu.pamukcu@hec.ca

Christopher W. Zobel
Virginia Tech
czobel@vt.edu

ABSTRACT

This paper introduces a machine learning tool for service systems, focusing on accurate classification of service requests and swift anomaly detection, particularly crucial during emergencies. Employing a Support Vector Machine model, this tool automatically classifies service calls into predefined categories with high accuracy, while effectively detecting irregular requests that require specific attention from operators. This approach streamlines resource management by reducing the manual categorization workload and enables early detection of emerging service needs. Examining Orange County, Florida 311 System data, with a specific focus on the COVID-19 period, we illustrate the tool's success in automatic request categorization and anomaly detection. Overall, this tool presents an effective automation approach to help with efficient resource management of service systems and proactive assessment of public service needs, promising to revolutionize service request management during crises. Future work will explore additional classification models for enhanced accuracy and integrate automated alerts for proactive disaster management.

Keywords

311 services, classification, anomaly detection, supervised learning, disaster management.

INTRODUCTION

The evolution of information technology-enabled service systems has revolutionized the way governments interact with their constituents. Public service requests offer individuals a direct means to voice their needs and concerns, serving as a vital bridge for citizen-government interactions (Schwester et al., 2009; Wu, 2020). Citizens now have the flexibility to engage with government agencies through various mediums such as online portals, mobile applications, and chat systems. These systems facilitate the seamless collection of requests and enable prompt and effective responses, thereby fostering a more responsive and citizen-centric governance approach. Given the widespread availability and accessibility of public systems, the necessity for streamlined and automated processes becomes imperative to meet the diverse and expanding needs of the population (Hashemi, 2022). Consequently, the automation of handling service requests in public systems emerges as a crucial step toward enhancing efficiency and responsiveness in governance.

Automation in public services refers to the integration of technology and streamlined processes to carry out tasks and functions within government agencies without direct human intervention. This involves the use of software, artificial intelligence, and other technological solutions to handle routine and repetitive tasks, facilitating the analysis of patterns in service requests, anticipation of citizen needs, and optimization of resource allocation within government agencies (Bianchini et al., 2006; Brogi et al., 2006; Cheng et al., 2022; Pamukcu et al., 2022). The overarching goal is to improve the efficiency of public service delivery and enable agencies to proactively address emerging challenges. The benefits of automation extend to quicker response times, a consistent service approach, and the reduction of human errors, fostering a reliable interaction between citizens and government agencies. Moreover, as automation takes charge of routine tasks, it frees up human agents to concentrate on more complex and critical issues that demand a personalized touch.

The importance of automation becomes even more evident during disasters (Shah et al., 2019). Community members often actively contribute to reporting emerging needs during crises, and automated systems can leverage crowdsourced data to comprehensively identify and categorize new service requirements, providing a more nuanced understanding of the evolving situation (Pamukcu et al., 2021). Accurate categorization of emerging needs and efficient resource utilization becomes imperative to ensure a dynamic and responsive system, even during surges in service requests. This heightened demand for accuracy and efficiency underscores the crucial role that automation can play in effectively managing and responding to the diverse and urgent needs that arise during disaster scenarios.

The recent coronavirus pandemic highlighted critical challenges in current public service request management practices as systems faced a surge in pandemic-related calls, particularly in the early months of the emergency (Filip et al., 2022). The overwhelming volume of calls resulted in an increased demand for system resources, including service operators and personnel hours (Pamukcu et al., 2021). Manual operations, such as assigning each request into its appropriate category, consumed a substantial amount of limited system resources. Additionally, systems began to receive irregular types of requests that did not align with existing categories, posing request categorization challenges, and potentially causing delays in response times as these inquiries required more careful attention.

An illustrative example is seen in the Orange County Florida 311 non-emergency system, where similar requests were inconsistently assigned to different service categories. To address this issue, the system created new categories to address newly emerged service needs under a unified classification. Despite early signs of the pandemic, with the first two cases in Florida reported on March 1 and the declaration of a state of emergency on March 9, a dedicated category for COVID-19 was not established until May 8, indicating a notable delay in adapting to emergent situations (Cutway, 2020). Similarly, the New York City 311 system introduced new categories such as “Social Distancing” and “Mask Covering” to accommodate pandemic-related concerns, highlighting the necessity for a more adaptive categorization process within these systems (Pamukcu et al., 2021). The examples highlight the ongoing need for refined and efficient automation processes, particularly in times of crises.

Automated systems, equipped with adaptive algorithms, effectively tackle challenges arising from the extensive need for human resources in manual operations. This adaptive capability ensures the system remains dynamic and responsive to the evolving nature of service requests. For example, the literature indicates that the incorporation of Natural Language Processing (NLP) in automated systems enhances their ability to comprehend and categorize unstructured data, such as text or voice messages, from affected individuals. Additionally, numerous studies emphasize that the application of machine learning and neural network methods can provide accurate service categorization (Bianchini et al., 2006; Brogi et al., 2006; Hashemi, 2022). This is particularly valuable in identifying newly arising type of service calls that may not be predefined in the system and appear as irregular requests or anomalies. Utilizing such models, systems can become capable of recognizing and categorizing these anomalies based on distinctive patterns and characteristics.

To tackle the challenges associated with categorizing service requests, particularly during crisis periods, our research employs automation techniques to introduce an innovative tool that integrates classification with anomaly detection. This tool automatically categorizes regular requests while promptly identifying anomalies, aiming to improve the adaptability and effectiveness of service management. The goal is to ensure that systems can respond efficiently to unprecedented events, thereby better serving the community in times of crisis.

In this study, we specifically investigate the 311 system of Orange County, Florida, to demonstrate the potential improvements in handling service requests through accurate classification and simultaneous anomaly detection, especially during disaster scenarios. Our analysis is supported by a comprehensive dataset of service requests obtained through collaboration with the Orange County 311 government. This partnership has afforded us access to valuable data, facilitating a thorough evaluation of the system's performance and the potential impacts of our proposed tool.

This study makes a valuable contribution to the efficient management of service systems, especially during crises. First, we introduce a novel and practical method for automatic request classification with high accuracy. We employ Support Vector Machines (SVM), a supervised machine learning technique, to classify service calls. The choice of SVM is motivated by its notable advantages, including rapid processing and robust performance, even with a relatively small dataset. The model is particularly beneficial for text classification tasks, often constrained by the availability of limited labeled samples. Consequently, our approach offers a practical and accurate solution for service providers, empowering them to enhance their operations through automated categorization.

Second, our method innovatively utilizes the decision function of SVM to incorporate anomaly detection capabilities. By adding an “extra category” to the predefined set, defined as “Other,” our model dynamically adjusts to unfamiliar types of calls. This adaptability ensures the system remains flexible and responsive to evolving service needs. Moreover, this enhancement significantly reduces the time needed to detect new categories, a crucial step in improving efficient service management, particularly during crises. This feature substantially improves the system's responsiveness and efficiency during crises, ensuring timely and appropriate service delivery.

RELATED WORK

In the literature, there is a notable gap in addressing challenges related to service request management, particularly with a focus on efficient and adaptive service categorization, especially during disasters. This summary encapsulates a few studies utilizing machine learning and artificial intelligence techniques on service data and relevant efforts in classification of crisis-related social media data to help with request management efforts of service systems.

Several studies focus on identifying service request patterns, volumes, and frequencies in different regions. For example, Cheng et al. (2022) employ machine learning techniques to forecast the number of non-emergency service requests in the Miami-Dade County and to identify the neighborhood traits within census tracts that influence call volumes, while Kontokosta et al. (2017) use gradient boosting regression to classify heat and hot water problems in New York City. Wang et al. (2017) explore the use of non-emergency service requests for decision support and socioeconomic trend prediction, employing k-means clustering to categorize census tracts.

Hashemi (2022) addresses automatic detection of service request types using recurrent neural networks, primarily focusing on classifying requests into preexisting categories. While our paper shares similarities with Hashemi (2022), our approach distinguishes itself by incorporating timely detection of anomalies, which demands more careful attention from service providers. Brogi et al. (2006) and Bianchini et al. (2006) discuss the use of semantic descriptions and ontologies in service discovery. Specifically, Brogi et al. (2006) present a matchmaking system for service compositions and Bianchini et al. (2006) propose a semantic-enriched framework for service description and discovery. Unlike previous work, our tool is designed to simultaneously identify outlier service requests, i.e., anomalies, that do not fit within any of the existing service categories while conducting automatic classification of requests.

In addition to efforts for classifying requests to service systems directly, recent studies have also focused on utilizing social media data for service providers' attention. With the increasing use of social media during disasters, relevant information extracted from such platforms has been found helpful for the crisis management efforts of local authorities. Li et al. (2018) utilized SVM among other techniques to classify tweets with high accuracy, but their approach lacks the anomaly detection concept that is helpful for service systems to differentiate the newly arising needs during crises from regular service needs. Similarly, Alam et al. (2019), Li et al. (2021), and Wiegmann et al. (2020) focused on classifying disaster-related content in social media data without the capability of detecting unusual disaster types.

Kersten et al. (2021) propose a machine learning model which improved the detection rate of crisis-related tweets, but this came with a higher false positive rate, leading to more non-crisis tweets being misclassified as crisis-related. Other studies, such as Aipe et al. (2018) and Buntain et al. (2022) were limited to classifying known crises and lacked the detection of unforeseen events. However, our tool extends the SVM model with a distinctive learning approach to identify unforeseen crisis-related requests like those specific to the COVID-19 pandemic, without prior training or information on such specific scenarios. This enables our model to adapt dynamically to emerging crises and enhance its effectiveness in crisis management.

In summary, while previous literature underscores the potential of automated techniques, semantic descriptions, and ontologies in service requests and relevant social media content, there is limited exploration of machine learning approaches for disaster management efforts in service systems. Our study aims to fill this gap by proposing a dynamic model for accurate service request classification and timely anomaly detection. By doing so, our tool contributes to the advancement of knowledge in this domain, especially in adaptive service management during emergencies. This distinctive approach sets our study apart from existing research, offering valuable insights into automation and efficient decision-making.

PROBLEM CONTEXT: 311 SERVICE SYSTEMS

Cities have increasingly embraced 311 systems for local non-emergency service management, driven by two key factors (Samuel, 2019). First, the implementation of 311 services aims to alleviate the workload on 911 operators by diverting non-emergency inquiries to dedicated 311 call centers. In doing so, calls related to non-emergency issues can be efficiently managed by operators with expertise in directing such concerns to the appropriate municipal departments. Secondly, these services serve as a convenient channel for residents to request serv ice and access information on a diverse range of municipal topics, spanning from garbage collection and snow removal to tax regulations and new local laws. This enhances citizen engagement and provides a centralized resource for obtaining relevant information. Consequently, 311 services often encompass a variety of predetermined categories for which residents might seek assistance. This structured approach ensures that residents receive prompt and specialized attention for their specific concerns, contributing to more effective municipal service delivery (O'Brien, 2016).

Over 300 cities and counties in North America have integrated 311 as a fundamental resource for municipal information and services (Winbourne, 2022). These systems function independently from each other, leading to a lack of uniformity in their operational frameworks. Different cities adopt various strategies in their service management based on their local needs, preferences, and agreements with business partners. For instance, while Chicago 311 fully incorporates automation technologies to streamline and categorize service requests(City of Chicago, 2018). Orange County Florida 311 favors manual categorization of requests and relies on human operators for the management of calls, showcasing the diverse approaches cities take in utilizing 311 services (Pamukcu et al., 2021). The significance of these systems in helping with local disaster management efforts was particularly highlighted during the COVID-19 pandemic when 311 served as a vital conduit for disseminating information critical to safeguarding the health and well-being of residents and their families (Pamukcu et al., 2021).

DATA AND METHODOLOGY

Data Description

For our study, we examine the Orange County, Florida 311 system to demonstrate accurate classification and prompt anomaly detection. Our analysis is supported by comprehensive service requests data from OCFL311, obtained through collaboration with the Orange County 311 managers. The dataset comprises 1,055,678 records spanning 2016 to 2021 with 14 predefined main categories of service calls. A record in the context of this study refers to an inquiry, which is created upon the receival of the service request. Each record is uniquely stored in the database, and it undergoes updates as necessary, such as follow-ups or closure upon resolution or forwarding to a specific municipal department. The deliberate choice of the Orange County 311 system was driven by its commitment to meticulous manual processing of service requests, with a primary focus on achieving the highest level of accuracy in request categorization.

Upon receiving service requests through various channels such as phone calls, chat, or mobile apps, each entry undergoes a comprehensive triage and processing phase facilitated by trained 311 operators. This careful handling is designed to guarantee the precise assignment of requests to their respective and appropriate categories. It's important to note that the manual triage of requests places a substantial demand on system resources, and the efficiency of this process directly impacts the overall performa nce of the system, as highlighted by Pamukcu et al. (2022). This underscores the significance of optimizing the request categorization process to enhance system efficiency and effectiveness.

The Orange County Florida 311 dataset offers comprehensive information about each request, including the exact location of the incident, detailed issue descriptions, along with the issue creation and completion dates, and resolution information. To enhance manageability and focus on essential aspects relevant to our study, we only retained critical information pertinent to our study, such as the main categories and detailed subcategories of the requests, along with descriptions and creation dates of the reported issues. These retained columns are instrumental for training our model, providing valuable insights into the nature and classification of service requests. The subsequent section will delve into the specifics of our data preprocessing methodology, providing a clear understanding of the steps taken to refine and prepare the dataset for our analytical processes.

Data Preprocessing

In this stage, we meticulously conducted data preprocessing as an integral part of our approach to text classification. Recognizing the importance of clean and well-structured data for the success of machine learning models, we employed a series of preprocessing steps. This included the removal of punctuation, conversion to lowercase, and elimination of stop words to refine the textual information and enhance its focus. We also addressed missing data to ensure the dataset's reliability and consistency. Tokenization and vectorization were applied to transform the text into a format suitable for our chosen machine learning model.

In order to transform text into a format suitable for our chosen machine learning model, we opted for the widely used Term Frequency-Inverse Document Frequency (TF-IDF) vectorization method. TF-IDF proved to be a robust choice, capturing the importance of terms in the context of the entire dataset, and providing a numerical representation that retained the significance of words while mitigating the impact of common terms (Shah et al., 2019). Furthermore, we labeled and encoded categorical labels, facilitating numerical representation. Our commitment to thorough data preprocessing played a crucial role in preventing overfitting and promoting efficient model training. The resulting clean and well-structured data contributed to the interpretability of our model results and facilitated the identification and resolution of potential issues.

Automatic Classification of Issue Descriptions with SVM

SVM is a supervised machine learning model that utilizes algorithms for classification problems (Cortes & Vapnik, 1995). The SVM model is trained on a set of labeled training data for each category, enabling it to categorize new text. In our investigation, we leveraged the SVM algorithm for the intricate task of multi-class classification in the context of issue description categorization. Originally designed for binary classification, SVM is extended to handle multiple classes through a one-versus-the-rest (OvR) approach (Cortes & Vapnik, 1995; Hsu & Lin, 2002). The OvR involves training a binary classifier for each class against the rest, and the class with the highest confidence score is assigned during prediction (Galar et al., 2011).

Specifically, the classification task was approached using the Linear Support Vector Classification (LinearSVC) algorithm, a variant of SVM tailored for high-dimensional datasets (Gualtieri & Cromp, 1999). The core hyperparameter governing the model's behavior is the regularization parameter. This parameter, set by default to 1.0, serves as the inverse of the regularization strength. Higher values of this parameter lead to reduced regularization, allowing the model to fit the training data more closely. Striking the right balance for this parameter is crucial as it determines how much emphasis the model places on fitting the training data versus preventing overfitting (Hastie et al., 2009).

The optimization process to prevent overfitting in LinearSVC involves minimizing a cost function that combines a hinge loss term for classification accuracy with a regularization term (Hastie et al., 2009). The Pegasos algorithm, a variant of stochastic gradient descent, is used to iteratively update the model parameters based on a subset of training examples, allowing the model to adapt to the complexities of the data (Shalev-Shwartz et al., 2007) The convergence of the training process was guided by adjusting the tolerance parameter. As the algorithm iterated over the dataset, the training process would stop when the change in the cost function (a combination of hinge loss and regularization) dropped below the specified tolerance. This ensured that the model achieved a satisfactory level of convergence, balancing accuracy on the training data with generalization to new, unseen examples (Apidopoulos et al., 2022).

Anomaly Detection for New Categories

In addition to our primary classification objective, our study emphasizes the early detection of new patterns through the integration of anomaly detection. This facet adds a crucial layer to our automatic classification system, enabling the timely identification of instances that deviated from established categories. The importance of early detection lies in the system's ability to proactively recognize emerging trends, ensuring a responsive approach to evolving data dynamics.

Anomaly detection, as a pivotal component, operated by identifying outliers within the dataset—instances that exhibited novel or unexpected patterns. These outliers, which are a new type of call, were promptly categorized as “Other” in a dynamic and iterative manner, marking a departure from traditional classification frameworks. The iterative nature of this process allowed our model to continuously refine the “Other” category as the dataset evolved, making it a proactive mechanism for adapting to shifts in patterns and trends.

In the technical intricacies of our automatic classification system, the decision function scores play a crucial role, offering a nuanced and data-driven approach to categorizing instances beyond predefined labels. The decision function scores, obtained through the LinearSVC algorithm, provide a measure of the distance of each instance from the decision boundary. These scores represent the model's confidence in its classification decision, with lower scores indicating proximity to the boundary.

The importance of finding the optimal decision score threshold cannot be overstated. This threshold delineates the boundary between instances classified within established categories and those categorized as "Other". The process of determining the threshold involved careful calibration to strike the right balance between sensitivity and specificity in anomaly detection. Sensitivity ensured the accurate identification of true anomalies, while specificity prevented the misclassification of instances merely representing variations within known categories.

In the quest for the best decision score threshold, we engaged in a systematic exploration, considering various values to assess their impact on the model's performance. This involved a trade-off analysis, as adjusting the threshold influenced the number of instances classified as "Other" and the potential impact on overall model accuracy.

The technical complexity of this process lies in its adaptability to changing data dynamics. As the dataset evolved, the decision score threshold underwent recalibration to accommodate new patterns and maintain the efficiency of anomaly detection. This dynamic adjustment allowed our automatic classification system to stay attuned to the subtle shifts in data patterns over time.

ANALYSIS AND RESULTS

As previously mentioned, we demonstrate the effectiveness of our approach in automatic request categorization and anomaly detection using the Orange County, Florida 311 System data. The system classifies requests into 14 main categories, as illustrated in Figure 1. The graph indicates the number of service requests and the proportional distribution of main service categories for the period between 2016 and 2021 after preprocessing the data. Notably, the Animals category experiences the highest call volumes, followed by Public Safety, and then Neighborhoods & Housing. This visualization provides insights into the relative frequency of service requests across different main categories during the specified time frame. Figure 1 also shows the imbalance across the 14 main categories, a challenging problem in the realm of text classification. Following data cleaning and preprocessing, which involved the removal of missing and incorrectly labeled data, the number of observations decreased from 1,055,678 to 600,362. This is a crucial step in the text classification before training the machine learning model.

Pandemic-related Requests Categorization in Early Phase

Our analysis is rooted from early signals indicating challenges in categorizing emerging service needs during the initial phases of the COVID-19 pandemic. Following the declaration of a state of emergency on March 9, the Orange County, Florida 311 system started to receive many pandemic-related calls. However, a dedicated category for COVID-19 was not established until May 8, 2020. Following this, the system introduced 23 new subcategories to consolidate pandemic-related inquiries under dedicated classifications.

In our preliminary analysis, we observe that similar requests were inconsistently assigned to different service categories. The absence of a specific category for this emergency led to the classification of pandemic-related inquiries into various existing categories, each managed by different departments. To identify the categories to which these inquiries were assigned, we filtered 311 requests based on the issue description attribute using keywords such as "covid," "corona," "pandemic," and "epidemic" directly related to the COVID-19 crisis. Our analysis of this subset revealed that the system had already received 3440 unique pandemic-related service requests, constituting 9.5% of the total requests, from the beginning of the pandemic until the creation of a new category to accommodate calls that did not fit well under any existing category. Similar types of requests were addressed under different main categories, specifically *Public Safety*, *Neighborhoods & Housing*, *Families, Health & Social Services,* and *General*.

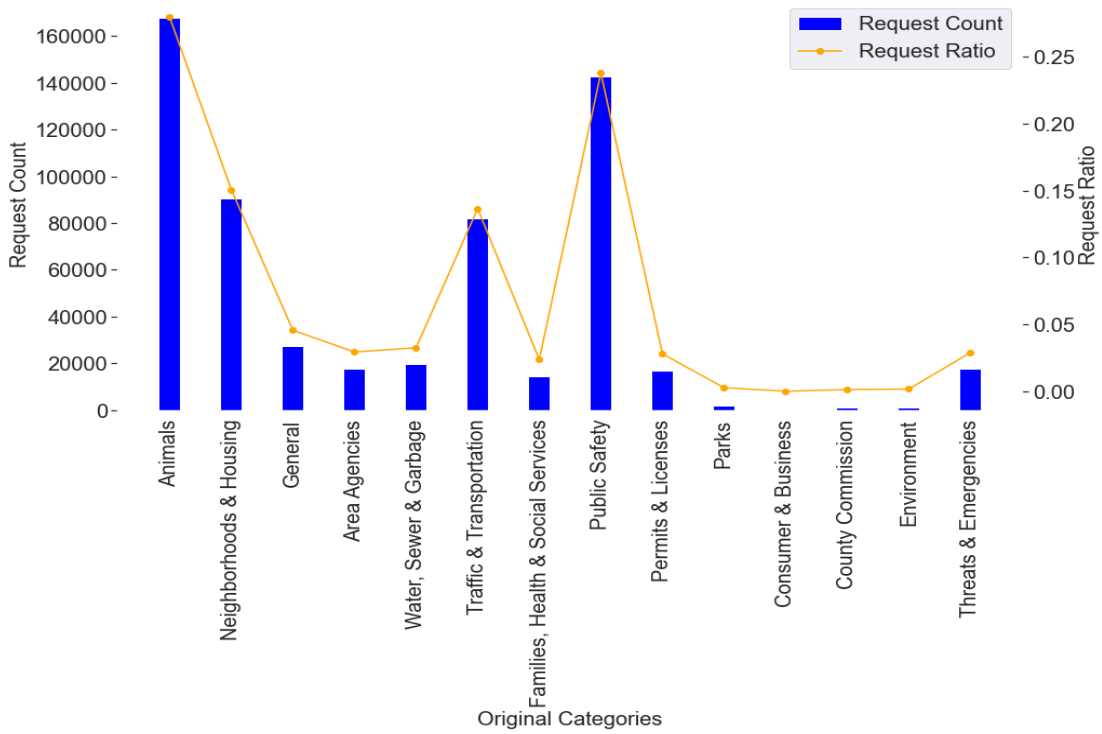

Figure 1. The proportional distribution of main service requests from 2016 until 2021

Support Vector Machine Model Results

The full preprocessed dataset was used in this stage to train and test the data with the LinearSVC algorithm. The dataset first has been divided into training and testing sets where 80% formed the training set and the remaining 20% formed the testing set. The TF-IDF vectorizer is then used to extract features from the data and finally the transformed features are used to fit our model. The testing set is then used for prediction and evaluation of the performance of our classifier model.

Table 1. SVM Model Results

Performance Metric	Result
Accuracy	92.19%
Precision	91.95%
Recall	92.19%
F1 Score	91.96%

Table 1 summarizes the results on the testing set, showcasing the proficiency of the SVM model in classifying service calls. With an accuracy of 92.19%, the model demonstrates remarkable reliability, correctly categorizing service calls with a high degree of certainty. Precision and recall, standing at 91.95% and 92.19% respectively, further underscore the model's precision in minimizing false positives and its comprehensive ability to capture relevant instances. The harmonized balance between precision and recall is encapsulated in an F1 score of 91.96%, highlighting the model's adeptness in maintaining equilibrium between exactness and completeness. Such robust performance metrics not only affirm the model's effectiveness in navigating classification challenges but also bolster its potential as a pivotal tool in the context of service request classification.

In order to thoroughly evaluate the performance of the proposed classification algorithm, we delved into a granular analysis of the model's predictions and the corresponding actual outcomes using the confusion matrix. Figure 2 provides a comprehensive breakdown, offering insights into the algorithm's performance across different true categories and their respective predictions. Within the confusion matrix, the diagonal elements stand as representations of instances correctly classified for each class. The varying shades along the diagonal convey the degree of success in true classification, with darker blue denoting higher accuracy. The figure showcases the model's robust performance across the majority of categories, except for *General* and *Water, Sewer & Garbage*. In these specific areas, the model's performance appears to exhibit some challenges, suggesting opportunities for further investigation and potential refinement to optimize classification accuracy.

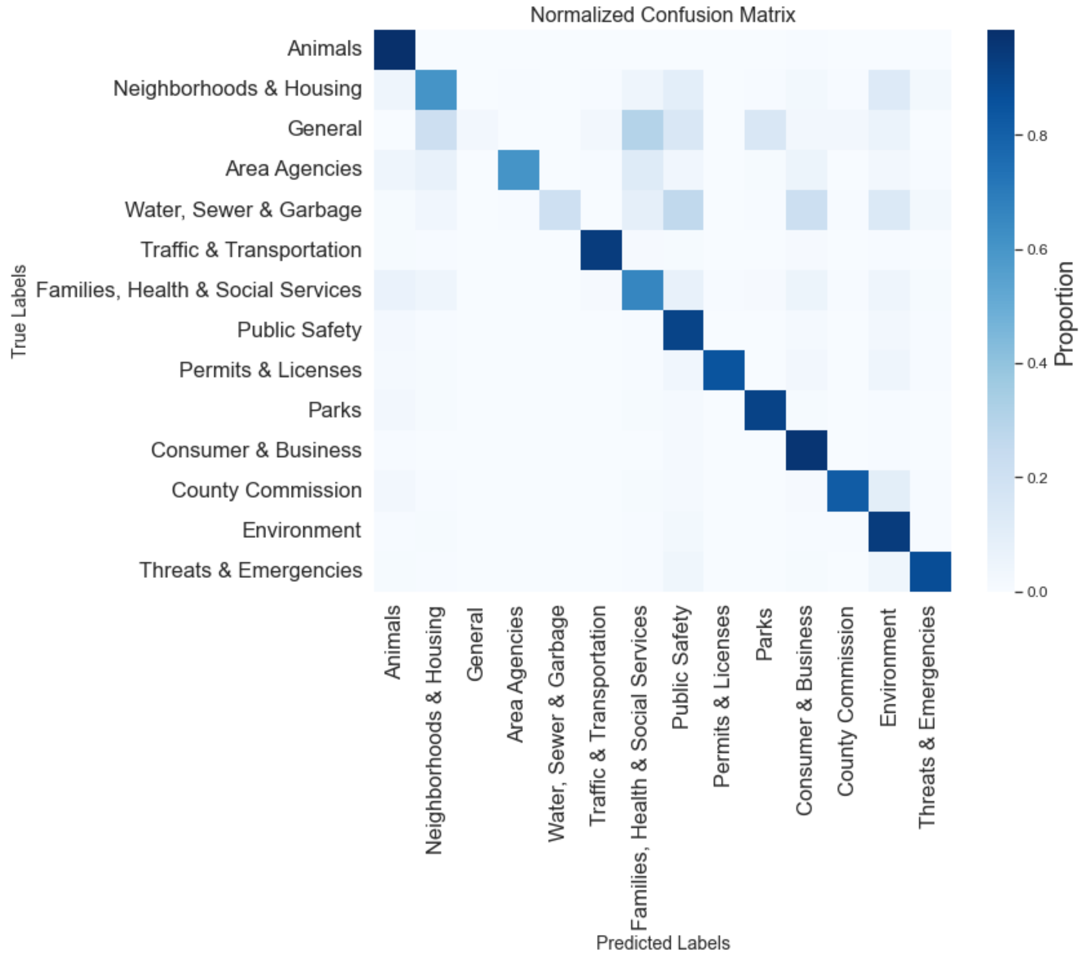

Figure 2. Confusion matrix for multiclass classification of Orange County, Florida 311 service categories

The complexity in classifying requests in the *General* category aligns with its nature, encompassing diverse requests related to information systems, incomplete calls, disconnected calls, undeliverable requests, or general inquiries not directly tied to any other category. Conversely, an examination of requests directed to the *Water, Sewer & Garbage* category reveals shared issues with several other categories, including concerns about water quality, hazardous waste, and standing water. This shared nature poses challenges for the model in accurately classifying these types of service requests. These insights contribute to a comprehensive understanding of the algorithm's strengths and areas for improvement, guiding future efforts to refine and optimize its performance across diverse service request categories.

Anomaly Detection for Unveiling Emerging Categories

In this part of our study, we employed the previously outlined anomaly detection methodology to identify irregularities within the service request patterns. This approach helps recognize unusual request types, signaling for new public needs and a possible emergence for additional request categories. To detect anomalies in service requests, the SVM model was trained on data spanning from January 1, 2016, until December 31, 2019. This timeframe was deliberately selected to encapsulate the anomalies that surfaced during the first year of the pandemic crisis, which broke out in Orange County, Florida on March 9, 2020.

Figures 3, 4, and 5 highlight these irregularities, starting from January 1, 2020, classified under "Other" category. The analysis reveals a notable number of anomalies in the service calls within the specified time frame. The surge in request volumes in this category immediately following the pandemic outbreak is indicative of atypical request patterns during the crisis. It serves as a clear signal for service operators to consider establishing new service categories to unify service requests related to the ongoing crisis. Illustrated by the COVID-19 crisis, the presented tool proves valuable in identifying an increase in anomalies within service requests, regardless of the specific emergency scenario.

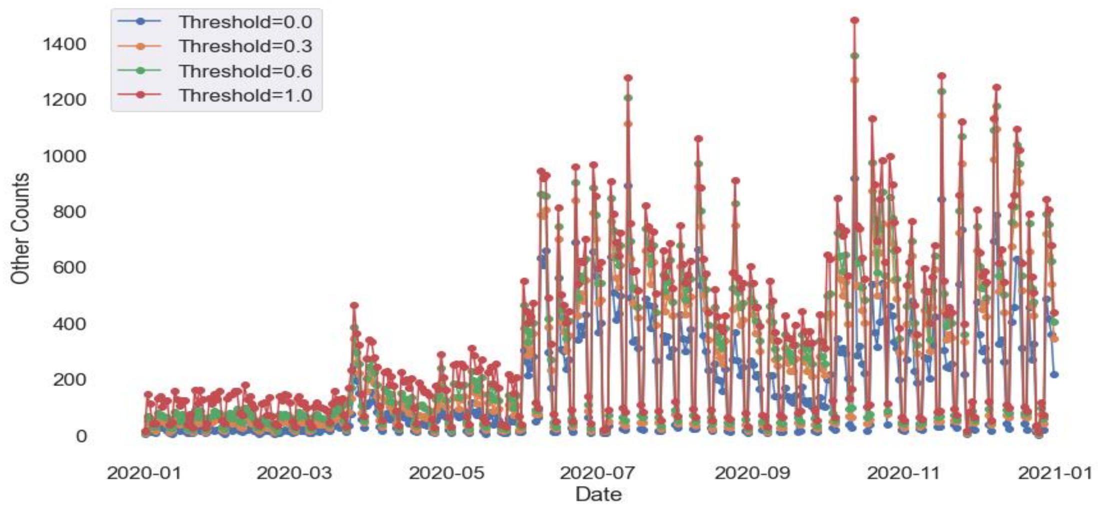

Figure 3. Number of requests classified in "Other" category with different selection thresholds.

To establish the optimal confidence threshold for classification for the "Other" category, we have examined four different thresholds ranging from 0.0 to 1.0, illustrated in Figure 3. In our anomaly detection model, the proximity of a request to each category is expressed by a confidence value, and a threshold is defined to assign requests to "Other" category if the confidence falls below the minimum threshold. This is crucial for capturing an anomaly when a request does not fit well into any of the existing categories, especially when the ratio of "Other" calls to the total number of calls is minimal.

Through a sensitivity analysis on various threshold values, we concluded that a confidence threshold of 0.0 establishes the optimal boundary. While the anomaly pattern remained consistent across different thresholds, our observations indicated that choosing higher thresholds resulted in the model incorrectly classifying requests under the "Other" category since the model became less sensitive. The selected threshold of 0.0 enhanced the accuracy of our model and effectively distinguished between well-known categories and anomalies, ensuring appropriate classification under the "Other" category.

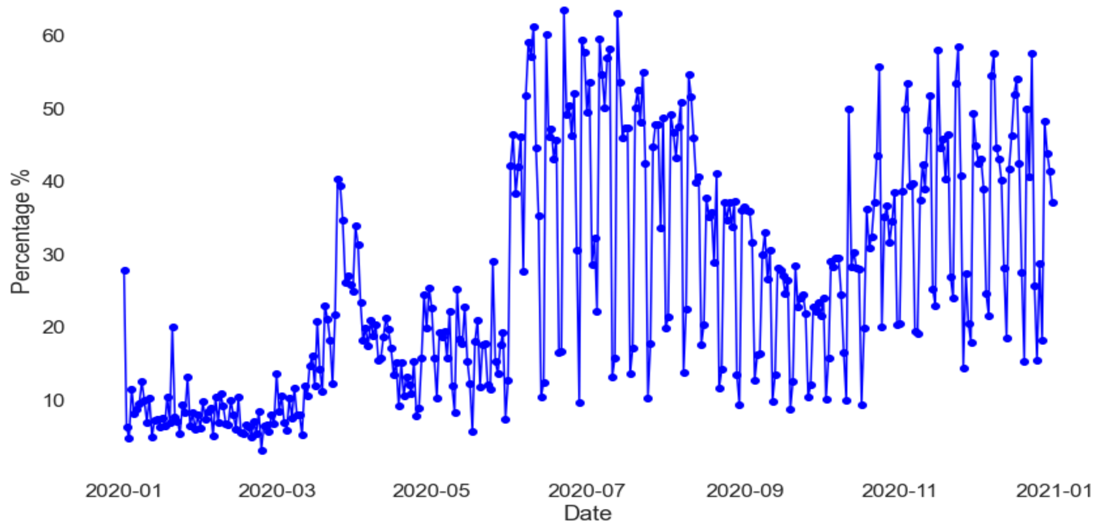

Figure 4. The percentage of requests in the "Other" category among the total number of service requests

Figure 4 illustrates the proportion of "Other" category requests in relation to the overall service inquiries. The data, trained up until January 1, 2020, shows a consistent pattern up to March 2020. Post-March 2020, however, there is a noticeable increase in variability, indicating a rise in service requests falling under the "Other" category.

This period coincides with the intense phase of the COVID-19 pandemic, leading to a surge in unique inquiries in the system. Consequently, the model began to assign these COVID-19 related inquiries to the “Other” category. This finding highlights the utility of our tool in identifying significant increase in unique calls, underscoring a potential need for introducing new categories within the service request system.

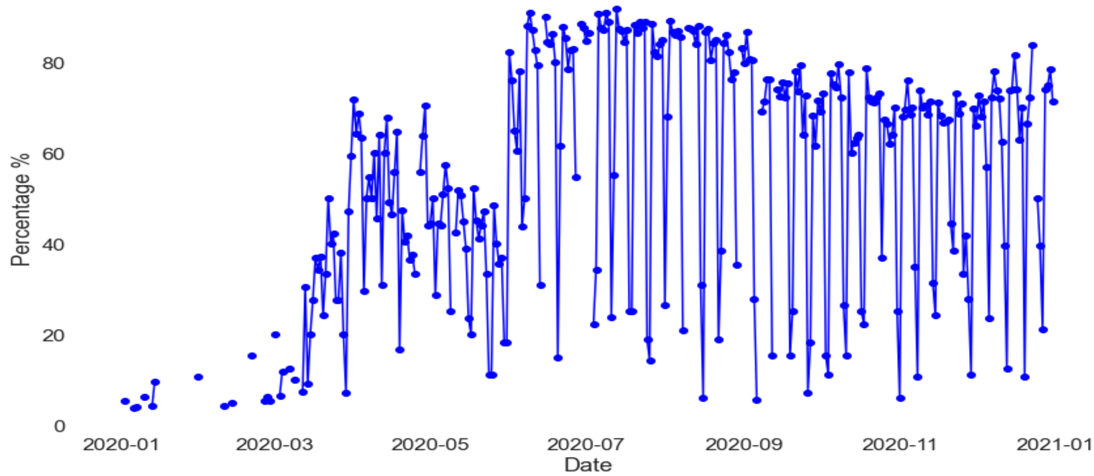

Figure 5. The percentage of pandemic-related anomalies within the total number of requests in the “Other” category

Moreover, to illustrate the proportion of pandemic-specific inquiries within the detected anomalies, we present the percentage of pandemic-related anomalies relative to the total number of requests in the “Other” category. Figure 5 indicates an increase in this percentage after March 2020, aligning with the emergence of new types of calls in response to the state of emergency declared in Orange County, Florida, on March 9, due to the COVID-19 outbreak. Another significant rise is observed in June, coinciding with the announcement of the Coronavirus Aid, Relief, and Economic Security (CARES) Act by the Orange County Government to provide financial support to residents, business owners, nonprofits, municipal and community partners in response to the pandemic (*Orange County Government, Florida*, n.d.)

Consequently, these calls were classified under “Other,” contributing to the overall rise in the number of “Other” service requests. These results affirm the effectiveness of our model in identifying such anomalies, which were later addressed by adding new service categories specific to these needs. Specifically, during the COVID-19 crisis, the Orange County Florida 311 system established 23 new subcategories, such as Orange CARES, Vaccine Outreach, and Testing, to group these pandemic-related requests under distinct subcategories within the *Public Safety* main category. Our analysis indicates the success of the presented model in promptly detecting emerging needs which can notify service providers about the emergence of new categories before the number of anomalies increases extensively.

Moreover, through periodic retraining of the model, the tool gains the ability to recognize whether the system has introduced new categories. This ensures that incoming calls are appropriately classified into recently established categories, rather than defaulting on the generic label of “Other.” This adaptability is crucial for staying responsive to evolving circumstances and augmenting the system's precision in categorizing service requests.

To highlight the advantages of periodically retraining the model, Figure 6 illustrates the variation in the number of service calls classified under the category “Other”, comparing the data trained before and after the onset of COVID-19, depicted in red and blue lines, respectively. Initially, the dataset was trained on records up to March 1, 2020, marked in red, which did not include specific categories for pandemic-related issues. Subsequently, the model was retrained with data up to April 1, 2021, during which time, new categories to capture COVID-19 related inquiries were established.

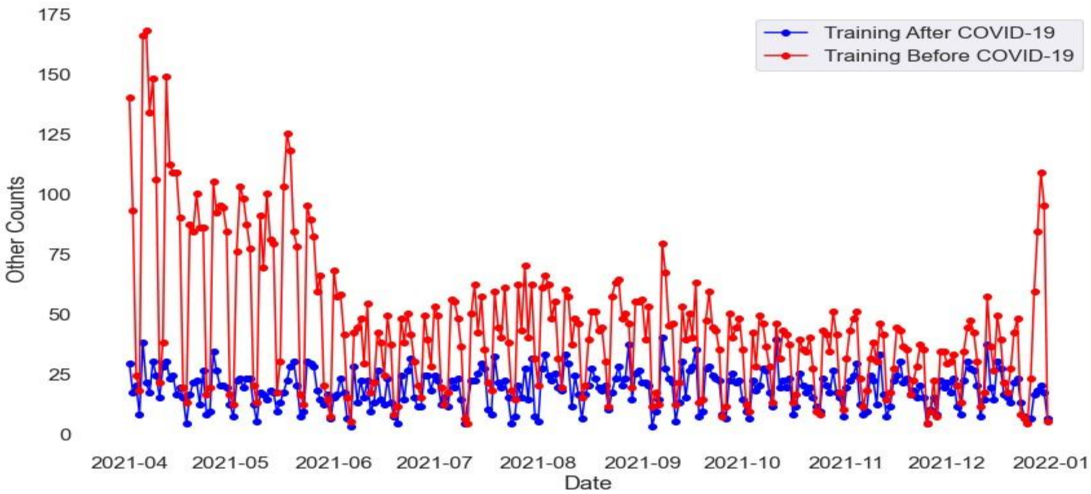

Figure 6. Number of requests classified under "Other" category with training before and after COVID-19

The visualization illustrates a high number of calls classified as "Other" with training before COVID-19, as indicated by the red line, due to the model interpreting all pandemic-related inquiries as anomalies. In contrast, the blue line shows much smaller calls in "Other" category and a consistent pattern after the model is retrained with data up to April 2021, encompassing the first year of the pandemic. The blue line exhibits a small and stable call volume, since COVID-related inquiries are now accurately classified into *Public Safety* main category. Notably, the remaining calls in the "Other" category suggest anomalies stemming from other factors that the model struggles to categorize.

This analysis demonstrates that periodic retraining enables the model to reassign requests previously labeled as "Other" to newly established categories if these inquiries become regular. Hence, the categorization of requests as "Other" is continually refined with each retraining cycle, ensuring it only reflects current anomalies. Thus, the model acts as a valuable tool for service operators, signaling the need to periodically review anomalies and consider creating new categories to effectively classify the evolving types of service calls.

DISCUSSION AND CONCLUSIONS

This work-in-progress paper emphasizes the significant potential of automatic classification of service requests and prompt detection of anomalies to identify emerging needs, especially during disasters. This capability is instrumental in supporting crisis management efforts by enabling public service providers to take timely actions during emergencies. The study's scientific contribution lies in bridging the gap identified in existing literature regarding machine learning approaches for disaster management efforts in service systems. Specifically, while previous studies have primarily focused on classifying service requests into preexisting categories, our proposed model addresses this limitation by incorporating timely anomaly detection.

Our preliminary investigations highlighted a critical issue—the delay in opening new categories for emergent types of service calls. This challenge was especially notable during the COVID-19 period, where it took an extended two months to introduce a category dedicated to COVID-19. Our tool markedly enhances this process, facilitating swift and precise categorization of requests and the identification of irregular requests. Notably, the application of the SVM model demonstrated a 92.19% accuracy rate in classifying service categories. The utility of the approach extends beyond reducing manual classification time for operators; it expedites the recognition and proper categorization of new types of categories. This streamlined approach enhances the overall process by not only automating classification of regular requests but also freeing up human agents to concentrate on detected irregular requests that necessitates a manual evaluation.

The contribution of the study extends beyond merely detecting anomalies related to specific crises; it offers a flexible and adaptable approach applicable to any service system facing disruptions due to evolving service needs during emergencies. Consequently, the presented tool is adaptable and responsive, capable of recognizing and managing newly emerged types of service requests. This underscores the model's extensive applicability in service request management, showcasing its potential to enhance the adaptability and responsiveness of public service systems, particularly in addressing evolving societal needs. This flexibility emphasizes the practicality of the tool in service call systems.

Although the current model excels in classifying main categories, this study faces limitations in classifying requests into subcategories due to the limited number of observations for numerous subcategories. Future research should prioritize improving the model's ability to classify subcategories, which has significant potential in achieving a more nuanced and precise service call categorization and identification of anomalies.

Further improvements could include exploring additional classification models to benchmark against our current method, determining if enhancements are necessary. Moreover, integrating automated alert systems could dynamically notify relevant authorities or response teams upon detection of new types of service needs. Such proactive approaches would ensure swift and targeted responses to emerging challenges, significantly contributing to a more effective overall disaster management strategy.

In conclusion, this study offers a substantial foundation for improving service call categorization systems, with potential widespread implications for emergency response and disaster management. Nonetheless, continuous advancements are essential to fully realize the tool's capabilities and to address the ever-evolving landscape of societal needs. As part of our future work, we plan to incorporate additional classification algorithms, explore different sets of hyperparameters, and implement cross-validation techniques. These efforts will ensure the robustness and generalizability of our findings, enabling a more comprehensive comparison of different modeling approaches and their impact on the classification task.

ACKNOWLEDGMENTS

This research was supported by the University of Vermont's Sustainable Development Policy, Economics, and Governance (SDPEG) interdisciplinary Ph.D. program, and by a grant from the National Science Foundation (NSF-SCC #1952792). The authors would like to thank Orange County, Florida 311 for their collaboration and for sharing their datasets. We also acknowledge the primary and equal contribution of the first and second authors, Hakan Unveren and Ayesh Abu Lehyeh, to this work.

REFERENCES

Aipe, A., Mukuntha, N., Ekbal, A., & Kurohashi, S. (2018). Deep learning approach towards multi-label classification of crisis related tweets. In Proceedings of the 15th ISCRAM Conference.

Alam, F., Ofli, F., & Imran, M. (2019). CrisisDPS: Crisis Data Processing Services. In Proceedings of the 16th ISCRAM Conference.

Apidopoulos, V., Poggio, T., Rosasco, L., & Villa, S. (2022). Iterative regularization in classification via hinge loss diagonal descent. *arXiv preprint arXiv:2212.12675*.

Bianchini, D., De Antonellis, V., Melchiori, M., & Salvi, D. (2006). Semantic-enriched service discovery. 22nd International Conference on Data Engineering Workshops (ICDEW'06).

Brogi, A., Corfini, S., Aldana, J. F., & Navas, I. (2006). Automated discovery of compositions of services described with separate ontologies. International Conference on Service-Oriented Computing. Berlin, Heidelberg: Springer Berlin Heidelberg.

Buntain, C., McCreadie, R., & Soboroff, I. (2022). Incident Streams 2021 Off the Deep End: Deeper Annotations and Evaluations in Twitter. In Proceedings of the 19th ISCRAM Conference.

Cheng, S., Ganapati, S., Narasimhan, G., & Yusuf, F. B. (2022). A machine learning-based analysis of 311 requests in the Miami-Dade County. *Growth and Change*, *53*(4), 1627-1645.

City of Chicago. (2018). *Mayor Emanuel Launches Modernized Chicago 311 System*. Press Room. https://www.chicago.gov/city/en/depts/mayor/press_room/press_releases/2018/december/Chicago311.html

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, *20*, 273-297.

Cutway, A. (2020). Timeline: The Spread of coronavirus in Florida. *Click Orlando.com*. https://www.clickorlando.com/news/local/2020/03/20/timeline-the-spread-of-coronavirus-in-florida/

Filip, R., Gheorghita Puscaselu, R., Anchidin-Norocel, L., Dimian, M., & Savage, W. K. (2022). Global challenges to public health care systems during the COVID-19 pandemic: a review of pandemic measures and problems. *Journal of Personalized Medicine*, *12*(8), 1295.

Galar, M., Fernández, A., Barrenechea, E., Bustince, H., & Herrera, F. (2011). An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes. *Pattern Recognition*, *44*(8), 1761-1776.

Gualtieri, J. A., & Cromp, R. F. (1999). Support vector machines for hyperspectral remote sensing classification. In 27th AIPR workshop: Advances in computer-assisted recognition (Vol. 3584, pp. 221-232). SPIE.

Hashemi, M. (2022). Automatic Type Detection of 311 Service Requests Based on Customer Provided Descriptions. *Applied Artificial Intelligence*, *36*(1), 2073717.

Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2). Springer.

Hsu, C.-W., & Lin, C.-J. (2002). A comparison of methods for multiclass support vector machines. *IEEE transactions on Neural Networks*, *13*(2), 415-425.

Kersten, J., Bongard, J., & Klan, F. (2021). Combining supervised and unsupervised learning to detect and semantically aggregate crisis-related twitter content. In Proceedings of the 18th ISCRAM Conference.

Kontokosta, C., Hong, B., & Korsberg, K. (2017). Equity in 311 reporting: Understanding socio-spatial differentials in the propensity to complain. *arXiv preprint arXiv:1710.02452*.

Li, H., Caragea, D., & Caragea, C. (2021). Combining self-training with deep learning for disaster tweet classification. In Proceedings of the 18th ISCRAM Conference.

Li, H., Li, X., Caragea, D., & Caragea, C. (2018). Comparison of word embeddings and sentence encodings as generalized representations for crisis tweet classification tasks. In Proceedings of ISCRAM Asia Pacific 2018 Conference.

O'Brien, D. T. (2016). 311 hotlines, territoriality, and the collaborative maintenance of the urban commons: Examining the intersection of a coproduction policy and evolved human behavior. *Evolutionary Behavioral Sciences*, *10*(2), 123.

Orange County Government, Florida. (n.d.). Retrieved 02/02/2024 from https://www.ocfl.net/AboutUs.aspx

Pamukcu, D., Zobel, C., & Ge, Y. (2022). A Data Envelopment Analysis-based approach for managing performance of public service systems during a disaster. In Proceedings of the 19th ISCRAM Conference.

Pamukcu, D., Zobel, C. W., & Ge, Y. (2021). Analysis of Orange County, Florida 311 system service requests during the COVID-19 pandemic. In Proceedings of the 18th ISCRAM Conference.

Samuel, L. (2019). *Cities Embrace New and Improved 311 Services*. Data-Smart City Solutions Retrieved 22/01/2024 from https://datasmart.hks.harvard.edu/news/article/cities-embrace-new-improved-311-services

Schwester, R. W., Carrizales, T., & Holzer, M. (2009). An examination of the municipal 311 system. *International Journal of Organization Theory & Behavior*, *12*(2), 218-236.

Shah, S. A., Seker, D. Z., Hameed, S., & Draheim, D. (2019). The rising role of big data analytics and IoT in disaster management: recent advances, taxonomy and prospects. *IEEE Access*, *7*, 54595-54614.

Shalev-Shwartz, S., Singer, Y., & Srebro, N. (2007). Pegasos: Primal estimated sub-gradient solver for svm. In Proceedings of the 24th International Conference on Machine Learning.

Wang, L., Qian, C., Kats, P., Kontokosta, C., & Sobolevsky, S. (2017). Structure of 311 service requests as a signature of urban location. *PloS one*, *12*(10), e0186314.

Wiegmann, M., Kersten, J., Klan, F., Potthast, M., & Stein, B. (2020). Analysis of detection models for disaster-related tweets. *Analysis of Detection Models for Disaster-Related Tweets*, 872-880.

Winbourne, J. (2022, October 18, 2022). *311 Services — A key Step in the Global Transition to e-Government Services* https://www.w-llc.com/311-services-a-key-step-in-the-global-transition-to-e-government-services/

Wu, W. N. (2020). Determinants of citizen-generated data in a smart city: Analysis of 311 system user behavior. *Sustainable Cities and Society*, *59*, 102167.