**REGULAR PAPER**

# Urban fire station location planning using predicted demand and service quality index

**Arnab Dey**[1] (ID) · **Andrew Heger**[2] · **Darin England**[3]

## Abstract

In this article, we propose a systematic approach for fire station location planning. We develop machine learning models, based on Random Forest and Extreme Gradient Boosting, for demand prediction and utilize the models further to define a generalized index to measure quality of fire service in urban settings. Our model is built upon spatial data collected from multiple different sources. Efficacy of proper facility planning depends on choice of candidates where fire stations can be located along with existing stations, if any. Also, the travel time from these candidates to demand locations need to be taken care of to maintain fire safety standard. Here, we propose a travel time-based clustering technique to identify suitable candidates. Finally, we develop an optimization problem to select best locations to install new fire stations. Our optimization problem is built upon maximum coverage problem, based on integer programming. We further develop a two-stage stochastic optimization model to characterize the confidence in our decision outcome. We present a detailed experimental study of our proposed approach in collaboration with city of Victoria Fire Department, MN, USA. Our demand prediction model achieves true positive rate of 80% and false positive rate of 20% approximately. We aid Victoria Fire Department to select a location for a new fire station using our approach. We present detailed results on improvement statistics by locating a new facility, as suggested by our methodology, in the city of Victoria.

**Keywords** Extreme gradient boosting · Facility planning · Fire risk · Optimization · Random forest · Risk prediction

## 1 Introduction

Urban fires adversely affect the socio-economic growth and ecosystem health of any community. In 2019, the National Fire Protection Association (NFPA) reported an estimated loss of $14.8 billion and 3700 civilian fire deaths in the U.S. [1]. According to the report, on average, a fire department responded to a fire incident every 24 s and in total, local fire departments responded to an estimated 1.3 million fires. Thus, building a safe community equipped with adequate fire safety measures is one of the most essential and challenging tasks of any fire department and local government. However, it is very difficult to quantify the measure of safety, which in turn results in complexity in determining if additional fire stations are required to be installed. This decision must be substantiated with all necessary spatio-temporal data, gathered from various different sources, such as accurate details of historical emergency incidents from the fire department, future patterns of land use from local government, parcel information from geographic information system (GIS) (maintained by both US Census Bureau and local authorities) and demographic information from the US Census Bureau. Combining all of this information together is a demanding task, due to differences in storage architectures across different sources. For example, the address of a real property is documented with different abbreviations and acronyms in different data sources, which makes the

✉ Arnab Dey
dey00011@umn.edu

Andrew Heger
aheger@ci.victoria.mn.us

Darin England
engl0124@umn.edu

1 Electrical and Computer Engineering, University of Minnesota, Twin Cities, Minneapolis, MN 55455, USA

2 Fire Chief, City of Victoria Fire Department, Victoria, MN 55386, USA

3 Industrial and Systems Engineering, University of Minnesota, Twin Cities, Minneapolis, MN 55455, USA

task of joining the data sets very difficult. Also, the accuracy of the information on historical fire incidents is poor and exacerbated by incompleteness. A few examples include address mismatches, missing dates, and missing timestamps. Further, even if the requirement can be empirically justified using historical spatio-temporal data, the determination of locations for new facilities is a demanding task that must be strategically chosen based on the performance of the existing facilities as well as future community growth and fire risk projections. Such prediction of future service demand is very difficult due to its dependency on many different underlying factors. Moreover, a correlation between the predictive features makes the task even more challenging. We emphasize that quantifying public safety measures, and embedding them to devise a systematic approach to locate new fire stations is limited in the existing literature. Moreover, the efficacy of such approaches from a data-mining perspective, including both confidential and publicly available datasets, is unexplored in existing works.

To this end, we provide an approach that utilizes both temporal and spatial data to select locations for future fire stations in urban settings. Our methodology uses a Random Forest [2] and Extreme Gradient Boosting [3] machine learning models and a Density-based spatial clustering of applications with noise (DBSCAN) algorithm [4]. For our experimental case study, we partnered with the City of Victoria Fire Department (VFD), MN, USA. We apply our proposed methodology to select the location of a future fire station for VFD. Note that our approach is not specific to VFD and can be used to analyze candidate locations for any urban area. The main contributions of this article are summarized below:

1. *Predictive demand model and service quality assessment*: We propose two machine learning (ML) models, based on Random Forest (RF) and Extreme Gradient Boosting (XGBoost), to predict future service demand of urban areas from spatial data. We present a detailed performance comparison between the models to choose the superior one. The chosen ML model is utilized to define a generalized service quality index ($SQI$) that measures the quality of service provided by existing fire stations at any demand point. We propose to use the SQI as a new national fire service standard to assess the performance of any fire facility, which can be tailored to a specific locality. We also present a travel time-based DBSCAN algorithm to identify candidate locations for new facilities. Our proposed methodology can be used to predict and analyze the performance of any fire department to improve the quality of service.
2. *Optimization model for facility location selection*: We propose an optimization model, based on $SQI$ estimates, to select future fire stations. Our optimization model accounts for spatial coverage, service prioritization, and

service redundancy, thus presenting an end-to-end strategic approach for fire station location selection. Further, we propose a two-stage stochastic optimization model utilizing our demand prediction model and travel-time base DBSCAN algorithm. While the SQI-based optimization model establishes the utility of our proposed SQI to be used as a new standardized index, the stochastic optimization model establishes confidence in the decision outcome of our proposed models, thus providing a comprehensive overview of the merits of fire station location selection methodology.
3. *Impact on the city of Victoria Fire Department*: The work is done in collaboration with Victoria Fire Department (VFD). VFD utilizes the detailed performance analysis of the existing fire station to properly allocate resources to improve the quality of service. It can also help VFD to identify properties and locations which require special attention in terms of fire safety measures. The predictive models and the facility location study are used by VFD for crafting the initial proposal for the new fire station to be approved by the local authorities.

## 2 Related work

Systematic approaches for selecting sites for critical infrastructure have been widely studied in the literature [5]. In [6], a bi-objective spatial optimization model is proposed to locate an urban fire station by minimizing the number of stations to be sited along with their distances from the service demand clusters to ensure a required area coverage. A similar approach is applied in [7] to prove the efficacy of such algorithms in guiding local government authorities to select a location for a new fire station. A multi-criteria decision-making approach for the planning of emergency facilities, tuned with GIS information, is proposed in [8]. Site selection based on the spatial efficiency of existing and future facilities is studied extensively in many prior works [9–11]. A GIS-based approach for emergency facility location is the primary focus of [12,13]. Various optimization models to solve facility location problems, both in the context of cost and service coverage optimizations, are widely studied in the literature [14–16]. In [17], an optimization methodology based on response time of the fire stations is proposed to select new facilities.

However, when there exists a constraint on the maximum number of facilities that can be sited, determining the weights on multiple options in decision-making can be challenging. Here, prioritization of service demand locations based on future risk prediction can be helpful [18]. Fire risk prediction has been a primary focus of many researchers; however, this work has mostly pertained to forest and wild land fires [19–23]. [24] presents an extensive study on the effect of climate

on fire prediction along with an analysis of the impact of different seasons on burned area from global perspective. In [25], the authors propose a neural network model to predict forest fires with spatio-temporal data.

Urban fire risk prediction has received little attention despite the critical nature of such applications. [26] provides a data-driven approach to predict commercial and residential building fires. In [27], a Random Forest algorithm is used to generate predicted fire risk scores to aid the local fire department in identifying properties that require fire inspections. Such data-driven predictive models are widely used in many prior works [28–31]. [32] proposes an interesting approach to the prediction of fire incidents that is based on a deep neural network. Here, the authors use spatio-temporal data from GIS to learn a generative model. However, the application of these models to locate new facilities has not been extensively studied. Moreover, other critical aspects of fire rescue service, such as response time [33], need to be combined with prediction models. In summary, a unified approach, to select new fire station locations in urban settings, that considers fire service demand predictions and temporal variables with proper characterizations of service priorities and redundancies is limited in the literature.

To this end, we propose an end-to-end fire station location selection methodology that respects predicted demand as well as travel time. We propose a Random Forest (RF) model to predict demand for urban fire and emergency services demand and combine it with travel time information to assess service quality. We quantify the quality of service using our proposed *Service Quality Index* and further utilize it to identify candidate locations for additional fire stations. Finally, an optimization model is presented to select the best option(s) among the candidate locations. In summary, we devise a systematic approach for locating a new fire station using GIS and temporal data. Our demand prediction model is inspired by [26,27]; however, instead of restricting the model to residential and commercial buildings, we consider all types of properties present in a typical city. We apply our proposed methodology to aid VFD with site selection for the new fire station in the city of Victoria. Historical fire incident data is provided by VFD which is augmented with GIS and other demographic data collected from multiple different sources. We substantiate the efficacy of our method with extensive experimental results.

The rest of the paper is organized as follows: The proposed SQI and the demand prediction models are explained in Sect. 3. Section 4 describes the algorithm to choose suitable candidate locations for new station installation. Two optimization problem formulations, to select new locations from the candidates, are introduced in Sect. 5. The efficacy of our proposed approach is illustrated with a detailed case study in Sect. 6. Section 7 presents the concluding remarks.

# 3 Service quality assessment

The service quality of existing fire stations is an important indicator that influences the choice of candidate locations for future facilities. However, in the context of urban fire safety, there is no standard approach to assess the service quality that considers demographic factors such as age of residents, income, property types, and population density. Apart from demographic factors, another important aspect to measure fire safety is travel time from the responding station to the incident location. A longer travel time can result in significant damage or even loss of life. In this context, we propose a generalized service quality index for each property within the service area of a fire station, considering both demographic factors and travel time, as described in the following sections.

## 3.1 Service quality index

Let $\mathcal{J}$ denote the set of demand locations (properties) and let $\mathcal{I}$ denote the set of fire stations within the city boundary. We define the Service Quality Index ($SQI$) of a demand location $j \in \mathcal{J}$ with respect to a fire station $i \in \mathcal{I}$ as follows:

$$SQI_{ji} := P(j)\hat{T}(j,i), \tag{1}$$

where $P(j) \in [0,1]$ denotes the probability of demand request at a location $j$ and $\hat{T}(j,i)$ denotes the normalized travel time from fire station $i$ to the demand location $j$. The normalized travel time can be calculated as follows:

$$\hat{T}(j,i) = \frac{T_{\text{actual}}(j,i)}{T_{\text{norm}}}, \tag{2}$$

where $T_{\text{norm}}$ is a normalization factor to ensure $\hat{T}(j,i) \in [0,1]$ and $T_{\text{actual}}(j,i)$ is the actual travel time from fire station $i$ to demand location $j$ which can be calculated using GIS. A lower value of $SQI_{ji}$ indicates better service. Note that, $SQI_{ji}$ depends on the probability of demand request $P(j)$ of property location $j$, which usually depends on factors like local population density, property type (residential, business etc.), property age, and demographic factors such as residents' age and median income. Moreover, $SQI_{ji}$ also considers the response time for emergency personnel to reach a location, a crucial factor in quality of service [34].

**Remark 1** In the trivial case of no existing fire station, $\hat{T}(j,i)$ is undefined. Here, one can take $SQI_{ji}$ values to be equal to $P(j)$ and the rest of the analysis follows.

**Remark 2** The trade-off between low demand probability at a location with a long travel time from $i$ and high demand probability at a location with a short travel time from $i$ ensures a balance between area coverage and emphasis on demand request. However, the interpretation of better (or

worse) service quality can be tuned per the facility planners' requirements as we describe next.

The availability of multiple fire stations ideally enhances the service quality at a location. However, it is very difficult to characterize, quantify, and predict the availability status of all the stations at the time of demand request. Also, note that the probability of a demand request usually depends on demographic and property characteristics, and hence is independent of the location of fire stations. Therefore, we factor out the dependency on a particular fire station by taking the minimum of $SQI_{ji}$ values over all fire stations $i \in \mathcal{I}$. Therefore, the Service Quality Index of a property $j$ can be written as

$$SQI(j) = \min_{i \in \mathcal{I}} SQI_{ji}. \tag{3}$$

The interpretation of better/worse service quality can be done by categorizing the demand locations into low, medium, and high quality locations based on their $SQI$ values. Let the sets $\mathcal{J}_l, \mathcal{J}_m, \mathcal{J}_h \subseteq \mathcal{J}$ denote the set of indices of the demand locations with low, medium and high quality of service which are defined as follows:

$$
\begin{aligned}
&j \in \mathcal{J}_h, \ \text{if } 0 \le SQI(j) < \tau_l, \\
&j \in \mathcal{J}_m, \ \text{if } \tau_l \le SQI(j) < \tau_h, \\
&j \in \mathcal{J}_l, \ \text{if } \tau_h \le SQI(j) \le 1,
\end{aligned}
\tag{4}
$$

where $\tau_l, \tau_h$ are user-defined constants in $\mathbb{R}$ with $0 < \tau_l < \tau_h \le 1$. Note that, the choice of $\tau_l$ and $\tau_h$ provides an interpretation of fire service quality at a location based on planning constraints. For example, one can calculate the values of $\tau_l, \tau_h$ based on maximum allowable travel time (which depends on the type of fire station [34]) and an acceptable demand request probability that is to be considered as high demand. Thus, our objective is to identify candidate locations to install additional fire stations such that the $SQI$ values of $j \in \mathcal{J}_l$ are improved.

## 3.2 Demand prediction model

Calculation of $SQI$ requires a fire service demand prediction model. In addition to the ability to accurately predict the probability of a service request for a future emergency event, it is also important to gain insights into the predictors and their effects on service demand. Such information can be utilized for resource planning and staff management. Here, tree-based models are suitable as they can be used to quantitatively assess the importance and the effect of the predictors. We propose to use Random Forest (RF) [2] and Extreme Gradient Boosting (XGBoost) [3] algorithms, along with a detailed performance comparison, to predict the probability

**Table 1** Categorization of properties

| Category | Included property types |
|---|---|
| Residential | $1 - 3$ units residential properties, apartments with $4+$ units |
| Commercial | Warehouses, retail outlets, factories etc. |
| Institution | Churches, schools, Govt. offices, police stations etc. |
| Park | Parks, lakes, wetlands, agricultural lands, golf courses, vacant lands etc. |

of a demand request $P(j)$ for each property $j \in \mathcal{J}$. In building the prediction models, we consider various demographic factors and property features as described in Table 2. Note that service demand prediction, with an objective of fire station location selection, is a static analysis. Unlike short-term demand prediction models based on online learning algorithms [35] in a dynamic environment, our model requires long-term demand prediction. Thus we include only spatial factors in our prediction models. The temporal factors, governing the dynamic nature of service demand, are analyzed in [36] to aid VFD in resource planning and staff management. The properties included in our analysis are categorized into four categories based on their usage as shown in Table 1. For our case study, the types of land use are obtained from [37].

The predictor and response variables of our prediction model are described in Table 2.

**Remark 3** The age of a property is calculated from the year it is built and the year when an incident is reported. For example, if a property built in 2010 reports a fire event in 2015, then 'Age'= 5. Multiple events at the same location at different points in time are treated separately. For the properties with no reported incidents, the age is calculated with reference to the year of writing the article (year 2021).

Once the prediction models are built, we obtain the demand request probabilities $P(j)$ using the class probability estimates of the samples corresponding to the positive labeled class (class of samples with response variables labeled as 1; Refer to Table 2). We present the detailed performance measures of our prediction models and variable importance in Sect. 6.2 for our case study.

Upon finding the demand request probabilities, it remains to compute travel time from a fire station to a property in order to completely characterize $SQI$ values for each property. We calculate travel time using the Open Street Map-Based Routing Service (OSRM) package [38] in R [39] with traffic flow disabled to account for the relaxed traffic rules for fire service vehicles. Once the $SQI$ values are calculated, we need to identify suitable geographic locations for additional fire stations. In the next section, we discuss our proposed methodology to select such candidate locations.

**Table 2** Demand prediction models feature description (Variable type notation: P ≡Predictor, R ≡ Response)

| Variable name | Type | Range | Description |
|---|---|---|---|
| Land value | P | $\mathbb{R}^+$ | Estimated land value ($\times \$10,000$). |
| Land size | P | $\mathbb{R}^+$ | Land size (acres). |
| Num units | P | $\mathbb{N}$ | Number of units in a property. |
| Prop. Age | P | $\mathbb{N}$ | Age of property at time of incident. |
| Resi. Age | P | $\mathbb{R}^+$ | Residents' Median age in the block where the property is located. |
| Population | P | $\mathbb{R}^+$ | Average population of the block where property is located. |
| Prop. Type | P | 0 − 3 (factor) | 0 ≡ residential, 1 ≡ commercial, 2 ≡ institution, 3 ≡ park |
| Demand | R | 0 − 1 (binary) | 1 if any incident occurred, 0 otherwise |

## 4 Determining candidate locations

The probability of a demand request depends on demographic and property characteristics and is independent of the location of a fire station. Thus, from (1), we note that improvement of $SQI$ values can be achieved by placing additional fire stations in close proximity to areas with properties having high $SQI$ values (poor service) with respect to the existing fire stations. Let $\mathcal{I}_c$ denote the index set of candidate locations for new fire stations. We aim to find out the geographic locations of the candidate fire stations $i_c \in \mathcal{I}_c$ such that $SQI(j)$, $j \in \mathcal{J}_l$, is minimized (service becomes better). In this context, our objective is to locate high density clusters of properties in $\mathcal{J}_l$. Once the clusters are found, we can choose to place the new fire stations within the clusters by considering other qualitative features such as road accessibility, terrain condition, and availability of vacant space. Note that to improve the fire service, finding the clusters that have high densities of properties $j \in \mathcal{J}_l$, should be based on the travel time rather than Euclidean distance (which is not suitable for measuring the distance between points on a sphere) or haversine distance. For example, even if two properties are close to each other in terms of their longitude and latitude, the time to travel from one to another may be large due to various factors such as road inaccessibility or traffic congestion. Moreover, the existence of outliers (for example, a few properties far from the main city), and arbitrarily shaped clusters are well-known problems in spatial clustering that cannot be handled by centroid-based clustering algorithms such as K-Means [40]. Another challenge associated with clustering algorithms such as K-Means is the required specification of the number of clusters. Although there exist various methods to empirically find the optimal number of clusters [41], they are not always reliable in spatial clustering [4]. Therefore, we propose to use the *Density-based spatial clustering of applications with noise* (DBSCAN) algorithm [4] with distance measure replaced with travel time. We group properties based on the travel time between pairs of properties. Let $n$ denote the total number of properties in $\mathcal{J}_l$. We first compute the travel time between each pair of properties in $\mathcal{J}_l$ using the OSRM package [38] in R [39]. The travel time information is



**Fig. 1** Catchment area with one existing and two candidate stations. Circles with $T_{\max}$ radius denote areas that can be reached from respective stations within $T_{\max}$ time
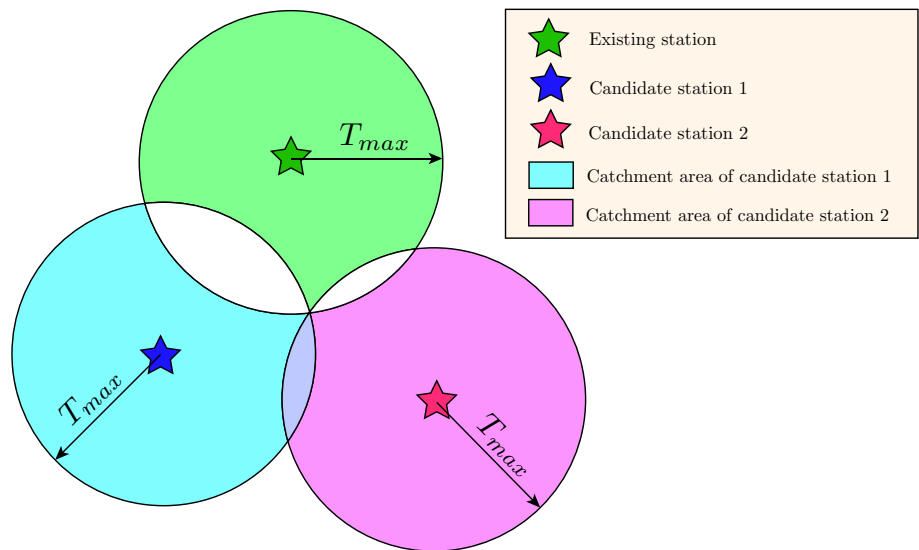
stored in a symmetric matrix $T \in \mathbb{R}_+^{n \times n}$, where each element $T[i, j]$ denotes the actual travel time from $i$ to $j$, $T_{\text{actual}}(i, j)$. The longitude and latitude values of each property are stored in $P \in \mathbb{R}^{n \times 2}$ such that the $i^{th}$ row of $P$ corresponds to the $i^{th}$ row of $T$. Our travel time-based DBSCAN requires two parameters, $\epsilon$ and $\delta$. $\epsilon$ denotes the $\epsilon$-neighborhood of a property $j$, which is defined as the set $\{k \in \mathcal{J}_l \mid T[k, j] \le \epsilon\}$. $\delta$ denotes the minimum number of properties required to be present in the $\epsilon$-neighborhood of $j$ to be considered a part of a cluster. Note that in travel time-based DBSCAN, the number of clusters is found during runtime and is not required to be specified a priori. Once the clusters are obtained, we choose to locate the candidate fire stations close to the centroids of the clusters (for our case study, we account for the factors such as availability of vacant space based on the future city development plan [37], road accessibility, terrain, etc.) The travel time-based DBSCAN algorithm is shown in Algorithm 1. Next, we present the optimization model to select the best location(s) among the candidate sites.

---

**Algorithm 1:** Finding Candidate Locations

**Function** FindCandidate *($P \in \mathbb{R}^{n \times 2}$, $T \in \mathbb{R}^{n \times n}, \epsilon, \delta$)*:
  $cluster\_id \leftarrow 0; outlier \leftarrow -1;$
  **for** $i = 1$ **to** $n$ **do** $L_n[i] \leftarrow undefined;$
  **for** $i = 1$ **to** $n$ **do**
    **if** $L_n[i] \ne undefined$ **then continue**;
    $N \leftarrow \{j : T[i, j] \le \epsilon\}$
    **if** $|N| < \delta$ **then**
      $L_n[i] \leftarrow outlier;$
      **continue;**
    **end**
    $cluster\_id \leftarrow cluster\_id + 1;$
    $L_n[i] \leftarrow cluster\_id;$
    $S_n \leftarrow \{s : s \in N, s \ne i\};$
    **foreach** $j \in S_n$ **do**
      **if** $L_n[j] = outlier$ **then**
        $L_n[j] \leftarrow cluster\_id;$
      **end**
      **if** $L_n[j] \ne undefined$ **then continue**;
      $N \leftarrow \{k : T[j, k] \le \epsilon\};$
      $L_n[j] \leftarrow cluster\_id;$
      **if** $|N| < \delta$ **then continue**;
      $S_n \leftarrow S_n \cup N;$
    **end**
  **end**
  **for** $c = 1$ **to** $cluster\_id$ **do**
    $C_f[c] \leftarrow centroid(\{P[k, :] : L_n[k] = c\});$
  **end**
**return** $C_f$;

---

# 5 Fire station location selection

Installation of new fire stations along with existing facilities can cause the service areas of the fire stations to overlap. It is desirable to locate new fire stations in such a way that

the overlap of service areas is minimized and the aggregated coverage area is maximized. In this context, we first introduce the concept of catchment area and then formulate an optimization problem utilizing the $SQI$ values and the catchment areas of each candidate location, to select the best location(s) among the candidates to install new fire station(s). Further, in a facility planning model, involving demand probabilities, it is required to characterize the stochasticity of decision outcome. Therefore, we propose a two-stage stochastic optimization formulation, based on a probabilistic reward function, to analyze the stochastic performance of our model. Note that, both of our proposed optimization formulations can be utilized to select the best location(s) to install new fire station(s) while the two-stage stochastic optimization model provides more insight into the confidence in the decision.

## 5.1 Catchment area

The fire service in a city usually maintains a performance bound on the time required to travel from the station to the incident location. The bound depends on the type of fire station [34] and the local government. Let $T_{\max}$ denote the specified bound on the travel time. With this constraint, the candidate locations of new fire stations $i_c \in \mathcal{I}_c$ should allow the number of demand points $j \in \mathcal{J}$ for which $T_{\text{actual}}(j, i_c) \le T_{\max}$ to be maximized. Essentially, the catchment area of a candidate location $i_c$ includes properties $j \in \mathcal{J}$ that cannot be accessed from any existing station $i \in \mathcal{I}$ within $T_{\max}$ but can be accessed from $i_c$ within the specified travel time limit of $T_{\max}$. An illustrative example of catchment areas with one existing fire station and two candidate fire stations is shown in Fig. 1.

In this example, the stations are located at the center of circles covering areas that can be reached within $T_{\max}$ units of time from the center. Note that the catchment areas of the candidate stations are the areas that do not overlap with the existing station. To ensure good area coverage along with $SQI$ improvement, we propose to place the future stations within the cluster of high $SQI$ properties that cover the maximum number of properties within their respective catchment areas. In the next section, we present the optimization model. It is derived from the maximum area coverage problem [42].

## 5.2 SQI-based optimization problem formulation

The best locations for future fire stations are those that maximize the number of properties with high $SQI$ (poor service by the existing fire stations) within their respective catchment areas. A property with higher $SQI$ (poorer service by existing fire stations) is given more priority relative to a property with lower $SQI$ (better service by existing fire stations). Therefore, we propose to weigh the properties according to

their $SQI$ values. Let us assume that it is required to locate at most $p$ number of additional fire stations. For each candidate location $i \in \mathcal{I}_c$ we consider a binary decision variable $x_i$. It takes a value of 1 if candidate location $i$ is selected and 0 otherwise. For each property $j \in \mathcal{J}$ we consider another binary decision variable $y_j$ which takes a value of 1 if that property is within the catchment area of any candidate location $i \in \mathcal{I}_c$, and 0 otherwise. For each property $j$, $N_j$ denotes the set of candidate stations whose catchment areas cover that property. In other words, the property $j$ can be reached within maximum allowable travel time from any of the candidate locations in $N_j$. Our objective is to obtain values for $x_i$ and $y_j$ that maximize the weighted sum $\sum_{j \in \mathcal{J}} SQI(j) y_j$. Hence, our proposed weighted maximum coverage optimization problem can be stated as follows:

$$\text{maximize} \sum_{j \in \mathcal{J}} SQI(j) y_j \tag{5}$$

$$\text{subject to} \sum_{i \in N_j} x_i \geq y_j, \quad \text{for all } j \in \mathcal{J}, \tag{6}$$

$$\sum_{i \in N} x_i \leq p, \tag{7}$$

$$x_i \in \{0, 1\} \text{ for all } i \in \mathcal{I}_c, \tag{8}$$

$$y_j \in \{0, 1\} \text{ for all } j \in \mathcal{J}, \tag{9}$$

where

$$\mathcal{J} = \text{ set of all demand locations within the city,}$$
$$\mathcal{I}_c = \text{Set of candidate fire stations,}$$
$$p = \text{Number of fire stations to be located,}$$
$$\hat{T}(j, i) = \text{Normalized travel time from fire station } i,$$
$$\text{to demand location } j,$$
$$\hat{T}_{\max} = \text{Normalized maximum allowable travel time,}$$
$$N_j = \{i \in \mathcal{I}_c \mid \hat{T}(j, i) \leq \hat{T}_{\max}\},$$
$$x_i = \begin{cases} 1 & \text{if } i \in N_j, \text{ for all } j \in \mathcal{J}, \\ 0 & \text{otherwise,} \end{cases}$$
$$y_j = \begin{cases} 1 & \text{if } \hat{T}(j, i) \leq \hat{T}_{\max}, i \in \mathcal{I}_c, \\ 0 & \text{otherwise.} \end{cases}$$

Let $\boldsymbol{x}^* := \{x_i^* : x_i^* \in \{0, 1\}, i \in \mathcal{I}_c\}$ and $\boldsymbol{y}^* := \{y_j^* : y_j^* \in \{0, 1\}, j \in \mathcal{J}\}$ denote the optimal values of the decision variables. The selected locations for the new fire stations are those for which $x_i^* = 1$.

## 5.3 Two-stage stochastic optimization problem formulation

Our proposed two-stage stochastic optimization problem is an iterative process involving a stochastic reward function $R_t(i)$ associated with each candidate location $i \in \mathcal{I}_c$ in iteration $t \in \mathbb{N}$. We associate a random variable $X_{j,t}$ with each demand location $j \in \mathcal{J}$ in $t^{th}$ iteration. We assume that, for all $t \in \mathbb{N}$ and $j \in \mathcal{J}$, the set of random variables $\{X_{j,t}\}$ are independent and identically distributed according to Bernoulli distribution with $\Pr(X_{j,t} = 1) = P(j) = 1 - \Pr(X_{j,t} = 0) = 1 - (1 - P(j))$. Note that, such a definition assigns a value of 1 or 0 to $X_{j,t}$ based on the demand request probability, $P(j)$. Let $\mathcal{C}_i := \{j \in \mathcal{J} \mid \hat{T}(j, i) \leq \hat{T}_{\max}, i \in \mathcal{I}_c\}$ denote the catchment area of candidate $i \in \mathcal{I}_c$. We define the reward function as

$$R_t(i) := \sum_{j \in \mathcal{C}_i} X_{j,t}. \tag{10}$$

Our proposed approach is as follows: at iteration $t \geq 1$, in the first stage, the candidate, $i_t \in \mathcal{I}_c$, having maximum expected reward till $(t-1)th$ iteration, $Q_{t-1}(i_t)$, is chosen. We compute the expected reward, till $tth$ iteration, corresponding to each candidate location as follows:

$$Q_t(i) := \frac{\sum_{k=1}^{t} \mathbb{1}_{(i_k=i)} R_k(i)}{\sum_{k=1}^{t} \mathbb{1}_{(i_k=i)}}, \tag{11}$$

where $\mathbb{1}_{(i_k=i)} = 1$ if candidate $i$ is chosen at iteration $k$, and 0 otherwise. Thus,

$$i_t = \text{argmax}_{i \in \mathcal{I}_c} Q_{t-1}(i). \tag{12}$$

In the next stage, all demand locations in its catchment area, $\mathcal{C}_{i_t}$, are triggered and the values $\{X_{j,t} \mid j \in \mathcal{C}_{i_t}\}$ are probabilistically determined and the reward $R_t(i_t)$ is evaluated as per (10). Once $R_t(i_t)$ is evaluated, we update the expected reward for all $i \in \mathcal{I}_c$ as per (11).

We employ an epsilon-greedy approach to enable exploration of all candidates to find the optimal candidate that maximizes the reward, that is, with probability $(1 - \epsilon)$ we choose $i_t$ as per (12) and with probability $\epsilon$, we randomly choose a candidate from $\mathcal{I}_c$. Upon termination of our proposed algorithm, based on a maximum iteration count $t_{\max}$, the candidate having the maximum expected reward is chosen as the optimal location to install a new fire station. Note that, if $p$ number of fire stations are to be located, where $p > 1$, the first $p$ candidates corresponding to expected rewards sorted in descending order, can be chosen. The sets $\mathcal{I}_c$ and $\mathcal{C}_i$ are derived as described in Sects. 4 and 5.1 respectively and $\epsilon$ is tuned empirically. The entire algorithm is summarized in Algorithm 2.

---

**Algorithm 2:** Finding Optimal Candidate

---

**Function** getOptimalCand($\{P(j), j \in \mathcal{J}\}, \epsilon, \mathcal{I}_c,$
$\{\mathcal{C}_i, i \in \mathcal{I}_c\}, p$)**:**

   **for** $i = 1$ **to** $|\mathcal{I}_c|$ **do**
      Initialize $R_0(i), Q_0(i)$;
      $\boldsymbol{Q}[i] \leftarrow Q_0(i)$;
   **end**
   **for** $t = 1$ **to** $t_{max}$ **do**
      $i_t \leftarrow \begin{cases} \text{argmax}_{i \in \mathcal{I}_c} Q_{t-1}(i), & \text{with } \Pr(1 - \epsilon) \\ \text{any } i \in \mathcal{I}_c \text{ with } \Pr(\epsilon) \end{cases}$ ;
      **for** $j = 1$ **to** $|C_{i_t}|$ **do**
         $X_{j,t} \sim \text{Bernoulli}(P(j))$;
      **end**
      Update $R_t(i_t)$ as per (10);
      Update $Q_t(i_t)$ as per (11);
      $\boldsymbol{Q}[i_t] \leftarrow Q_t(i_t)$;
   **end**
   $\boldsymbol{I}[\cdot] \leftarrow \text{argsort}_i \boldsymbol{Q}[i]$;
**return** $\{\boldsymbol{I}[1], \boldsymbol{I}[2], \ldots, \boldsymbol{I}[p]\}$;

---

We now present a detailed case study of our entire fire station location methodology described in the previous sections.

# 6 Case study: City of Victoria, MN

To apply our proposed methodology, we partnered with the city of Victoria fire Department (VFD) to select a suitable location for a new fire station. The city of Victoria is located in Carver county of Minnesota, USA. The city enjoys a solid economic and agricultural foundation. This has catalyzed rapid urbanization and population growth in the past few years. The increased acceptance of remote-work situations will further add to the city's growth. Per the US Census Bureau [43], the population of the city rose from 7400 in 2010 to 9170 in 2017. The current population (at the time of writing) is approximately 10,000 and this number is expected to increase by 68% by the year 2040 [37]. Currently, Victoria is being served by one fire station. The expected growth and economic development have caused concern among the city's leadership that the current infrastructure may be insufficient to provide adequate public safety services, including those services provided by the fire department. Therefore, VFD initiated a resilient community project (RCP) in 2019 to assess the performance of the existing facility and to identify a location for an additional fire station. The new facility will enhance (not replace) the capabilities of the existing fire station. A data exploratory analysis, as a part of the RCP initiative, is performed in [44] to find spatio-temporal demand statistics. Following the statistics and the city guidelines, as presented in [44], we apply our location selection methodology to provide a comprehensive guidance on future station

location. Our database features and the detailed description of our results are presented in the next sections.

## 6.1 Description of data

The entire database used in this study is comprised of data from multiple different sources. Data containing the details (date, locations, response time etc.) of historical incidents that occurred between January 2010 and October 2020 are provided by VFD. In particular, the dataset contains: (1) incident types (fire, explosion and overheat, Rescue and emergency medical service, Hazardous condition, service call, good intent call, false alarm, natural disaster, and special intent type), (2) address of incident location, (3) date and time, (4) Response time, (5) fire service vehicle used. As there are reported incidents for these properties between 2010 and 2020, these properties are designated as demand requesting locations. We use Carver county's open data portal [45] to further augment this dataset with properties with no reported incidents, and these properties are designated as no demand requesting locations. The entire dataset contains 5334 samples of which 47.9% have reported historical incidents. Parcel information for each property location is also obtained from [45]. The parcel information includes address, land size (acres), estimated land value ($), property usage category (see Table 1), and year of construction. This dataset is augmented with block-wise average population and the median age of residents obtained from the US census bureau and Federal Communications Commission [46,47]. For the spatial analysis, the latitude and longitude of the property locations are obtained from the US Census Bureau Geocoding Service [48]. However, some locations cannot be geocoded due to incorrect address entries, such as zip code and address line mismatch, which are primarily caused by human errors. These entries are matched by address lines in the Carver County Active911 database [49] to obtain the geographic locations. The US census bureau geocoder along with the Active911 database reduce the number of missing data due to common errors. Every record is matched by property address to create a fully processed database. There are several sources of missing data, mainly due to misspellings and abbreviations of the addresses (such as 'highway' is termed as 'hwy' / 'hgwy' / 'highway' and sometimes misspelled as 'hgihway', 'drive' is represented as 'Dr'/'drv'/'drive' etc.). Some of the common errors are fixed in the data cleaning step; however, several different types of errors lead to some amount of missing data. (4.2% of the entire dataset). We performed our proposed station location selection analysis on this dataset and the results are presented in the subsequent sections.

**Fig. 2** Locations of all incidents occurred during 2010–2020 (till Oct): Incidents are concentrated in southern region and in residential areas. Existing facility is marked with black star icon and the locations are marked with red dots. Blue solid line indicates fire district boundary. Black dashed line indicates Victoria city boundary (color figure online)



★ Victoria fire department
⬚ Fire district boundary
⋮⋮⋮ Victoria city boundary
● Incident location
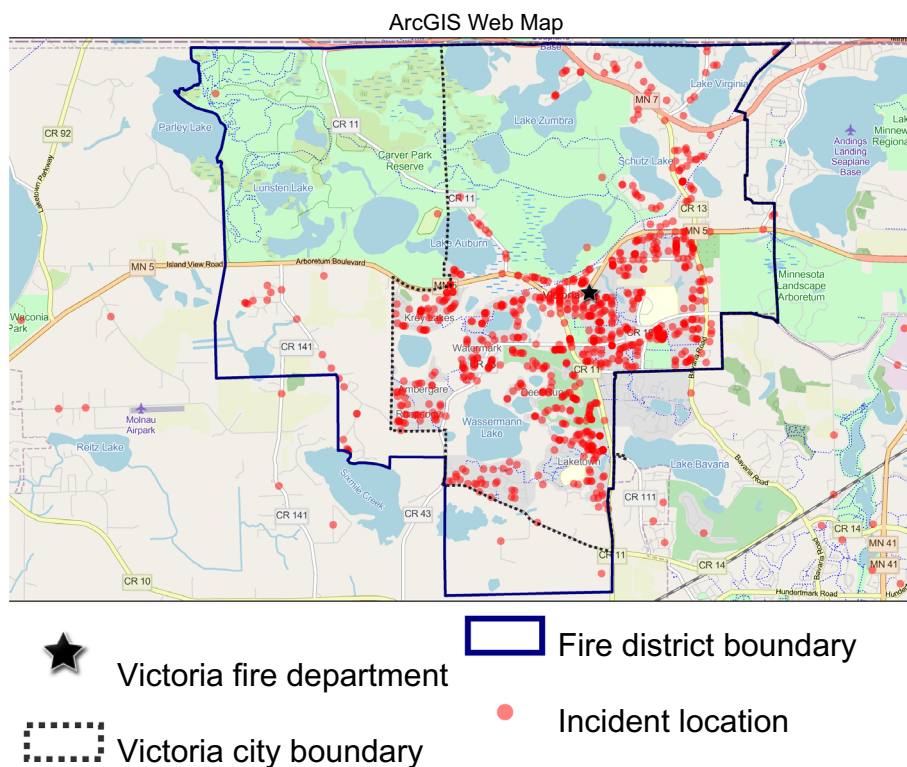
## 6.2 Demand prediction using random forest and XGBoost

Figure 2 shows the locations of all incidents that occurred in 2010–2020 (until Oct 2020) on an open street map. At the time of writing, the City of Victoria has dense residential areas in the southeast [37]. The concentration of the incident locations in the southeast, as can be seen in Fig. 2, indicates that the service demand is mostly concentrated in the residential areas. All of the maps are created using ArcGIS [50]. The shapefiles of the fire district and city boundary are taken from [51,52]. The incidents outside the fire district are typically serviced in cooperation with other fire departments, while some of the events that are outside the city boundary, but inside the Victoria fire district, are serviced by the VFD.

For the rest of our analysis, we consider only those properties that are within the VFD fire district and served by VFD only, in case of emergency events. We combine the properties with a history of reported incidents and other properties with no historical event. The entire dataset, with the sample features shown in Table 2, is used to build our proposed prediction models for demand request prediction. We implement our RF model using the Scikit-learn package [53] and XGBoost model using the xgboost package in Python. 80% of the total number of samples are used for training the model and the remaining 20% are used for performance testing and fine-tuning. Tables 3 and 4 show the set of parameters used for grid search, with 10-fold cross vali-

**Table 3** Grid parameters for Demand prediction RF model training

| Parameters | Grid search parameters | Selected parameter |
| --- | --- | --- |
| Criterion | Entropy, Gini index | Gini index |
| Max tree depth | 2, 4, 6, 8, 10 | 8 |
| Number of trees | 100, 200, 300, 400 | 300 |
| No. of features randomly chosen | $\sqrt{\text{No. of total features}}$ | 3 |
| Min samples in Leaf nodes | 30, 40, 50 | 30 |

dation, to find the best set of parameters of RF and XGBoost model respectively. For our RF model, we set the minimum number of samples required to split a node of a tree to 2, and enable bootstrapping from the entire training sample set while building the trees. The selected parameters in the rightmost column resulted in the best accuracy. The performance of the prediction models are presented in Table 5.

The ROC curve [54] is shown in Fig. 3a. It can be seen that both of our models perform very well in predicting the demand request of different property types. Further, our extreme gradient boosting model performs significantly better than the RF model in terms of accuracy and true and false positive rates. RF model exhibits a true positive rate (TPR) of 69% and 65% for the training and the test set respectively while with XGBoost they are improved to 89% and 80%

**Table 4** Grid parameters for Demand prediction XGBoost model training

| Parameters | Grid search parameters | Selected parameter |
|---|---|---|
| Learning rate | 0.01, 0.1, 0.2, 0.3 | 0.2 |
| Max tree depth | 2, 4, 6, 8, 10 | 4 |
| Number of trees | 100, 200, 300, 400 | 100 |
| Gamma | 0, 1, 2, 3, 4 | 3 |
| Alpha | 0.001, 0.002, 0.003 | 0.002 |
| Lambda | 1, 2, 3 | 1 |
| Subsample | 0.7, 0.8, 0.9 | 0.8 |
| Random feature Ratio per tree | 0.5, 0.6, 0.7 | 0.6 |

**Table 5** Demand prediction model performance measures

| Performance measure | RF | XGBoost |
|---|---|---|
| Training accuracy | 74% | 87% |
| Test accuracy | 70% | 78% |
| Out-of-Bag score | 69% | NA |
| True positive rate: Training set | 69% | 89% |
| False positive rate: Training set | 22% | 11% |
| True positive rate: Test set | 65% | 80% |
| False positive rate: Test set | 26% | 21% |
| AUC: Training set | 0.83 | 0.94 |
| AUC: Test set | 0.79 | 0.85 |



**(a)**



**(b)**

**Fig. 3** Demand Prediction ML model performance: **a** ROC curve. AUC values are $\geq 0.7$, which indicates the efficacy of our proposed model. **b** Feature importance shows that property age and residents age are the two most important factors governing the demand request

respectively. The false positive rate (FPR) values of 11% and 21%, of XGBoost model, are relatively low and significantly better than those of the RF model. While TPR is associated with performance guarantees of our $SQI$-based stochastic optimization-based planning algorithms, FPR is closely related to the cost effectiveness of VFD and local government, as it can result in unnecessary investments in safety measures. We derive impurity-based feature importance, as illustrated in Fig. 3b, to explain the effect of demographic and property features on the service demand request. Here, we observe that the most important features in determining service demand are property age, median age of residents in the block where the property is located, land size, population at the block level, and estimated land value. We note that this additional information can be utilized by the city of Victoria to further analyze future service demand with projected growth [37]. As our XGBoost model performs better than the RF model, we use the class probability estimates of our XGBoost model as the demand request probabilities for the calculation of $SQI$. The raw class probabilities are scaled to have values between 0 to 1 using min–max normalization. We categorize the demand request probabilities into three categories as shown in Table 6, in order to tune $\tau_l$
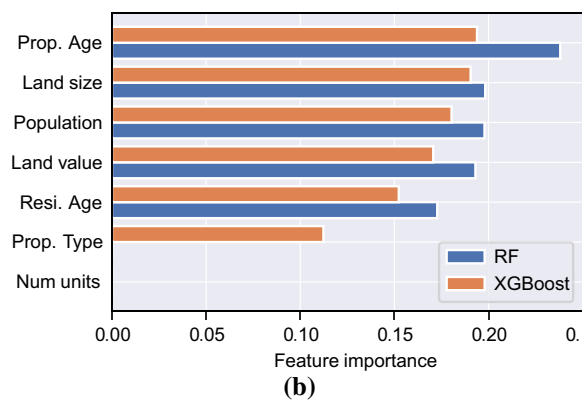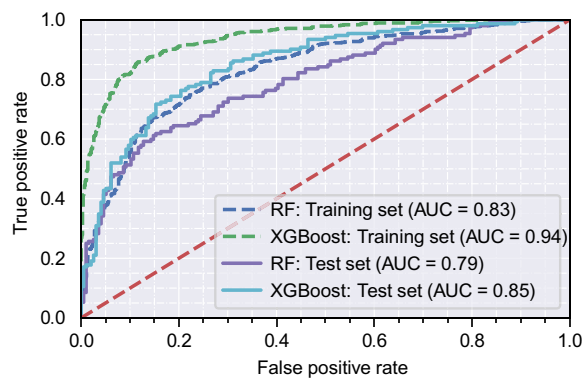
**Table 6** Demand Request Probability of Properties

| Demand probability | Number of properties (%) |
|---|---|
| Low ($0 \leq Pr < 0.35$) | 18.1 |
| Medium ($0.35 \leq Pr < 0.65$) | 61.9 |
| High ($0.65 \leq Pr \leq 1$) | 20 |

and $\tau_h$ that are required for (4). Table 6 shows the number of properties falling under each category as a percentage of total number of properties within the VFD fire district. The spatial distribution of the same is presented in Fig. 4.

After obtaining the demand request probabilities, the travel time from VFD to each property are required in order to compute the $SQI$ values and the catchment areas. In the next sections, we present the comprehensive results of our new fire station location selection methodology based on $SQI$ values and two-stage stochastic optimization model as described in Sects. 4 and 5.

**Fig. 4** Spatial property distribution based on probability of demand request: Existing facility is marked with black star icon and low, medium, and high service quality properties are marked in red, orange, and green dots respectively. Blue solid line indicates fire district boundary. Black dashed line indicates Victoria city boundary (color figure online)
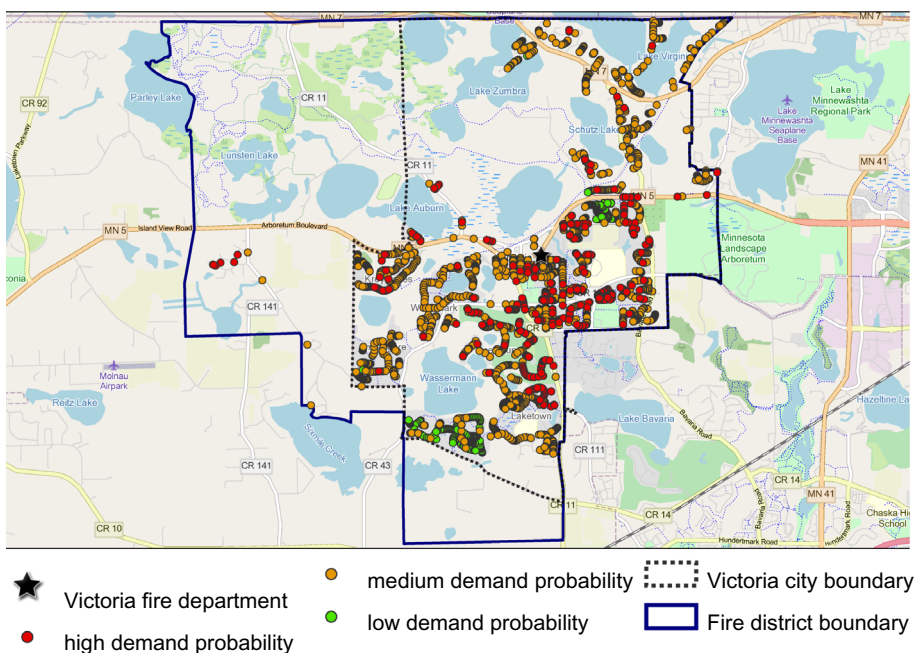


**Table 7** Location selection parameter values

| Parameter | Values | Units |
| --- | --- | --- |
| $T_{\max}$ | 4 | Minutes |
| $T_{\text{norm}}$ | 20 | Minutes |
| $\tau_l$ | 0.05 | None |
| $\tau_h$ | 0.16 | None |
| $\epsilon$ | 2 | Minutes |
| $\delta$ | 80 | None |
| $p$ | 1 | None |

**Table 8** $SQI$ percentages

| Quality of service | Number of properties (%) |
| --- | --- |
| Low ($\tau_h \leq SQI \leq 1$) | 12 |
| Medium ($\tau_l \leq SQI < \tau_h$) | 72.9 |
| High ($0 \leq SQI < \tau_l$) | 15.1 |

## 6.3 Fire station location selection

Once the probability of service demand of all properties are predicted using the prediction model and the travel time from VFD to each property is obtained from OSRM [38], we calculate the $SQI$ values of the properties using (3). VFD has a performance goal that travel time is not to exceed 4 min. Therefore, we have taken $T_{\max} = 4$. The other parameters that are required for Algorithm 1 and the for the SQI-based optimization model, described in Sect. 5.2, are listed in Table 7. Note that the parameters $\tau_l, \tau_h$ are derived from the demand request probability categorization shown in Table 6 and Eq. (1).

We calculate the $SQI$ values and categorize the properties into low, medium, and high service quality properties as per (4). The total number of properties falling under each category as a percentage of total number of properties within the fire district is presented in Table 8.

The geographical distribution of properties is shown in Fig. 5. Here, a visual comparison between Figs. 4 and 5

is useful to provide an interpretation of the remark made in Sect. 3.1. It is observed that most of the properties with medium demand request probabilities, but far from the existing fire station, are marked as receiving low quality of service while the properties with high demand but close to the existing fire station are marked as receiving better service.

Now, we apply Algorithm 1 on the properties marked with red in Fig. 5 to find the candidate locations. Algorithm 1 is implemented using the Scikit-learn package [53] in Python. The result is shown in Fig. 6 along with the 4-min drive time polygons of the candidates. We only show the approximate locations of the candidates with circles on the map to conform to the confidentiality of actual locations as requested by VFD. Based on Algorithm 1, we find three candidate locations in the fire district: (a) northeast, (b) southwest, and (c) south. For the rest of our analysis, we denote the candidates by:

1. Northeast (NE)
2. Southwest (SW)
3. South (S).

Note that the non-overlapped 4-min travel time polygons between SW, S, and NE with the existing station form the

**Fig. 5** Service Quality with existing fire station: Existing facility is marked with black star icon and low, medium, and high service quality properties are marked in red, orange, and green dots respectively. Blue solid line indicates fire district boundary. Black dashed line indicates Victoria city boundary (color figure online)
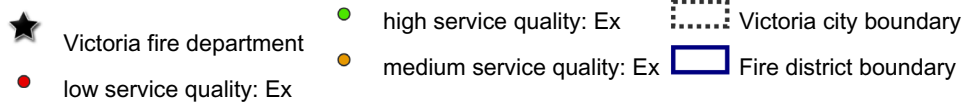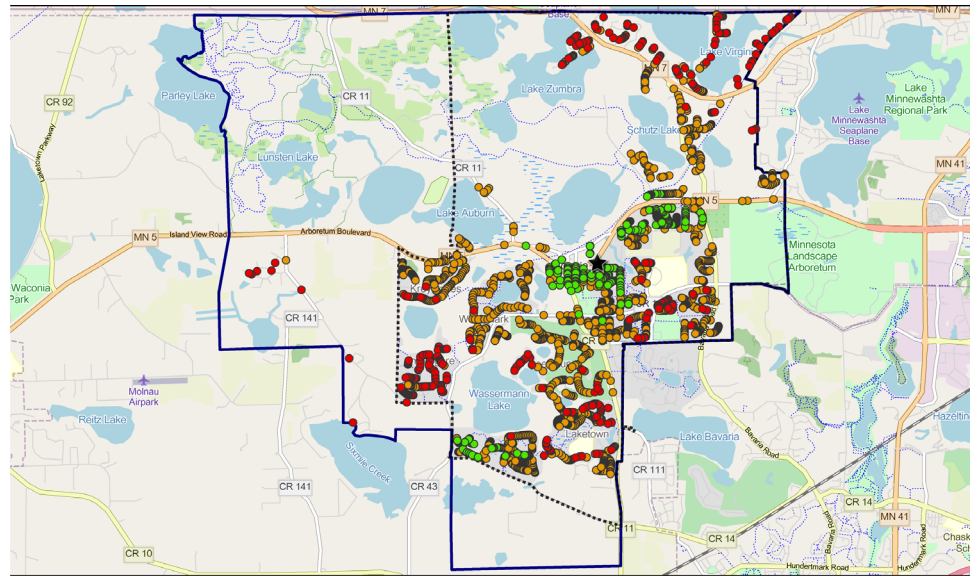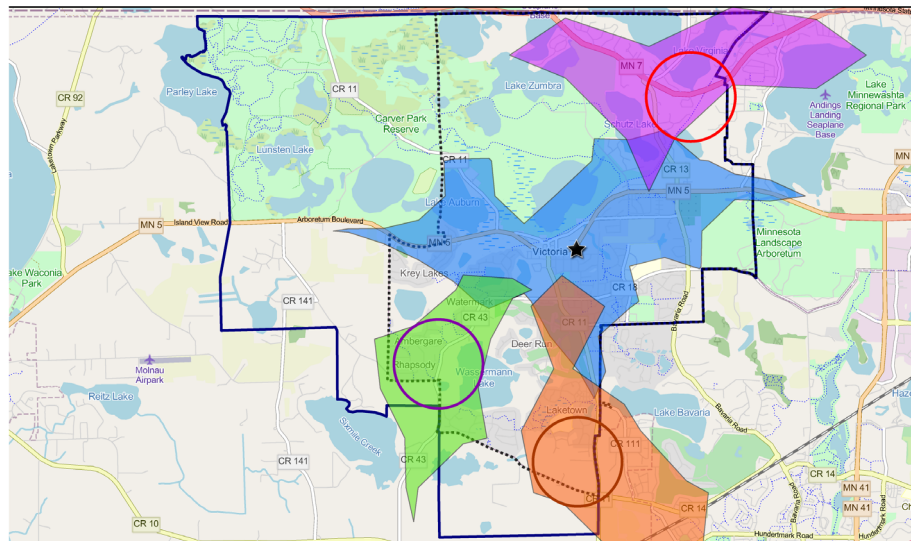


★ Victoria fire department

● low service quality: Ex

● high service quality: Ex

● medium service quality: Ex

⋮ Victoria city boundary

▭ Fire district boundary

**Fig. 6** Catchment areas of candidate locations based on 4-min drive time polygon: Overlap is least for SW thus indicating least redundancy in service along with the existing facility. Existing facility is marked with black star icon and the candidate locations are shown with large circles to indicate broader zones. Blue solid line indicates fire district and black dashed line indicates Victoria city boundary (color figure online)



○ Option: south-west

○ Option: south

○ Option: north-east

■ 0-4 min: option south-west

■ 0-4 min: option south

■ 0-4 min: option north-east

■ 0-4 min: existing fire station

★ Victoria fire department

⋮ Victoria city boundary

▭ Fire district boundary

**Table 9** *SQI* percentages

| Service quality | Ex (%) | Ex+SW (%) | Ex+S (%) | Ex+NE (%) |
|---|---|---|---|---|
| Low | 12 | 9 | 9.4 | 9.6 |
| Medium | 72.9 | 67.2 | 70.4 | 71.8 |
| High | 15.1 | 23.8 | 20.2 | 18.6 |

catchment areas of the candidates. We aim to apply our SQI-based optimization methodology, as described in Sect. 5.2, with respect to these catchment areas. Properties within the catchment areas of SW, S, and NE are obtained using 'sf' package [55] in R. VFD seeks a location for one additional fire station, i.e. $p = 1$. We remark that although the project with VFD requires *one* additional station, our methodology can be applied without modification to more general settings for any value of $p$. The optimization problem is solved using ILOG CPLEX Optimization Studio V20.1 [56]. The solution to the optimization problem indicates that SW candidate is the one that maximizes the number of properties with high *SQI* (lower service quality, marked red in Fig. 5) within its catchment area. A detailed comparison of the number of properties in each *SQI* category as a percentage of total number of properties is presented in Table 9. The existing fire station is abbreviated as 'Ex' and the expected performance, in case of an additional fire station with the existing one, is tabulated. Note that while the existing station provides high quality service to 15.1% of properties, an additional fire station at SW can enhance it to 23.8%. Thus, we can expect an improvement of 53% over the existing performance. However, the installation of an additional station at S or NE location results in an approximate $23 - 33\%$ improvement over the existing one. Similarly, in the case of low service quality properties, SW location provides an improvement of approximately 25%, while S and NE both provide an approximate 20% improvement.

To analyze the improvement in *SQI* from the installation of an additional fire station at the SW location, we show histograms of *SQI* values of all the properties considering only the existing fire station and considering the existing station along with SW candidate, as per (3) (see Fig. 7a). We see that the addition of the SW candidate location increases the density towards the head of the distribution, thus improving the number of properties with low *SQI* (better service quality).

Similarly, Fig. 7b, c show the *SQI* histogram comparison between SW, NE, and SW, S candidates. We observe that the improvement by installing SW candidate, which is visible from the increased density towards the head of the distribution, is clear from all the plots.
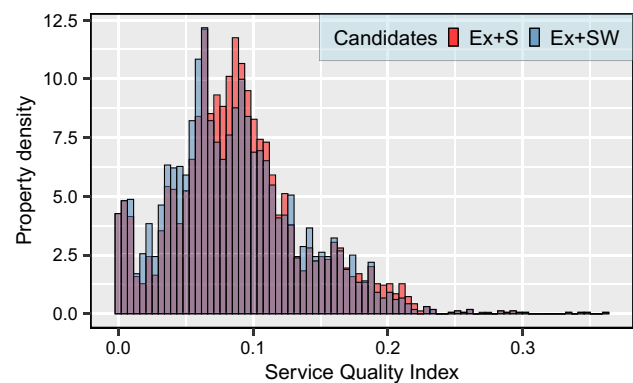
We conclude our analysis by illustrating the geographical distribution of properties categorized, as per (4), with both the existing fire station and the additional fire station at SW location. The plot is shown in Fig. 8.
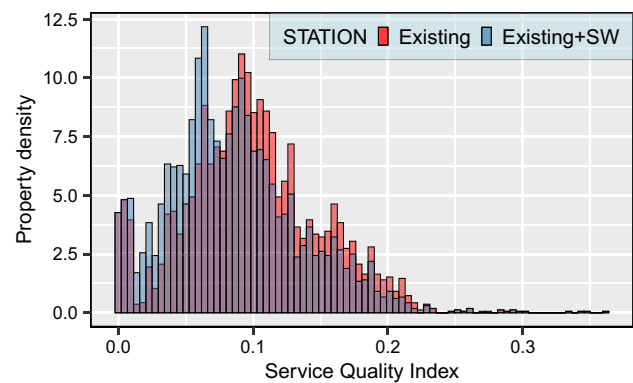


(a)
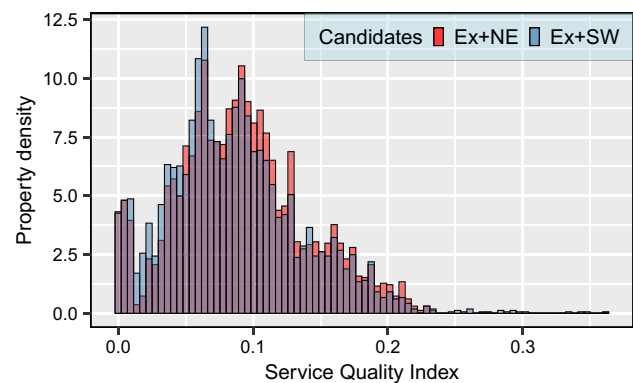


(b)



(c)

**Fig. 7** **a** SW candidate along with existing fire station (marked in blue) increases the density towards the head of the distribution, thus improving number of properties with better service quality, compared to that with only existing station (marked in red). **b** SW candidate along with existing station (marked in blue) improves number of properties with better service quality compared to NE candidate with the existing station (marked in red). **c** Similar improvement can be seen by comparing the histograms corresponding to SW with existing station (marked in blue) and S with existing station (marked in red) (color figure online)

**Fig. 8** Visual comparison with Fig. 5 reveals that installation of additional fire station at SW location would significantly improve the number of properties with high fire service quality (marked in green dots). Existing facility is marked with black star icon and the potential future fire station at SW location is shown with large circle to indicate broader zone. Blue solid line indicates fire district boundary. Black dashed line indicates Victoria city boundary (color figure online)
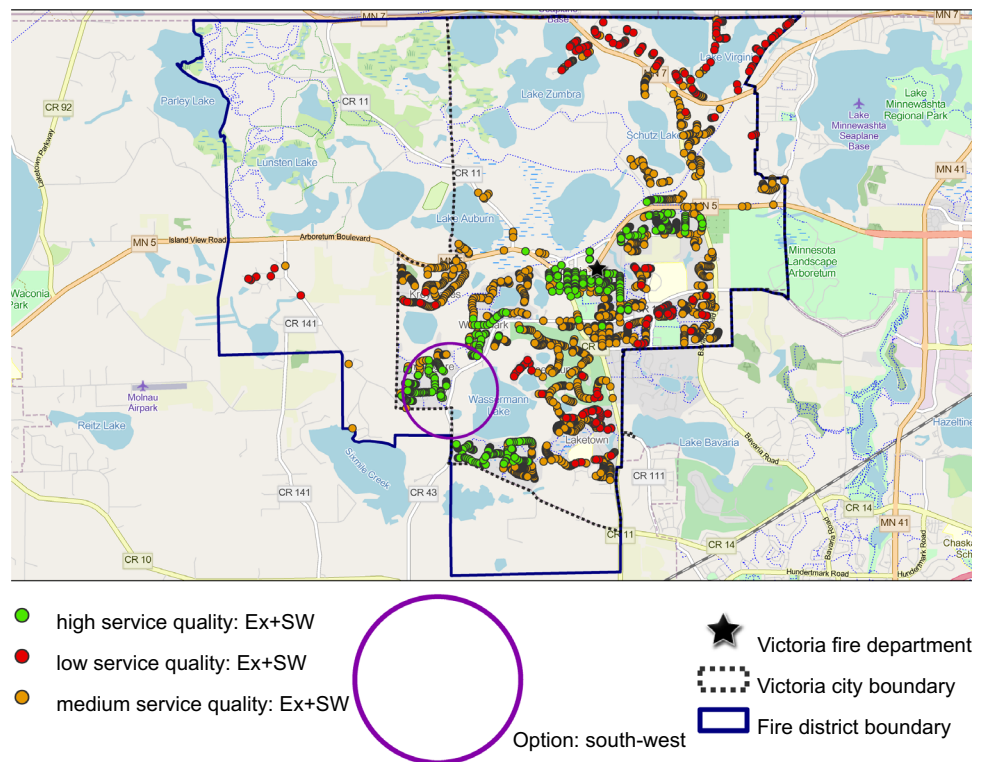


A visual comparison of Figs. 5 and 8 reveals that the installation of the SW candidate significantly reduces the number of properties with poor service quality. Note that the concentration of poor service quality properties in the northeast and southern parts of the city is difficult to avoid with the installation of only one additional fire station due to longer travel times from either the existing facility or from the SW candidate location. The longer travel time is mainly attributed to the presence of lakes, terrain conditions, and the road network. Next, we employ our proposed two-stage stochastic optimization model, as described in Sect. 5.3, to characterize the confidence in our decision to locate a new fire station at the SW location, which is apparent from our previous discussion. We run our proposed Algorithm 2 in 400 episodes each with a maximum iteration number $t_{max} = 200$. The catchment areas and the candidate locations are same as described in the previous sections. The $\epsilon$ value is taken to be 0.7. Figure 9 shows the density histogram plot of the expected rewards of the candidates SW, S, and NE for 400 episodes. The histograms are plotted with a bin size of 40. It can be observed that the SW location accumulates the highest expected reward followed by S and NE respectively. Further, the narrower histogram corresponding to the SW location indicates a lower standard deviation of the expected rewards, indicating higher confidence in the decision to choose SW as the optimal location to install a new facility. Therefore, we propose the SW location to VFD for the installment of an additional fire station.
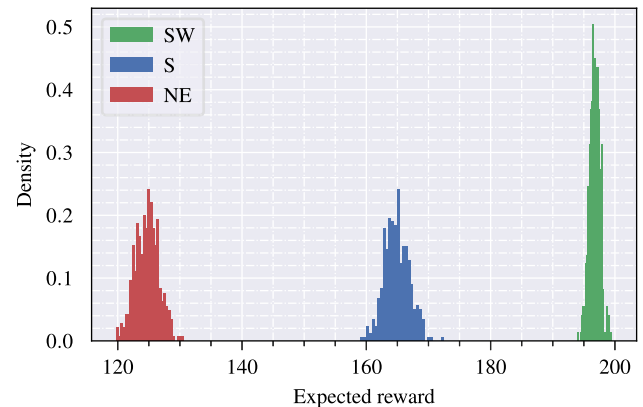


**Fig. 9** Histogram of expected reward accumulated by the candidates SW, S, and NE for 400 episodes each with 200 iterations. SW has the maximum expected reward. Variation of expected reward is significantly lesser in case of SW, compared to S and NE, as can be observed by narrow standard deviation of the histogram corresponding to SW. Thus, selection of SW location as the optimal choice to install a new fire station, as determined by our proposed algorithm has the highest confidence

## 7 Conclusions

In this work we present a systematic approach for locating new fire stations, taking into account demand prediction, travel time, service area coverage, and service redundancy. We propose a machine learning model based on the random forest and extreme gradient boosting algorithms to predict

service demand, which can then be used to enhance community safety. We present a detailed comparison between the performance of the models to select the better model for further study. We observe that the extreme gradient boosting model outperforms the random forest model. With the help of the machine learning model, we corroborate the influence and the importance of the key factors governing fire service demand associated with the properties. The model is further utilized to define a service quality index to quantify the quality of service, which accounts for both demand prediction and travel time from fire stations to demand locations. The quality index is utilized in a travel time-based DBSCAN algorithm to identify candidate locations for new fire stations. Finally, we present a service quality index-based optimization model that is derived from the maximum coverage optimization problem, considering demand prioritization and service area redundancy. We apply our proposed methodology to select a location for a new fire station in the city of Victoria, MN, USA. Further, we employ a two-stage stochastic optimization model to characterize the confidence in our decision outcome. We recommend that a future fire station be located in the southwest portion of the city. This location would best serve the community and improve the public safety standard.

## Declarations

**Conflict of interest** The authors declare that they have no conflict of interest.

**Code availability** Not applicable.

**Ethics approval** Not applicable.

**Consent to participate** Not applicable.

**Consent for publication** Not applicable.

## References

1. Ahrens, M., Evarts, B.: Fire loss in the united states during 2019. National Fire Protection Association (NFPA) report (2020)
2. Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (2001)
3. Chen, T., Guestrin, C.: Xgboost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794 (2016)
4. Ester, M., Kriegel, H.P., Sander, J. et al.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: kdd, pp. 226–231 (1996)
5. Turkoglu, D.C., Genevois, M.E.: A comparative survey of service facility location problems. Ann. Oper. Res., pp. 1–70 (2019)
6. Yao, J., Zhang, X., Murray, A.T.: Location optimization of urban fire stations: access and service coverage. Comput. Environ. Urban Syst. **73**, 184–190 (2019)
7. Aktaş, E., Özaydın, Ö., Bozkaya, B., et al.: Optimizing fire station locations for the Istanbul metropolitan municipality. Interfaces **43**(3), 240–255 (2013)
8. Nyimbili, P.H., Erden, T.: Comparative evaluation of gis-based best-worst method (bwm) for emergency facility planning: perspectives from two decision-maker groups. Nat. Hazards **105**(1), 1031–1067 (2021)
9. Church, R.L., Li, W.: Estimating spatial efficiency using cyber search, gis, and spatial optimization: a case study of fire service deployment in los angeles county. Int. J. Geogr. Inf. Sci. **30**(3), 535–553 (2016)
10. Chevalier, P., Thomas, I., Geraets, D., et al.: Locating fire stations: an integrated approach for Belgium. Socioecon. Plann. Sci. **46**(2), 173–182 (2012)
11. Murray, A.T.: Advances in location modeling: Gis linkages and contributions. J. Geogr. Syst. **12**(3), 335–354 (2010)
12. Soltani, A., Balaghi, R., Rezaei, M., et al.: Spatial analysis and urban land use planning with emphasis on hospital site selection, case study: Isfahan city. Bull. Geogr. Socio-economic Ser. **43**(43), 71–89 (2019)
13. Abdullahi, S., Mahmud, A.R,B., Pradhan, B.: Spatial modelling of site suitability assessment for hospitals using geographical information system-based multicriteria approach at Gazvin city, Iran. Geocarto Int. **29**(2), 164–184 (2014)
14. Erlenkotter, D.: A dual-based procedure for uncapacitated facility location. Oper. Res. **26**(6), 992–1009 (1978)
15. Toregas, C., Swain, R., ReVelle, C., et al.: The location of emergency service facilities. Oper. Res. **19**(6), 1363–1373 (1971)
16. Abareshi, M., Zaferanieh, M.: A bi-level capacitated p-median facility location problem with the most likely allocation solution. Transp. Res. Part B: Methodol. **123**, 1–20 (2019)
17. Bolouri, S., Vafaeinejad, A., Alesheikh, A., et al.: Minimizing response time to accidents in big cities: a two ranked level model for allocating fire stations. Arab. J. Geosci. **13**(16), 1–13 (2020)
18. O'Connor, C.D., Calkin, D.E., Thompson, M.P.: An empirical machine learning method for predicting potential fire control locations for pre-fire planning and operational fire management. Int. J. Wildland Fire **26**(7), 587–597 (2017)
19. Sevinc, V., Kucuk, O., Goltas, M.: A Bayesian network model for prediction and analysis of possible forest fire causes. For. Ecol. Manage. **457**(117), 723 (2020)
20. Eden, J.M., Krikken, F., Drobyshev, I.: An empirical prediction approach for seasonal fire risk in the boreal forests. Int. J. Climatol. **40**(5), 2732–2744 (2020)
21. Choi, M.Y., Jun, S.: Fire risk assessment models using statistical machine learning and optimized risk indexing. Appl. Sci. **10**(12), 4199 (2020)

22. Bui, D.T., Van Le, H., Hoang, N.D.: Gis-based spatial prediction of tropical forest fire danger using a new hybrid machine learning method. Eco. Inform. **48**, 104–116 (2018)
23. Salehi, M., Rusu, L.I., Lynar, T. et al.: Dynamic and robust wildfire risk prediction system: an unsupervised approach. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 245–254 (2016)
24. Turco, M., Jerez, S., Doblas-Reyes, F.J., et al.: Skilful forecasting of global fire activity using seasonal climate predictions. Nat. Commun. **9**(1), 1–9 (2018)
25. Cheng, T., Wang, J.: Integrated spatio-temporal data mining for forest fire prediction. Trans. GIS **12**(5), 591–611 (2008)
26. Singh Walia, B., Hu, Q., Chen, J., et al.: A dynamic pipeline for spatio-temporal fire risk prediction. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 764–773 (2018)
27. Madaio, M., Chen, S.T., Haimson, O.L., et al.: Firebird: predicting fire risk and prioritizing fire inspections in Atlanta. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 185–194 (2016)
28. Agarwal, P., Tang, J., Narayanan, A.N.L., et al.: Big data and predictive analytics in fire risk using weather data. Risk Anal. **40**(7), 1438–1449 (2020)
29. Jin, G., Zhu, C., Chen, X. et al.: Ufsp-net: a neural network with spatio-temporal information fusion for urban fire situation prediction. In: IOP Conference Series: Materials Science and Engineering, IOP Publishing, p. 012050 (2020)
30. Lau, C.K., Lai, K.K., Lee, Y.P., et al.: Fire risk assessment with scoring system, using the support vector machine approach. Fire Saf. J. **78**, 188–195 (2015)
31. Salehi, M., Rashidi, L.: A survey on anomaly detection in evolving data: [with application to forest fire risk prediction]. ACM SIGKDD Explor. Newsl. **20**(1), 13–23 (2018)
32. Jin, G., Wang, Q., Zhu, C., et al.: Urban fire situation forecasting: deep sequence learning with spatio-temporal dynamics. Appl. Soft Comput. **97**(106), 730 (2020)
33. Jaldell, H.: How important is the time factor? saving lives using fire and rescue services. Fire Technol. **53**(2), 695–708 (2017)
34. Association, N.F.P., et al.: NFPA 1720, standard for the organization and deployment of fire suppression operations, emergency medical operations, and special operations to the public by volunteer fire departments. National Fire Protection Association (2020)
35. Liu, K., Li, X., Zou, C.C. et al.: Ambulance dispatch via deep reinforcement learning. In: Proceedings of the 28th International Conference on Advances in Geographic Information Systems, pp 123–126 (2020)
36. Dey, A.: Planning the future fire station in the city of Victoria (2020)
37. Council, V.C.: Our Victoria Tomorrow. City Council (2019)
38. Giraud, T. et al.: osrm r package. https://cran.r-project.org/web/packages/osrm/index.html, [Accessed 2 Nov 2020] (2020)
39. R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/ (2019)
40. Bishop, C.M.: Pattern recognition. Mach. Learn. **128**(9) (2006)
41. Alpaydin, E.: Introduction to Machine Learning. MIT Press (2020)
42. Church, R., ReVelle, C.: The maximal covering location problem. In: Papers of the Regional Science Association, Springer, pp. 101–118 (1974)
43. Bureau, U.S.C.: City and town population totals: 2010-2019. https://www.census.gov/data/datasets/time-series/demo/popest/2010s-total-cities-and-towns.html, [Accessed 2 Nov 2020] (2019)
44. Wu, E.: Planning the future fire station in the city of Victoria https://conservancy.umn.edu/handle/11299/216227, [Accessed 1 Oct 2020] (2020)
45. County, C.: Carver county's open data parcels. https://data-carver.opendata.arcgis.com/datasets/open-data-parcels, [Accessed 1 November 2020] (2020)
46. Census: Census data. https://data.census.gov/cedsci/all?g=0500000US27019.100000, [Accessed 1 Feb 2021] (2020)
47. FCC: Federal communications commission api documentation for developers. https://geo.fcc.gov/api/census/, [Accessed 1 Feb 2021] (2020)
48. Bureau, U.S.C.: U.S. census bureau census geocoding service. https://geocoding.geo.census.gov/, [Accessed 1 November 2020] (2020)
49. County, C.: Carver county's open data portal. https://data-carver.opendata.arcgis.com/datasets/active-address-points-active911, [Accessed 1 November 2020] (2020)
50. ArcGIS: Arcgis online. https://www.arcgis.com/, [Accessed 1 Feb 2021] (2021)
51. County, C.: Fire district shapefiles, carver county's open data portal. https://data-carver.opendata.arcgis.com/datasets/administrative-fire-districts, [Accessed 1 November 2020] (2020)
52. Council, C.C.: City township boundary, carver county's open data portal. https://data-carver.opendata.arcgis.com/datasets/administrative-city-township-boundaries, [Accessed 1 November 2020] (2020)
53. Pedregosa, F., Varoquaux, G., Gramfort, A., et al.: Scikit-learn: machine learning in python. J. Mach. Learn. Res. **12**, 2825–2830 (2011)
54. James, G., Witten, D., Hastie, T., et al.: An Introduction to Statistical Learning, vol. 112. Springer (2013)
55. Pebesma, E.: Simple Features for R: Standardized Support for Spatial Vector Data. R Journal **10**(1), 439–446. https://doi.org/10.32614/RJ-2018-009 (2018)
56. Studio, I.I.C.O. (2020) V20.1. International Business Machines Corp, Armonk, USA