Received: 29 September 2020 | Revised: 7 September 2021 | Accepted: 27 September 2021

DOI: 10.1111/grow.12578

growth and change WILEY

SPECIAL ISSUE ARTICLE

A machine learning-based analysis of 311 requests in the Miami-Dade County

Shaoming Cheng[1] | **Sukumar Ganapati**[1] | **Giri Narasimhan**[2] | **Farzana Beente Yusuf**[2]

[1]Department of Public Policy and Administration, Florida International University, Miami, Florida, USA

[2]School of Computing and Information Sciences, Florida International University, Miami, Florida, USA

Correspondence
Shaoming Cheng, Department of Public Policy and Administration, Florida International University, Miami, FL, USA.
Email: scheng@fiu.edu

Funding information
Generous financial support of NSF Grant, Grant/Award Number: #1924154

Abstract

This paper illustrates the application of machine learning algorithms in predictive analytics for local governments using administrative data. The developed and tested machine learning predictive algorithm overcomes known limitations of the conventional ordinary least squares method. Such limitations include but not limited to imposed linearity, presumed causality with independent variables as presumed causes and dependent variables as presume result, likely high multicollinearity among features, and spatial autocorrelation. The study applies the algorithms to 311 non-emergency service requests in the context of Miami-Dade County. The algorithms are applied to predict the volume of 311 service requests and the community characteristics affecting the volume across Census tract neighborhoods. Four common families of algorithms and an ensemble of them are applied. They are random forest, support vector machines, lasso and elastic-net regularized generalized linear models, and extreme gradient boosting. Two feature selection methods, namely Boruta and fscaret, are applied to identify the significant community characteristics. The results show that the machine learning algorithms capture spatial autocorrelation and clustering. The features generated by fscaret algorithms are parsimonious in predicting the 311 service request volume.

1 | INTRODUCTION

The Federal Communications Commission designated the 311 phone number for public agencies to provide non-emergency public services in 1997. The allocation of 311 was designed to reduce the volume of 911 calls to police departments, which are mainly for emergency life-saving calls. Local governments (i.e., cities and counties) use the 311 customer centers as one-stop centers to fulfill non-emergency service requests, such as bulk trash pickup, pothole repairs, city service inquiries, or complaints. Residents who need these services can simply call the 311 customer center, instead of searching and identifying which department to call for a service. The service requests are routed to the appropriate department through a customer relationship management system on the backend. There are over 300 local governments with such 311 customer centers and they provide critical services to the local communities (Ganapati & Scutelnicu, 2015). In the Covid-19 pandemic context, the 311 centers have become important centers of disseminating information about testing sites, food security, mental health, childcare, and other resources (Descant, 2020).

The 311 customer centers are new sources of big administrative data. Administrative data are the data derived from the operation of government administrative systems such as registration, licensing, as well as delivery of services (e.g., in education, healthcare, taxation, criminal justice, or housing) (Elias, 2014). These administrative data are generated for specific departmental purposes rather than research purposes. Researchers typically are not involved in the data's design, collection, structure, or content. The administrative data are powerful, but underutilized resources. They can shed light on operational efficiency, service delivery, social inequality, and effectiveness of public policies and programs (Einav & Levin, 2013).

There is an urgent need to develop analytical approaches and methods for using and applying administrative data for decision-making processes. The strength of the "big data" is its volume, variety, and velocity (Connelly et al., 2016). Although administrative data are not as "big" as other data sets collected from social media posts, commercial transactions, or cell phone tracking, the administrative data are still complex and large for social scientists. Manipulating the administrative data for decision making requires advanced capacity in data management, transformation, and analytics. For example, New York City's 311 service request records for any given year are well over one gigabyte (GB), which is beyond the capacity of ordinary statistical software such as Excel or SPSS.

The purpose of the paper is to exemplify a machine learning predictive algorithm for public policy, using the 311 data. The paper analyzes the spatial and temporal patterns of service requests to uncover the factors that predict the call volume. Although we use the 311 data from Miami-Dade County as an empirical illustration for this paper, our aim is more to provide a methodological foundation to generalize the algorithm's application for analyzing administrative data. The principles of analysis can be extended 311 administrative data from other cities.

The developed and tested machine learning predictive algorithm overcomes known limitations of the conventional ordinary least squared method. Such limitations include but not limited to imposed linearity, presumed causality with independent variables as presumed causes and dependent variables as presume result, likely high multicollinearity among features, and spatial autocorrelation. All the limitations are explicitly and successfully addressed in the developed predictive algorithm to show the usage and advantages of the predictive capacity. Specifically, random forest and other tree-based algorithms will be tested which relax the assumption of a linear relationship and can predict relying on associations rather than presumed causality. More importantly, two feature selection methods are applied and tested to show means to address potential multicollinearity or overfitting when high-dimensional and correlated data are used. Spatial data science methods are also employed to demonstrate that the developed predictive algorithm is capable of coping with spatial autocorrelation.

Identifying the patterns of service requests and their determinants is important for answering crucial policy questions about deploying resources and public services more equitably and efficiently to *all* residents. Residents are increasingly engaged in asking for public services. Government's allocation of services is increasingly linked to such preference expression. For example, residents living in politically connected neighborhoods may demand more of public services from their elected officials, and the local government may respond to such neighborhoods preferentially. There are equity questions surrounding such distribution of municipal resources and provision of public services across neighborhoods—governments are not expected to demonstrate a systematic bias. Minority and/or economically disadvantaged neighborhoods often do not get adequate service simply due to their non-participation in requesting services from local governments (Thomas & Streib, 2013). Local governments are not aware of public service disruptions when residents do not participate in demanding services. Even when 311 centers are designed to make governments more accessible and to facilitate engagement and participation in local governance, governmental response time to service requests may vary between the different types of neighborhoods. Analysis of the 311 administrative data offers a refined understanding of spatial and temporal patterns of the service requests for identifying participation gaps, enhancing governmental preparedness, and achieving equitable and efficient public service provision.

This paper does not directly examine propensity of and its disparities among different population groups or variations in infrastructure quality/conditions across communities. Impacts of propensity to report and infrastructure quality/conditions across communities, however, are likely be incorporated in the predictive outcome because they are highly associated with community characteristics which are the key features for predicted volumes and distributions of 311 service requests. Both propensity to request and communities' infrastructure conditions are strongly correlated with community factors. For example, low propensity of participation among minority populations is correlated with distribution of minority groups, poverty rate, unemployment rate, housing characteristics, or education attainment. Similarly, infrastructure quality, such as likelihood of a pothole, is correlated with community environment too, that is, minority and economic challenged communities would have inferior community conditions. When diverse community attributes are incorporated in the developed and tested machine learning predictive algorithm, their influences on resident participation are captured in the prediction of counts of 311 requests.

Two related research questions guide this paper in applying the machine learning algorithms. The first question is: to what extent can we predict the residents' participation in 311 service requests across neighborhoods? Here, we apply the algorithms to predict the total annual counts of 311 service requests across Census tracts in Miami-Dade County. Answering the question is useful in understanding the geographic variations in service requests across neighborhoods (a Census tract approximates a neighborhood in this context). The second question is: what are the community level determinants that predict service requests in a neighborhood? The machine learning algorithms are used in a novel way to answer this question: the algorithm is trained using the year 311 data from 2017 in order to predict annual counts of 311 requests in a prior year (2014) and the subsequent year (2018). The algorithm takes into consideration a range of community characteristics, including demographic, social, economic, and employment features. It also accounts for spatial autocorrelation, which is an important consideration for considering spatial spillovers in small(er) geographic units.

The machine learning algorithm starts with identifying spatial autocorrelation and clustering of 311 requests by Census tracts. The Census tracts are used to approximate neighborhoods since they are generally cohesive small areas with a sense of *community*. Zip codes are much larger in geographic size and may have more variations within, so that they may not be reflective of neighborhoods. More granular geographic units that the Census tracts, such as Census block groups, can also be used for analysis.

The spatial exploratory analysis is then followed by a feature selection process which is to evaluate and select "important" features of community characteristics that determine the 311 service requests. The characteristics considered include race and ethnicity, gender, income, housing, education, employment, and age. Two feature selection algorithms (Boruta and fscaret) are compared for how accurately they predict the volume of 311 service requests. The features generated by these two algorithms are applied independently to demonstrate the algorithm's efficacy in feature selection and in predicting 311 service requests. The machine learning algorithms used for predicting the volume of service requests include random forest, support vector machines (SVM), lasso and elastic-net regularized generalized linear models, and extreme gradient boosting. These predictive algorithms are selected because they are the most commonly used approaches in machine learning. An ensemble prediction is also implemented based on each learning algorithm's separate prediction.

Selecting the 311 service requests for 2017 as training data and applying it for predicting 2014 and 2018 service requests requires explanation. Testing the prediction of 2018 service requests is justifiable since these requests could be hypothesized to be influenced by the previous year's requests. However, the temporal closeness could also arguably result in temporal correlation. Hence, to enhance robustness of the prediction from the trained algorithm, a retrospective data set from a prior year (2014) is used, which would be more exogenous and temporally independent from a future year. Last but not least, predictive errors are examined for spatial autocorrelation and clustering to show the extent to which spatial autocorrelation and clustering are controlled in the algorithm.

2 | LITERATURE REVIEW

The open data and open government movement have gathered momentum worldwide as public agencies have made their data available online (e.g., data.gov in the United States). Local, state, and federal government agencies mandatorily collect vast amounts of data. The data are of different types (quantitative, spatial, textual, image, voice) and from various sources (record keeping, surveys, public meetings, sensors). As a result, public agencies are vast repositories of administrative data (Ganapati & Reddick, 2012, 2014). Yet, these data have limited use for researchers because they are unstructured and have various degrees of integrity, quality, and accessibility (Wang & Shepherd, 2020; Wang et al. 2018). The data are also too big and complex for typical social science researchers (Connelly et al., 2016). If the data are standardized and made accessible and if appropriate public policy analytical algorithms are developed and readily available, public policy research and practice may be greatly enhanced from new insights gained from analysis of the vast and complex administrative data.

The 311 service requests data are emblematic of the big administrative data from local governments. These data are collected and maintained by 311 customer contact centers across many local governments in the United States. These 311 centers are modeled after 911, which has been long used for public safety and emergency services by the law enforcement agencies within local governments. Local governments adopted 311 initially to reduce the volume of non-emergency calls coming to 911 centers. Since then, the 311 centers have emerged as a hub for all local government services, fielding both information and service requests from residents.

The 311 centers receive requests for information (e.g., how much do I owe on my property tax) and services (e.g., paying the property tax, picking up trash, fixing potholes, etc.). Residents traditionally made such requests to the 311 centers via phone calls only, but modern 311 centers allow the requests to be initiated through phone, apps, or website. Information requests comprise about 70% of the request volume and service requests comprise the rest 30%. The 311 center employs professionals who respond to the information and service requests. They forward the service requests to the relevant governmental departments or agencies through a customer relationship management system. The working hours of these 311 centers vary from 24/7 to activation for limited time periods during emergencies (e.g., in the Covid-19 pandemic situation).

Studying the pattern of 311 service requests is important not only because they connect resident demand/need directly with governmental service delivery, but also because they are important means of coproduction of public goods and services. Ostrom (1996, p. 1,073) defined coproduction as “the process through which inputs used to provide a good or service are contributed by individuals who are not ‘in’ the same organization.” Coproduction is manifest when residents request information, provide assistance in public service delivery, and interact with public agencies on obtaining public services (Whitaker, 1980). Coproduction normatively supports democratic governance and public accountability through long-term interaction between residents and governments (Jakobsen & Andersen, 2013; Meijer, 2011). From an operational or managerial perspective, coproduction improves the efficiency and effectiveness of public services (Levine & Fisher, 1984). Residents are no longer passive recipients of public services who await local governments to identify and then fix a service outrage or disruption. They are instead encouraged to actively report their service requests and service issues in a process of working collaboratively with local governments to locate and resolve an issue in a timely and satisfactory manner. The rapid proliferation of 311 centers represents the resurgence of coproduction in urban governance, public administration, and service delivery (Nabatchi et al., 2017).

Existing literature on 311 service requests centers on civic participation and distributional equity. Residents from different racial, ethnic, and social-economic groups may have distinct attitudes toward making 311 service requests (Chatfield & Reddick, 2018; Clark et al., 2020). Older people, for example, normally prefer to call the city for specific services. Technologically savvy young professionals could be comfortable with making the service requests through online apps and social media. Age could thus be a distinctive factor in determining participation via phone apps or website compared with traditional phone calls (Pak et al., 2017). Moreover, digital divide among the socially disadvantaged and low-income residents could impede them from making the 311 service requests. Low participation in 311 requests may further exacerbate individual and community disparities in service delivery. Individual residents who file 311 requests tend to receive expedited service restoration, because closing a request is often a performance indicator for the public organization (Xu & Tang, 2020). Communities with disproportionate share of socially disadvantaged and economically challenged residents may, as a whole, receive low service delivery and/or slow service restoration, because a service issue or disruption is less likely to be reported and be made aware to local governments (Pak et al., 2017).

The extant literature, however, does not provide a method for analyzing and predicting the 311 requests. Forecasting the 311 service requests will enable and empower local governments to proactively allocate resources, identify service disparity, and achieve service efficiency and equity across communities and residents. Filling this methodological gap is important for local governments to use the 311 data effectively to make decisions and evaluate their service provision. It is in this context that this paper aims to advance a machine learning model for predicting 311 request data, using the Miami-Dade County as an example. The prediction algorithm takes into consideration the community characteristics that influence the service requests.

3 | DATA

The data on 311 requests are readily available from Miami-Dade County's open data hub (https://gis-mdc.opendata.arcgis.com/). The 311 data from the periods of 2017, 2018, and 2014 are utilized for analysis in this paper. The data for the year 2020 had not been made available by the time of writing this paper. The 2020 data are also a deviation from past years because of the Covid-19 pandemic. Although current data sets and analysis do not directly pertain to Covid-19, algorithms used in this paper can be adapted and expanded in the future for analyzing Covid-19 requests with other types of requests.

Each 311 request provides the location information (latitude and longitude coordinates) based on the address where public service was requested and needed. Figure 1 presents a choropleth map showing tract-level spatial distribution of 311 requests based on Jenks natural breaks. In addition, the 311 service request data contain detailed descriptions of requested services, targeted and actual days of service completion, and modality of service requests. For each year, all 311

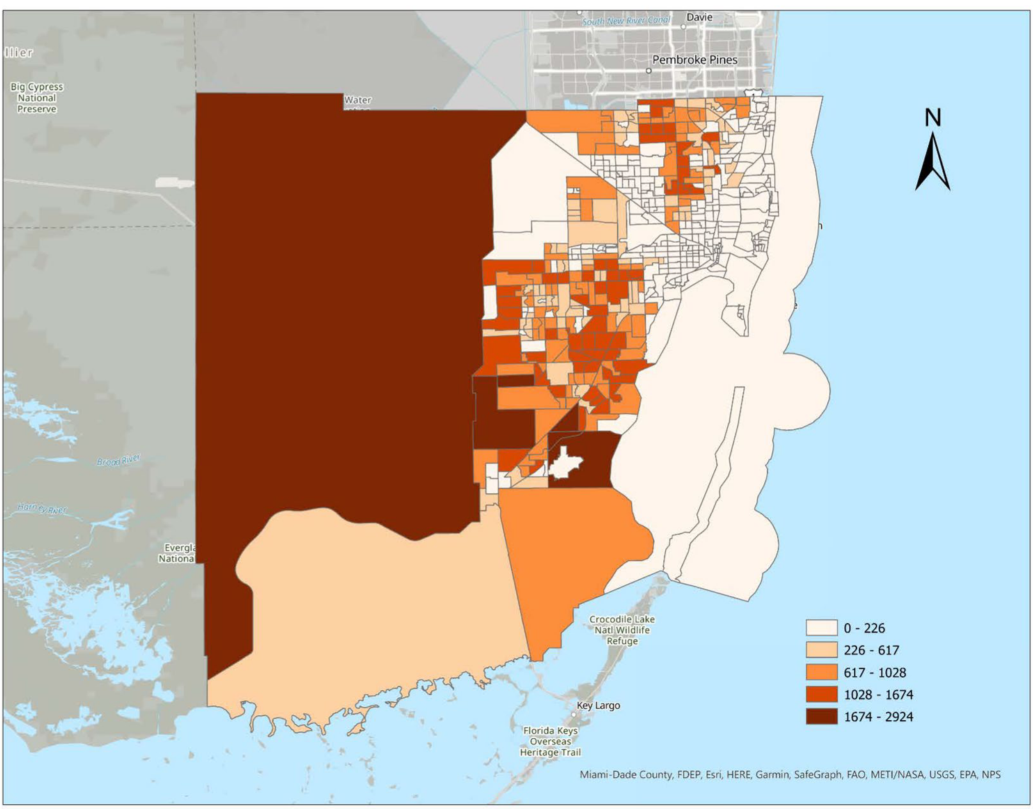

FIGURE 1 Jenks natural break distribution of 311 requests across census tracts

requests are aggregated into the Census tract level. Using machine learning algorithms to predict the numbers of total 311 requests across Census tracts is therefore the main focus of this paper. To examine the determinants of the 311 service requests, the Census Bureau on economic, demographic, employment, and housing information are also added to the 311 data set (for more info, https://geolytics.com/census-data). Understanding and accurately predicting 311 requests across Census tracts are important for public policy makers and professionals to be better prepared and more effectively allocate resources in anticipation of incoming 311 requests across Census tracts.

4 | METHODS AND ALGORITHM BUILDING

4.1 | Spatial autocorrelation and clustering

The spatial unit of analysis is Census tract. Exploratory data analysis on spatial autocorrelation and spatial clustering are carried out with *Geoda*, a free and open source software tool that is designed to explore and model spatial patterns (downloadable at http://geodacenter.github.io/). Global Moran's *I* indicator is first calculated to assess the overall extent of spatial autocorrelation regarding total counts of 311 requests among all Census tracts within the Miami-Dade County. Then, the local indicator of spatial association (LISA), using local Moran's *I*, is calculated in accordance with Anselin (1995). Unlike global Moran's *I* indicator, which assumes spatial homogeneity, LISA explores localized spatial clusters by calculating a local Moran's *I* index for each individual geographic unit and testing each local index's statistical significance.

A LISA cluster map shows the statistically significant locations which are color coded by four types of spatial autocorrelation, that is, high-high, low-low, high-low, and low-high. The first two types are generally referred to as spatial clusters while the latter two as spatial outliers. The spatial clusters are identified when the value at a location (either high or low) is more similar to its neighbors' spatially weighted average value than it would be under spatial randomness. The spatially weighted average values may vary when different spatial interaction relationships are imposed through the creation of different spatial weighting matrixes. On the contrary, spatial outliers signal dissimilarity between a location and its neighbors.

4.2 | Algorithms utilized

For answering the two related research questions, the paper first explores the predictability of total counts of 311 requests by applying and comparing four commonly used machine learning algorithms, namely, random forest (RF), lasso and elastic-net regularized generalized linear models (GLMNET), SVM with radial basis and polynomial functions, respectively (svmRadial and svmPoly), and extreme gradient boosting (XGBoost). Two variants of the SVM algorithm are also used, each of which utilizes a different kernel function to show how the choice of the kernel function may potentially affect predictability. In total, five models are thus employed. These methods are briefly described below.

Random forest (RF) is a tree-based ensemble learning model for classification and regression by randomly constructing a multitude of decision trees (forest). It is flexible, easy to use, and known for its superior performance and robust outcomes. In addition, random forest has an easily understandable predictive mechanisms compared with other robust but "black box" types of algorithms, such as neural network or even the support vector machine (Duro et al., 2012).

Although the generalization error that occurs in a Random Forest depends on the predictive strength of the decision trees in the models, ensemble models are suggested to provide robust models with a high predictive accuracy when compared with individual classifiers (Svetnik et al., 2003). As an ensemble model, random forest relies on aggregated votes of each individual trees and their results. Votes from all trees are pooled, and results with the maximum votes are finally computed. Random forest has been widely and successfully applied in various fields, such as finance, sports, medicine, and corruption across countries (Lima & Delen, 2020; Sharda et al., 2017).

Lasso and elastic-net regularized generalized linear models (GLMNET) are another set of algorithms that are widely used. They are efficient because they can apply entire lasso (short for *least absolute shrinkage and selection operator*) and/or ridge regularization into linear and logistic regression models. Both lasso and ridge techniques are shrinkage tools that impose penalty functions to overcome model overfitting and thus enhance the prediction accuracy (Friedman et al., 2010). The shrinkage effects are particularly significant when input features are highly correlated, because some of the correlated features may be redundant and thus overfit a model. By shrinking the number of (correlated) features and their estimated effects, lasso and ridge regularization methods can produce more robust algorithms with improved predictive accuracy.

SVM was first introduced and expanded by Boser et al. (1992) and Cortes and Vapnik (1995). It can also be used for both classification and regression predictions. SVM transforms original finite-dimensional feature space into a much higher- dimensional space and then construct a hyperplane that maximizes prediction. In SVM, both linear and non-linear kernel functions can be applied to convert original input variables into high dimensional feature spaces. Two kernel functions will be used in this paper and they are Polynomial and Radial Basis functions (svmPoly and svmRadial). They are both commonly used kernel functions in various kernelized learning algorithms (Chang et al., 2010).

Extreme gradient boosting (XGBoost) is a scalable decision tree-based ensemble algorithm that uses a gradient boosting framework. It is flexible and effective but requires minimal amount of computational resources. It was created by Chen and Guestrin (2016), but has been widely used in data science and machine learning, particularly after it successfully won in a series of Kaggle challenges[1] as well as in the annual Data Mining and Knowledge Discovery competition organized by ACM Special Interest Group on Knowledge Discovery and Data Mining (KDDCup). Built upon a gradient boosting framework, XGBoost can combine any singular models, which are often referred to as base learners or boosters, into an ensemble model. Commonly used boosters are linear models (xgbLinear) and tree-based models (xgbTree), though other boosters are readily available, such as splines and radial basis functions (Hastie et al., 2009). Both xgbTree and random forest are tree-based algorithms, but they differ significantly. Random forest builds and trains each tree independently, using a random sample of the data. This randomness helps to make the model more robust than a single decision tree and is also less likely to overfit on the training data. xgbTree builds one tree at a time in a forward stepwise manner, identifies weak-learner trees which have high prediction errors, and improves on existing weak learners.

4.3 | Feature selection

For the second research question, that is, what are the community level determinants that predict service requests in a neighborhood? This paper uses all the aforementioned four families of algorithms (five models) for training and testing the determinants. The 311 service requests data are first aggregated at the Census tract level and then combined with Census Bureau's social, economic, and demographic information. Two feature selection algorithms, namely "Boruta" and "fscaret" are used to identify the features that are the determining factors of the service requests. Feature selection is important because very often researchers and practitioners have high-dimensional data, in which many features are correlated and potentially redundant. Reducing dimensionality and extracting information that is uncorrelated and non-redundant allows to train a model faster because of low(er) complexity of the model, improves model accuracy, and reduces model overfitting.

Boruta is a tree-based algorithm wrapped around the RF algorithm. Boruta duplicates each feature in a data set by reshuffling its values in a column. The duplicated variables with randomly shuffled values are referred to as "shadow values." It trains a RF model using the extended data set (both original and shuffled features) for a predetermined maximum number of iterations (e.g., 500 times). In each iteration, it employs a feature importance measure (by default, Mean Decrease Accuracy) to evaluate if the importance of each original feature is higher (more important) than the highest of the shadow features. The Boruta algorithm stops either when all original features are confirmed as important or rejected as unimportant, or it reaches the maximum number of RF runs. Additional RF runs can be added for any undecided features or the inconclusive features can be decided based on their respective probability of being important or unimportant.

The *fscar*et algorithm is also a wrapper, but it differs from Boruta as it wraps around all the models with the caret package instead of only RF models. It is an ensemble procedure which aggregates variable importance evaluation results of individual, selected regression algorithms, such as RF, SVM, and XGBoost. Szlek (2018) indicated that there are over 100 models in fscaret and technically they are all collectively applied in one model training run. As selected models are built using all input features against the outcome variable, their respective variable importance results would be called upon and then scaled against their prediction errors. An error scaled variable ranking is produced for each of the selected models. Different models are likely to rank feature importance differently and therefore the ensemble approach can draw evidence and support from uncorrelated learning algorithms and generate an aggregated importance ranking list. Unlike Boruta which relies exclusively on RF and thus may favor tree-based models, fscaret is more versatile and provides a more balanced feature selection that is independent of any particular learning algorithm.

4.4 | Algorithm calibration

All five selected algorithms were calibrated with the R package of "caret," which is short for "Classification And REgression Training". It was created in 2005 by Max Kuhn from Pfizer. Caret is transformative because it unifies distinct packages and algorithms and makes the process of training, tuning and evaluating machine learning models consistent and easy. It wraps around more than 200 existing machine learning models and packages,[2] which often vary greatly in syntax and estimated parameters, and provides standardized and comparable model learning and prediction, using a common set of caret functions.

In this paper, the 311 requests data in 2017 are used to train and test the above algorithms and then apply the algorithms for predicting 2018 and 2014 data. The 2017 data set is randomly divided by a 70–30 split, that is, 70% of the original data set is used for training and the remaining 30% for testing. Furthermore, the n-time repeated, k-fold cross-validation method is used. The k-fold method randomly splits the training data into mutually exclusive k number of subsets.

The n-time repeats suggest how many times the k-fold cross validation will be repeated. In this cross-validation method, the $k-1$ folds of the data are used to build the model and the remaining fold is used to test the model. Cross validation is an effective tool for mitigating heterogeneous training and testing subsets and biased results caused by one single random split (Delen et al., 2012). We employed ten rounds ($k = 10$) of cross validations on the training data set, and these ten-fold cross validations are repeated five times ($n = 5$). A total of 50 runs are thus conducted for the training data set for each algorithm. The average model from all 50 rounds is used with the testing data set for evaluating accuracy. This process is repeated for the ensemble algorithms.

4.5 | Algorithm comparison and ensemble prediction

As discussed earlier, each of the five predictive models, namely, rf, glmnet, svmRadial, svmPoly, and xgbLinear, undergo n-repeats k-fold cross-validation during model training and tuning. So, each model has a total of ($n \times k$) sets of trained parameters and model accuracy indicators of regression models (such as, MAE (Mean Absolute Error), RMSE (Root Mean Squared Error), and R-squared (Coefficient of Determination)). Also, for each model, the distribution of a single accuracy indicator is constructed within the 95% confidence interval range. By plotting distributions of the accuracy indicators of all models and comparing them, we can assess their accuracy and show if the differences are statistically significant (e.g., by using a Student's t-test).

After the ensemble algorithms are trained and evaluated, they are combined to predict the outcome variable. The ensemble prediction is more advantageous than individual model predictions because it produces better results by decreasing generalization error (Hastie et al., 2009). Accuracy of ensemble prediction would improve if uncorrelated models are combined. A typical ensemble approach involves computing the arithmetic mean of predictions derived from various component models or a weighted average after (arbitrary) weights are assigned to the different model predictions. In this paper, we have instead relied on "stacking" ensemble multiple models. The stacking approach organizes models into top and bottom layers. Models in the bottom layer make predictions based on original input features; while those in the top layer take predictions of the bottom layer models as their input and predict the final output (Kuhn & Johnson, 2013). Any model can be used for top and bottom layers. In this paper, all five algorithms, RF, GLMNET, svmPoly, svmRadial, and xgbLinear, are used in the bottom layer, and xgbTree is used for the top layer model.

5 | RESULTS AND DISCUSSIONS

5.1 | Results of spatial autocorrelation and clusters

The global Moran's I indicator for total counts of 311 requests in 2017 is 0.589, suggesting a positive spatial autocorrelation across all 519 Census tracts that compose Miami-Dade County. A spatial boxplot is presented in Figure 2 highlighting the outliers and their geographic locations. There are six Census tracts whose numbers of 311 calls exceed three standard deviations from the mean. The map next to the boxplot indicates the location of the six outlier tracts.[3] They are roughly located in the eastern side of the County. The two largest Census tracts along the eastern border of the County cover, respectively, the Francis S. Taylor Wildlife Management Area and the Everglades National Park, both of which belong to the Greater Everglades Ecosystem of tropical wetlands (which are largely uninhabited areas).

FIGURE 2 Spatial box plot (six outliers)

LISA results are presented in the Figure 3 showing spatial clusters and outliers in the numbers of 311 request calls. According to the LISA map, there is a divide between the eastern/middle and western parts of Miami-Dade County. On the eastern and middle parts, there is a high-high pattern, suggesting hot spots of 311 requests. On the contrary, on the other side of the County, there is a low-low pattern, indicating cold spots. Besides the hot and cold spots, the remaining 226 Census tracts do not have statistically significant differences compared with their neighboring tracts in terms of 311 request calls received. This lack of statistical significance suggests limited spatial autocorrelation, that is, the 311 calls are spatially random.

Residuals of 2017 requests after ensemble predictions are shown in a LISA map in Figure 3. It shows that there are 416 Census tracts (80% of all tracts) that do not have statistically significant differences compared with their neighboring tracts in terms of 311 request calls received. This number is considerably greater than 226 tracts before ensemble predictions, thus indicating a great level of spatial autocorrelation has been controlled by ensemble predictions and by input features used in ensemble predictions.

5.2 | Results of feature selection

A total of 48 original features are inputted into the two feature selection algorithms. The set of 48 features include variables capturing social, economic, demographic, housing, employment, and migration characteristics of all Census tracts within the Miami-Dade County. In addition, each Census tract's longitudinal and latitudinal coordinates and its previous years' number of 311 requests are incorporated to control potential spatial and temporal autocorrelation, respectively.

The Boruta algorithm generated 28 confirmed features (Table 1 and Figure 4), based on the comparison between a feature and a "shadow value," which is a manufactured feature of reshuffled values of the original features. As discussed before, Boruta feature selection relies exclusively on RF. The fscaret algorithm generated six features based on their relative variable importance. When comparing these two sets of selected features (Table 1), it is evident that Boruta-informed features are much more extensive and includes all the features generated by fscaret. Features informed by the fscaret algorithm center on housing and locational characteristics while those informed by Boruta span income, employment, demographic, and employment factors.

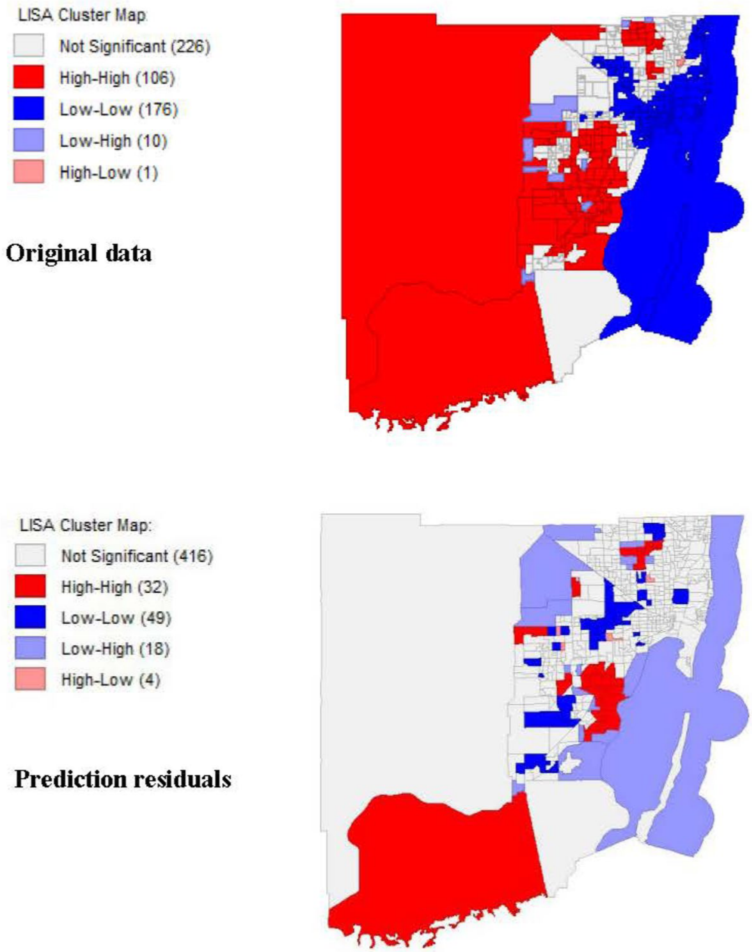

FIGURE 3 LISA map of 311 requests in 2017 across census tracts

Both sets of selected features are calibrated with the various algorithms for predicting the numbers of 311 requests in 2017 and subsequent years. The efficiency and effectiveness of the two types of feature selection are compared according to model predictability using two different sets of selected variables. For any algorithm, the set of selected features that gives lower prediction errors are considered to be more efficient than those compared with the other set of variables. If, however, two sets of selected features generate the same or similar predictive accuracy, the set with fewer selected features will be arguably more efficient because it is parsimonious. In addition, because multiple learning algorithms are applied in this paper, average predictive accuracy of all models will be used to assess efficiency of the two sets of selected features.

5.3 | Results of predictive accuracy

For all five algorithms, namely, RF, GLMNET, svmPoly, svmRadial, and xgbLinear, that are being trained and tested in the paper, Table 2 presents their comparative powers of predictive accuracy. The top section of the table presents accuracy of models using features selected by

TABLE 1 Comparison of selected features informed by two algorithms

Boruta informed features, 28 in total	fscaret informed features, six in total
Longitudinal coordinate of tract centroid	Longitudinal coordinate of tract centroid
Latitudinal coordinate of tract centroid	Latitudinal coordinate of tract centroid
Lagged request count of a tract	Lagged request count of a tract
# of owner-occupied housing units	# of owner-occupied housing units
# of detached single housing units	# of detached single housing units
# of employment in government	# of employment in government
# of total housing units	
# of vacant housing units	
# of rental housing units	
Median household income	
Median housing value	
Total population	
Female population	
Male population	
White population	
Black population	
Hispanic population	
Median age	
Male population in labor force	
Female population in labor force	
Self-employment population	
Population in poverty	
Married population	
Area of tract	
Growth rate of housing units	
Growth rate of detached single housing units	
Growth rate of vacant housing units	
Growth rate of employment in government	

the fscaret algorithm, while the bottom section shows accuracy of models using Boruta selected features. The table depicts that all models produce comparatively similar predictive accuracy regardless of which feature set is used. For example, for the 2017 test data set, xgbLinear generate a *R*-squared value of 0.946 using fscaret selected features and a value of 0.942 with Boruta selected features. However, because the number of Boruta selected features (28) is much greater than that of fscaret feature set (6), models using the fscaret feature set are more parsimonious and therefore more efficient in providing predictive accuracy. In addition, for either of the selected feature sets, the ensemble prediction accuracy which relies on a combination of the five individual algorithms does not differ significantly from that of well-performing individual models. For example, with the fscaret feature set, the ensemble prediction registers a *R*-squared value of 0.950, while RF and xgbLinear, respectively, have a value of 0.944 and 0.946.

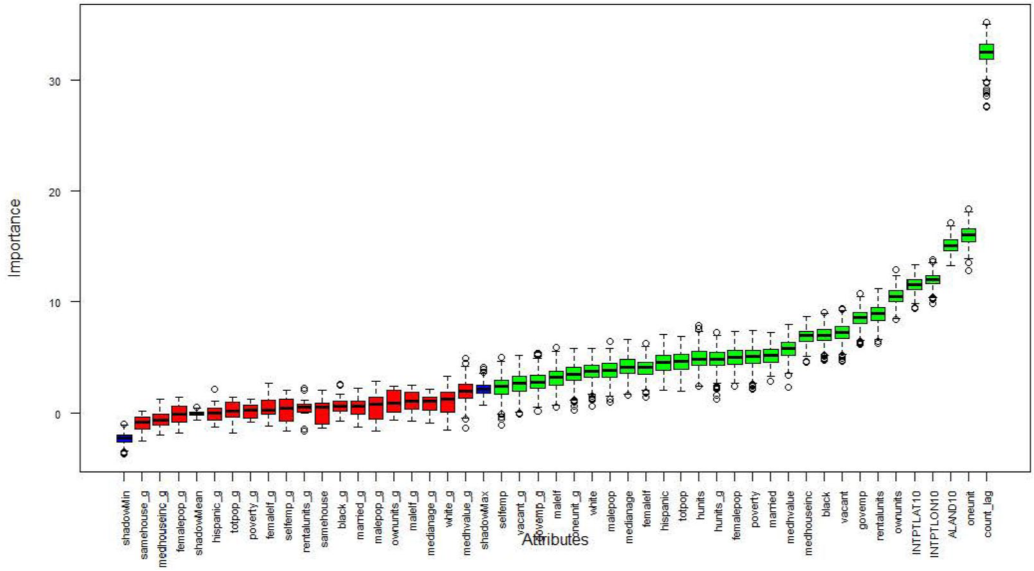

FIGURE 4 Feature selection results of Boruta algorithm. Features (red) that are not confirmed or selected by the Boruta algorithm consist of growth rate of % of residents stayed in the same houses, growth rate of median household income, growth rate of female population, growth rate of Hispanic population, growth rate of total population, growth rate of population in poverty, growth rate of female population in labor force, growth rate of population in self-employment, growth rate of rental housing units, number of residents staying in the same houses, growth rate of married population, growth rate of male population, growth rate of owner-occupied housing units, growth rate of male population in labor force, growth rate of median age, growth rate of white population, and growth rate of median housing value

TABLE 2 Predictive power (R^2) of algorithms

	rf	glmnet	svmRadial	svmPoly	xgbLinear	Ensemble
fscaret feature selection						
2017 test dataset	0.944	0.945	0.863	0.949	0.946	0.950
2018 full dataset	0.953	0.945	0.900	0.939	0.940	0.938
2014 full dataset	0.949	0.942	0.900	0.938	0.933	0.941
Boruta feature selection						
2017 test dataset	0.943	0.947	0.830	0.830	0.942	0.948
2018 full dataset	0.955	0.930	0.885	0.885	0.947	0.936
2014 full dataset	0.951	0.940	0.861	0.861	0.934	0.944

Across all five algorithms, as shown in Figures 5 and 6, GLMNET and svmPoly have the highest prediction accuracy, which are followed by xgbLinear and RF. The model of svmRadial has the lowest prediction accuracy. In addition, based on 50 repeated cross-validation samples (10 cross validations and five repeats), GLMNET and svmPoly also have the smallest 95% confidence intervals, followed by xgbLinear and RF. svmRadial does not only have the lowest accuracy, but also has the widest confidence interval, suggesting that its prediction is widely spread out and is thus unreliable.

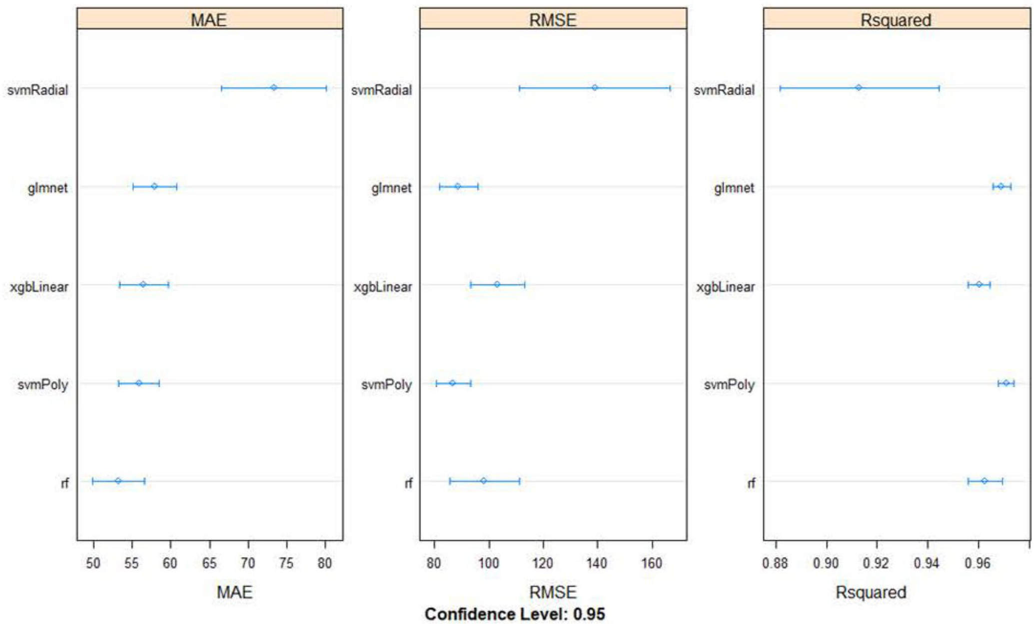

FIGURE 5 Algorithm comparison based on fscaret selected features ($n = 50$, six features)

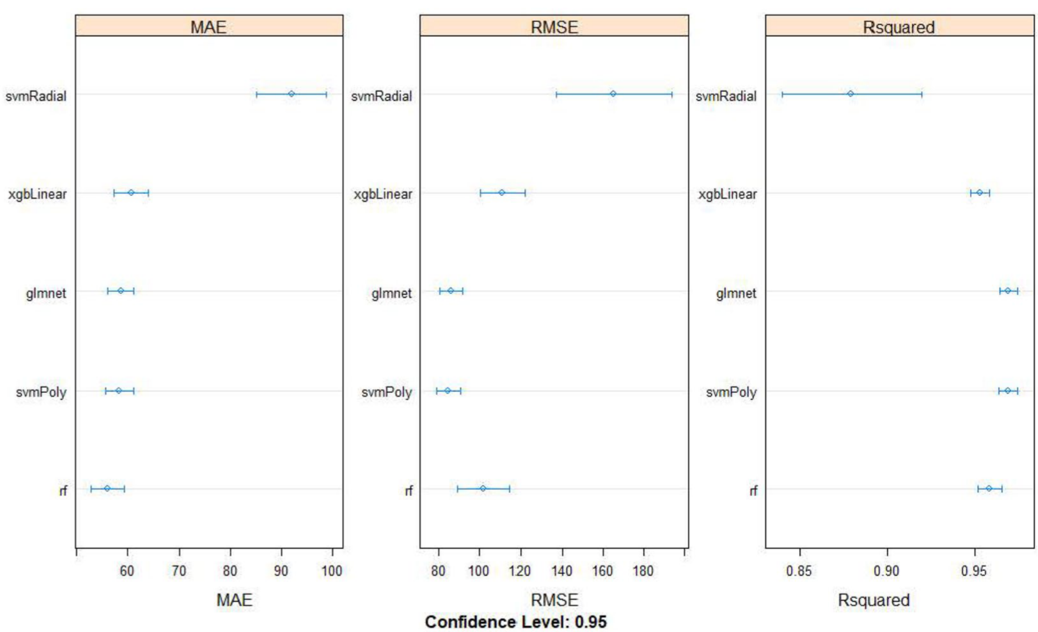

FIGURE 6 Algorithm comparison based on Boruta selected features ($n = 50$, 28 features)

All five trained models and the ensemble model with year 2017 data are applied with year 2018 and 2014 data. The purpose is to test if the trained models can predict counts and patterns of 311 requests with new and unseen data in 2018 and 2014. Results of this exercise is important because accurate prediction of the future with existing data is the essential for allocation of resources based on 311 service requests. Table 2 presents that prediction accuracy of the 2018 and 2014 data is consistent with that of the 2017 training data set. Models using either feature set selected by Boruta or fscaret are consistent in predictive accuracy, except two SVM models, svmPoly and svmRadial, which have considerably lower accuracy than other models. The accuracy of ensemble prediction is comparable with other individual models.

6 | CONCLUSIONS AND POLICY IMPLICATIONS

This paper illustrated the application of machine learning algorithms to administrative data for predictions. Local government administrators, policy makers, and researchers can analyze and predict 311 requests based on community characteristics and changes. 311 requests are unique to understand resident participation, service delivery and governmental efficiency and equity, because 311 requests allow residents to request information, provide assistance in public service delivery, and interact with public agencies on service variety and output. The predictive capacity could enable local governments to proactively allocate resources, identify service disparity, and achieve service efficiency and equity across communities and residents.

Four families of machine learning algorithms to predict 311 service requests have been examined in this paper: random forest, SVM, lasso and elastic-net regularized generalized linear models, and extreme gradient boosting. Their respective predictive accuracy is then compared with that of an ensemble prediction which aggregates outcome of each individual algorithm. For the test data set of 2017 311 requests, the ensemble approach had the highest predictive accuracy. But for 311 requests of 2018 and 2014, the random forest method had the highest predictive accuracy.

The algorithms were able to capture spatial autocorrelation and clustering. The spatial analysis of service requests in 2017 shows the global Moran's I test to be 0.589. About half of all 519 tracts in Miami-Dade are statistically significant in the LISA test. The analysis of residuals after the ensemble prediction, however, suggests that 416 Census tracts (80% of all tracts) become statistically insignificant in the LISA test. Incorporating each Census tract's longitudinal and latitudinal coordinates is thus effective for accounting spatial autocorrelation and clustering in machine learning predictive algorithms.

Two feature selection algorithms were applied in this paper, namely Boruta and fscaret. They generated distinct numbers of selected features—Boruta generated 28 features while fscaret generated six features. While Boruta-informed selected features span over locational, economic, demographic, employment, and housing factors, features selected by fscaret center on locational and housing characteristics. Both sets of selected features were used in the paper with various learning algorithms to examine their comparative predictive accuracy. Features selected by the fscaret algorithm produce more parsimonious models and are arguably better in providing predictive accuracy.

The predictive algorithms used in this paper is a methodological stepping stone which can be further expanded to examine community changes in shorter time intervals, such as quarterly, monthly, or weekly. Insights gained from the application of algorithms to the administrative data can provide objective performance indicators for various public services. In the context of 311, the algorithms provide predictive insights into the volume of service requests, and which features influence the volume of service requests. Examining the features can shed light on structural or systemic deficiencies in service provision. Residents and communities can hold government agencies accountable by examining the patterns of the 311 calls. The machine learning algorithms are new tools in this accountability process. They extend the traditional methods of regression analysis to identify parsimonious features that predict the volume of 311 requests. These tools are likely to play an increasingly important role in analyzing administrative data and government provision of services.

The paper also holds at least two policy insights from the perspective of 311 customer centers. The first insight is that fscaret generated parsimonious categories for predicting the factors influencing the volume of 311 service requests. The fscaret features are centered on locational and housing attributes. These fscaret-based attributes predict as much as, if not higher than the features identified through Boruta, which identified five times as many features. The results show a lack of importance of economic and demographic variables on predictive accuracy of 311 service requests. This could signal to policy makers that socioeconomic and demographic differences across communities do not correlate with neighborhood variations of 311 requests. This is preliminary evidence that there is no significant disparity in engagement with and participation in coproduction with local governments on public goods and services, across diverse communities, and demographic groups, especially those in disadvantaged socioeconomic status. In other words, fscaret-informed features and resulting predictive accuracy could suggest that there is equitable participation in 311 requests across diverse demographic and socioeconomic status and background.

The second policy insight is that the predictive algorithms could enable local governments to proactively allocate resources, identify service disparity, and achieve service efficiency and equity across communities and residents. The algorithms build upon commonly available community characteristics from the Census Bureau. They take into consideration spatial patterns and clustering, which could obviate the need for separate examination of spatial autocorrelation. The predictive capacity enables local governments to be pro-active in allocating resources to so that 311 requests can be fulfilled in an efficient manner across all neighborhoods. The 311 customer centers can thus proactively improve the efficiency and effectiveness of public services.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available at Miami-Dade County's Open Data Hub, https://gis-mdc.opendata.arcgis.com/. 311 service requests of the Miami-Dade County are publicly available at https://gis-mdc.opendata.arcgis.com/datasets/311-service-requests-miami-dade-county-2017.

ORCID

Shaoming Cheng https://orcid.org/0000-0003-0768-8373

ENDNOTES

[1] Kaggle, a subsidiary of Google LLC, is an online community of data scientists and machine learning practitioners. For more info, https://www.kaggle.com/competitions.

[2] Max Kuhn's caret package page, https://topepo.github.io/caret/available-models.html, indicates there are 238 models available.

[3] GeoDa is particularly helpful because of its visualization functionality of "brushing" across different graphs and charts, allowing users to view the data simultaneously from distinct perspectives. In this specific example, the six outliers are selected and highlighted in the boxplot and they are highlighted automatically, thanks to "brushing," in the GIS map window.

REFERENCES

Anselin, L. (1995). Local indicators of spatial association-LISA. *Geographical Analysis*, *27*(2), 93–115. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x

Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. *Proceedings of the Fifth Annual Workshop on Computational Learning Theory – COLT*, *'92*, 144. https://doi.org/10.1145/130385.130401

Chang, Y., Hsieh, C., Chang, K., Ringgaard, M., & Lin, C. (2010). Training and testing low-degree polynomial data mappings via linear SVM. *Journal of Machine Learning Research*, *11*, 1471–1490.

Chatfield, A. T., & Reddick, C. G. (2018). Customer agility and responsiveness through big data analytics for public value creation: A case study of Houston 311 on-demand services. *Government Information Quarterly*, *35*(2), 336–347. https://doi.org/10.1016/j.giq.2017.11.002

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. https://doi.org/10.1145/2939672.2939785

Clark, B. Y., Brudney, J. L., Jang, S., & Davy, B. (2020). Do advanced information technologies produce equitable government responses in coproduction: An examination of 311 systems in 15 US cities. *American Review of Public Administration*, *50*(3), 315–327. https://doi.org/10.1177/0275074019894564

Connelly, R., Playford, C., Gayle, V., & Dibben, C. (2016). The role of administrative data in the big data revolution in social science research. *Social Science Research*, *59*, 1–12. https://doi.org/10.1016/j.ssresearch.2016.04.015

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297. https://doi.org/10.1007/BF00994018

Delen, D., Cogdell, D., & Kasap, N. (2012). A comparative analysis of data mining methods in predicting NCAA bowl outcomes. *International Journal of Forecasting*, *28*, 543–552. https://doi.org/10.1016/j.ijforecast.2011.05.002

Descant, S. (2020). Cities reimagine 311 service as pandemic shifts the paradigm. *Government Technology*. https://www.govtech.com/analytics/Cities-Reimagine-311-Service-as-Pandemic-Shifts-the-Paradigm.html

Duro, D. C., Franklin, S. E., & Dubé, M. G. (2012). A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. *Remote Sensing of Environment*, *118*, 259–272. https://doi.org/10.1016/j.rse.2011.11.020

Einav, L., & Levin, J. D. (2013). The data revolution and economic analysis. *National Bureau of Economic Research*. https://www.nber.org/papers/w19035

Elias, P. (2014). Administrative data. In A. Dusa, D. Nelle, G. Stock, & G. Wagner (Eds.), *Facing the future: European research infrastructures for the humanities and social sciences* (pp. 47–48). Scivero.

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, *33*(1), 1–22. https://doi.org/10.18637/jss.v033.i01

Ganapati, S., & Reddick, C. G. (2012). Open e-government in U.S. state governments: Survey evidence from chief information officers. *Government Information Quarterly*, *29*, 115–122. https://doi.org/10.1016/j.giq.2011.09.006

Ganapati, S., & Reddick, C. G. (2014). The use of ICT for open government in U.S. municipalities. *Public Performance & Management Review*, *37*(3), 365–387. https://doi.org/10.2753/PMR1530-9576370302

Ganapati, S., & Scutelnicu, G. (2015). Open innovations in the public sector: Open 311. In P. Julnes, & E. Gibson (Eds.), *Innovations in the public and nonprofit sectors* (pp. 74–90). Routledge/Taylor and Francis.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning*. Springer.

Jakobsen, M., & Andersen, S. C. (2013). Coproduction and equity in public service delivery. *Public Administration Review*, *73*(5), 704–713. https://doi.org/10.1111/puar.12094

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. Springer.

Levine, C. H., & Fisher, G. (1984). Citizenship and service delivery: The promise of coproduction. *Public Administration Review*, *44*, 178–189. https://doi.org/10.2307/975559

Lima, M., & Delen, D. (2020). Predicting and explaining corruption across countries: A machine learning approach. *Government Information Quarterly*, *37*, 1–15. https://doi.org/10.1016/j.giq.2019.101407

Meijer, A. (2011). Networked coproduction of public services in virtual communities: From a government-centric to a community approach to public service support. *Public Administration Review*, *71*(4), 598–607. https://doi.org/10.1111/j.1540-6210.2011.02391.x

Nabatchi, T., Sancino, A., & Sicilia, M. (2017). Varieties of participation in public services: The who, when, and what of coproduction. *Public Administration Review*, *77*(5), 766–776.

Ostrom, E. (1996). Crossing the great divide: Coproduction, synergy, and development. *World Development*, *24*(6), 1073–1087. https://doi.org/10.1016/0305-750X(96)00023-X

Pak, B., Chua, A., & Moere, A. V. (2017). FixMyStreet Brussels: Socio-demographic inequality in crowdsourced civic participation. *Journal of Urban Technology*, *24*(2), 65–87.

Sharda, R., Delen, D., & Turban, E. (2017). *Business intelligence, analytics, and data science: A managerial perspective* (4th ed.), Pearson.

Svetnik, G., Liaw, A., Tong, C., Culberson, J. C., Sheridan, R. P., & Feuston, B. P. (2003). Random forest: A classification and regression tool for compound classification and QSAR modeling. *Journal of Chemical Information and Computer Sciences*, *43*, 1947–1958.

Szlek, J. (2018). *A short fscaret package introduction with examples*. Retrieved on August 10, 2020 from https://cran.r-project.org/web/packages/fscaret/vignettes/fscaret.pdf

Thomas, J. C., & Streib, G. (2013). The new face of government: Citizen-initiated contacts in the era of E-government. *Journal of Public Administration Research and Theory*, *13*(1), 83–102.

Wang, D., Chen, C., & Richards, D. (2018). A prioritization-based analysis of local open government data portals: A case study of Chinese province-level governments. *Government Information Quarterly*, *35*(4), 644–656. https://doi.org/10.1016/j.giq.2018.10.006

Wang, V., & Shepherd, D. (2020). Exploring the extent of openness of open government data: A critique of open government datasets in the UK. *Government Information Quarterly*, *37*(1), 1–10. https://doi.org/10.1016/j.giq.2019.101405

Whitaker, G. P. (1980). Coproduction: Citizen participation in service delivery. *Public Administration Review*, *40*(3), 240–246. https://doi.org/10.2307/975377

Xu, C. K., & Tang, T. (2020). Closing the gap or widening the divide: The impacts of technology-enabled coproduction on equity in public service delivery. *Public Administration Review*, *80*(6), 962–975.

How to cite this article: Cheng, S., Ganapati, S., Narasimhan, G., & Yusuf, F. B. (2021). A machine learning-based analysis of 311 requests in the Miami-Dade County. *Growth and Change*, 00, 1–19. https://doi.org/10.1111/grow.12578