Predictive Modeling of Building Fire Risk

Metro21: Smart Cities Initiative

Predictive Modeling of Building Fire Risk

Designing and evaluating predictive models of fire risk to prioritize property fire inspections

A METRO21 RESEARCH PROJECT, CARNEGIE MELLON UNIVERSITY

METRO21: SMART CITIES INITIATIVE
The Metro21: Smart Cities Initiative takes a forward-looking creative approach to bringing people, policy and technology together to significantly improve the quality of life for metropolitan area citizens. The multidisciplinary effort employs research, development and deployment tactics with key partners to create and test smart city solutions. CMU's collaborative relationship with the City of Pittsburgh, Allegheny County and other government agencies has produced successful outcomes that are already being implemented in additional metro areas. Cities worldwide face the same imminent challenges. The work of Metro21 provides a global model for innovative future cities.

ACKNOWLEDGMENTS
This was a Metro21: Smart Cities Initiative project, conducted under the supervision of Michael Madaio from January to October of 2017, with generous financial support provided by the Hillman Foundation. Special thanks are due to our partners at the Bureau of Fire: Chief Darryl Jones, Chief Norman Auvil, Chief Thomas Cook, Lt. Jason Batts, Inspector Chris Skertich, and our partners at the Department of Innovation and Performance: Laura Meixell, Geoffrey Arnold, and Maxwell Cercone, among many others. Student contributors to this project were: Qianyi Hu and Bhavkaran Singh, Fangyan Chen, Jeffrey Chen, Nathan Kuo, Jessica Lee, Palak Narang, Amaya Taylor, and Sophia Yoo. Thank you to Rick Stafford and Jeff Chen for the stimulating discussion and support.

AUTHOR: Michael A. Madaio

CODE REPOSITORY: https://github.com/CityofPittsburgh/fire_risk_analysis

SUGGESTED CITATION: Metro21: Smart Cities Initiative (2018). Predictive Modeling of Building Fire Risk: Designing and evaluating predictive models of fire risk to prioritize property fire inspections. *Metro21 Research Publication.*

For inquiries about the contents of this report or the code in the repository, contact: mmadaio@cs.cmu.edu.
For general inquiries about Metro21, media relations, or partnership opportunities please contact: karenlightman@cmu.edu.

www.metro21.cmu.edu

Contents

1. Executive Summary

1.1 Motivation

In 2016 alone, there were 475,500 structure fires in the United States, causing 2,950 civilian deaths, 12,775 civilian injuries, and $7.9 billion in property damage [1]. Municipal fire departments, as the Authority Having Jurisdiction (AHJ), are responsible for enforcing applicable fire codes to reduce the risk of structure fires. The City of Pittsburgh's Bureau of Fire (PBF), like many other municipal fire departments, conducts regular inspections of various types of commercial, industrial, and governmental (here, "non-residential") properties to ensure that they comply with the city's Code of Ordinances [12] for fire prevention and safety. Such inspections are intended to detect code violations that, if left uncaught, may result in loss of life, as seen in recent high-profile fire incidents such as the Grenfell Towers in London [3], the "Ghost Ship" warehouse fire in Oakland in 2016 [15], or the Deutsche Bank fire in New York in 2007 [6]. However, many municipal fire departments are unable to inspect all commercial properties in their city on an annual basis [13], and must therefore determine which subset of properties warrants annual inspection. To determine this, many fire departments rely on a legacy system of inspections conducted on the basis of pre-existing permits, or, at best, a rule-based heuristic for determining which properties warrant inspection. However, these existing approaches to determining inspection priority do not involve a risk-based evaluation of the fire risk of the properties to be inspected. As such, existing approaches to fire inspection could be significantly enhanced by the adoption of risk-based data-driven processes for identification, selection, and prioritization of new properties to inspect [11, 10].

1.2 Methodology

To address this gap in fire inspection prioritization, we have developed a predictive model to determine property-level fire risk (i.e. likelihood of a given property having a fire incident in a given 6-month time period). To develop our model, we use data from historical fire incidents from 2009-2017, property assessment and valuation data from Allegheny County, and non-fire-related inspections and violations (e.g. noise and sanitation) from the City of Pittsburgh Department of Permits, Licenses, and Inspections. We join these data at an address level, and use machine learning models to predict the likelihood of a fire for a given address for a 6-month window, by "training" the model on the first 7.5 years of data on historical fire incidents, and evaluating the model on a test set which was not used for training. The output of this predictive model is a set of risk scores from 1-10 (lowest risk to highest) for the 20,806 non-residential properties in the City of Pittsburgh. We have displayed these results on a private instance of the city's geo-information visualization tool, "Burgh's Eye View," for fire inspectors and fire chiefs to use to identify high-risk properties of various property usage types. We also provided Pittsburgh Bureau of Fire with a dashboard to inspect the fire risk scores, to easily compare risk scores of neighborhoods, fire districts, or property types, and output those data as an image or spreadsheet with detailed property information to use in their strategic planning.

1.3 Findings

Our predictive model is able to predict the presence of a fire incident (any code 100) in a particular address for a given 6-month window with a predictive performance better than the prior state-of-the-art. For any given 6-month window, we are able to accurately detect over half (55%) of the fire incidents that occurred (aka, the "recall" metric of the model). Considering the small number of code-100 fire incidents in a 6-month window (~50) compared to the number of non-residential properties in the city (~20,000), if we were to guess randomly, we would only be correct 0.21% of the time, instead of 55%. As this is the first risk-based model to be deployed in Pittsburgh (and, to our knowledge, the 3rd in the country, after New York and Atlanta), we cannot compare this performance to PBF's existing risk assessment. At the time of this writing, we find that 164 properties are classified as high risk (score 7-10), 596 as medium risk (score 4-6), and the remaining properties as low risk (score 1-3). The most common property types in the high-risk properties are apartments of 40+ units (18 properties), apartments of 5-19 units (15 properties), properties owned by the Board of Education (14 properties), charitable exemption (9), and independent living for seniors (8). Our predictive model also identified the most predictive features of fire risk, which included unintentional alarm or smoke detector activation, gas leaks, or electrical wiring problems in the 6-months preceding a predicted fire, as well as the lot area and fair market building value of the property, and other features.

1.4 Implications

The fire risk prediction model is currently deployed to the Pittsburgh Bureau of Fire's server, where it re-runs every week, generating new risk scores based on the most up-to-date fire and property data. PBF fire chiefs and inspectors are able to use the fire risk scores to inform their day-to-day inspections and high-level inspection planning. As with any predictive model, the fire risk scores are an estimate of the likelihood for a particular outcome (i.e. fire incident) to be true, and contain some uncertainty (discussed in more detail in Section 7). These risk scores are intended to be used as part of a comprehensive process of risk assessment and decision-making, which may incorporate not only the likelihood of a fire occurring, but also the potential for loss of life or property damage. Therefore, these risk scores should complement, not replace, existing PBF risk assessment and strategic planning for fire inspection, as there may be factors unknown to the model (e.g. the mobility of residents, as in independent living facilities for seniors) that may be critical to informing fire inspection procedures. We believe predictive models for fire risk can be an essential asset to a modern, data-driven approach to fire risk reduction in urban environments, augmenting the practitioner knowledge of the fire inspectors and fire chiefs and helping contribute to greater municipal public safety.

2. Related Work

2.1 Fire Inspection Prioritization

As described in Section 1, the number of inspectable properties (i.e. non-residential) in urban areas far outstrips the capacity of most municipal fire departments to conduct regular fire inspections of all such properties. For a clear visualization of this disparity, see Figures 2.1 and 2.2 to see, respectively, the distribution of currently inspected properties in Pittsburgh (in green), and the full set of potentially inspectable (i.e. non-residential) properties in Pittsburgh (in blue). However, not all of these non-residential properties require inspection, and certainly not all with the same level of priority.

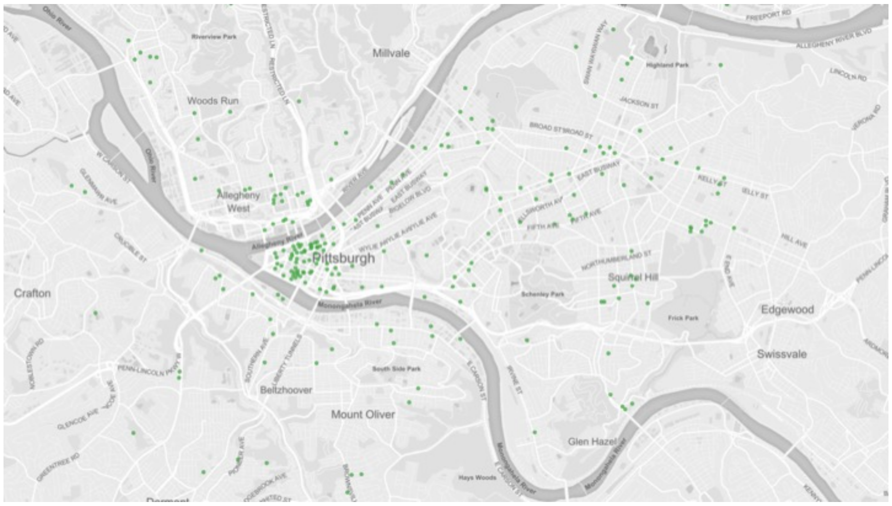


Figure 2.1: Inspected properties in Pittsburgh (in 2016)

Therefore, municipal fire departments face a choice about how to best prioritize their inspection efforts. To do this, many fire departments rely on stipulations in their municipal fire codes, typically

Figure 2.2: Non-residential properties (Commercial, Governmental, Industrial, Utility) in Pittsburgh

an adaptation of the National Fire Protection Association (NFPA)'s fire code, which identifies the property usage types (e.g. restaurants, apartments, etc) or the characteristics (e.g. contains flammable or explosive materials, etc) of the properties requiring inspection [12]. However, the fire code does not provide inspectors with a method for determining which properties from those usage types require inspection at greater priority or frequency. As a result, municipal fire departments are often left with a legacy inspection protocol where properties are inspected only after they have previously had fire incidents, or where inspectors have some reason to believe they may have some fire safety violations [11].

Prior work from Garis and Clare (2014) has developed a set of heuristics for determining the frequency of commercial property fire inspections, using characteristics about properties under consideration [9]. They scored each property by its level of compliance on prior inspections and by a set of risk metric components such as building classification, age, and presence of sprinklers. However, as they acknowledge, the weights and selection of those components were chosen by hand based on their fire code, and not based on historical data about features that were highly predictive of fires, which we utilize in our work. This approach, while better than a legacy approach to inspection, or one without any frequency prioritization at all, may be subject to bias from the creators of the handcrafted weights and rules, and is not likely to be flexible and improve over time with new data.

2.2 Urban Fire Risk Analysis

To address this gap, several municipal fire departments have begun to adopt risk-based inspection practices, using available data of historical fire incidents and property conditions to build predictive models of structural fire risk through data science and machine learning methods. See Figure 2.3 for a visualization of the building fires in Pittsburgh over the last 8 years.

Prior work in data-driven urban fire risk analysis, such as [5, 8], has often conducted their analyses at the regional or census block level, rather than the property- or address-level, which is the level that most municipal fire inspectors are assigned to inspect. For instance, [5] undertook a randomized controlled trial of community fire risk education efforts, to experimentally determine risk levels of residential communities. However, their method for identifying the high-risk areas

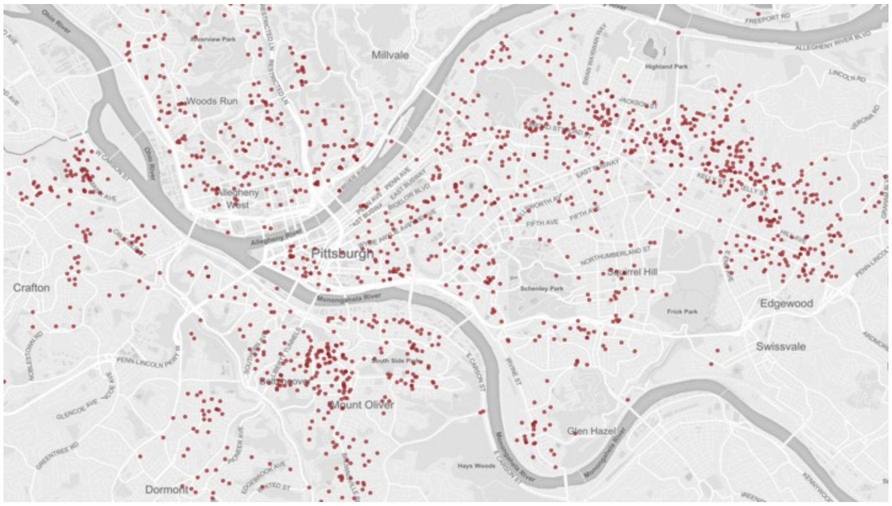


Figure 2.3: 8 years of building fires in Pittsburgh (2009-2017)

was to create a map of the distribution of residential structure fires and draw ellipses to capture the areas of densest concentration of fire incidents. A more statistical approach, as seen in [8]'s work on optimizing smoke-alarm inspections, joins data from the American Community Survey and American Housing Survey to predict census blocks most likely to have homes without functioning smoke alarms, using a Random Forest algorithm.

We differ from [8] in that our model generates a fire risk score instead of the likelihood of not having a smoke alarm, we generate a risk score for individual properties, rather than census blocks, and we target non-residential properties that are able to be inspected, rather than residential (see Section 6 for a discussion about residential risk prediction efforts).

2.3 Risk-based Fire Inspections

The nearest precedents for our work here are (1) the "Risk-Based Inspection System" developed by the New York Mayor's Office of Data Analytics with the Fire Department of New York (FDNY) [6] and (2) the "Firebird" fire risk prediction framework developed by the Data Science for Social Good program in partnership with the Atlanta Fire Rescue Department (AFRD) [10, 11].

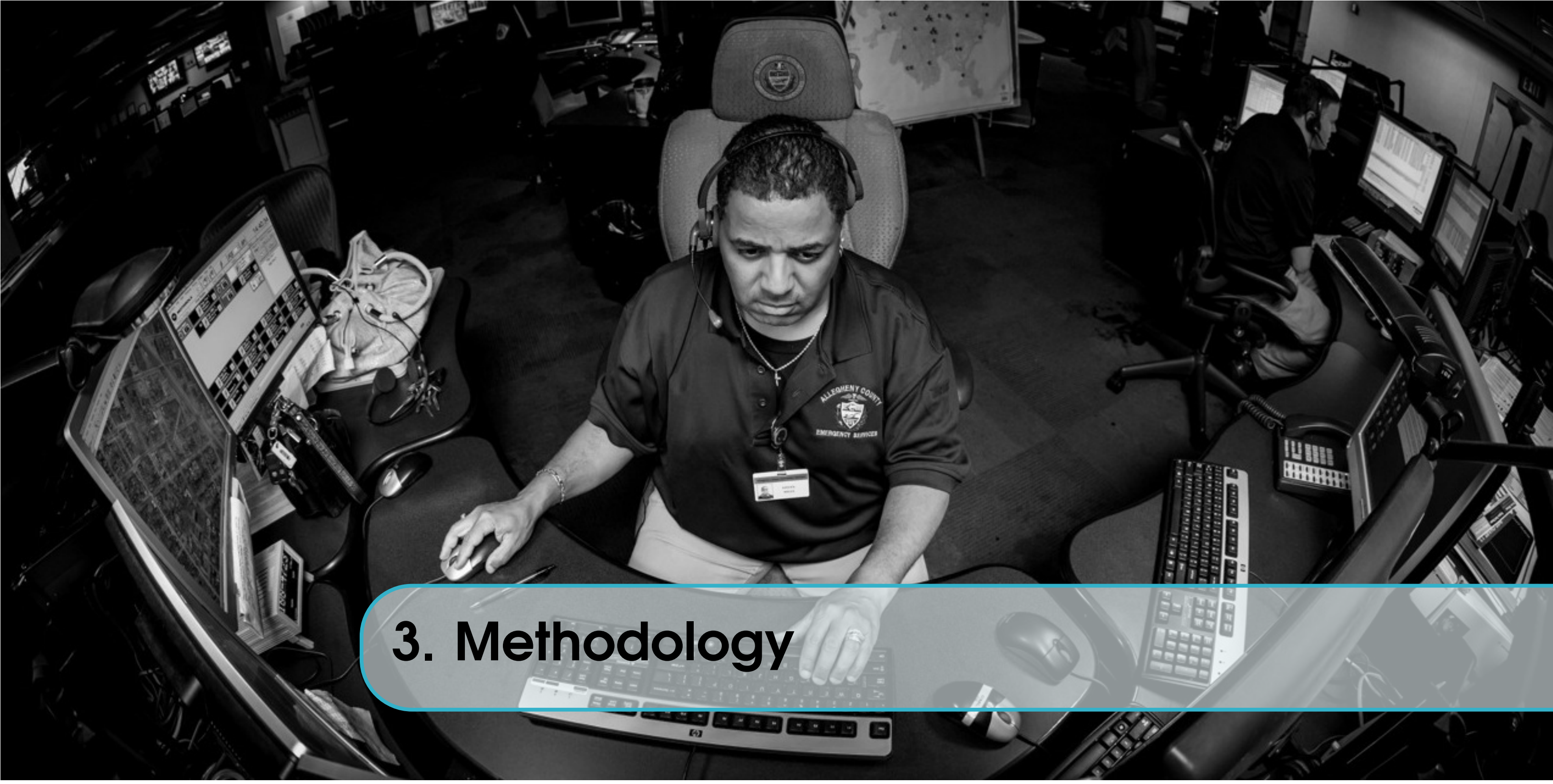
In New York City, in response to high-profile fire incidents such as the Deutsche Bank fire, the Mayor's Office of Data Analytics launched a "risk-based inspection system" in 2013 using data from structural features and behavioral indicators to predict the fire risk of a building and prioritize the property inspections appropriately [6]. They built a data-driven model to identify structures at greatest fire risk, to better prioritize FDNY's inspection process, using a set of structural and behavioral information about those properties. However, because the model was proprietary, and no details about its model construction, accuracy, or other model performance metrics were made public, it is difficult for other municipalities to benefit from this work or to compare their work against a performance benchmark from FDNY.

In 2015, the Atlanta Fire Rescue Department (AFRD) partnered with Georgia Tech's Data Science for Social Good (DSSG) program to develop "Firebird", an open-source framework for identifying and prioritizing commercial property inspections [10, 11]. Atlanta faced a challenge similar to many municipalities – the need for a reliable, accurate method for prioritizing the commercial properties requiring inspection, based on the Fire Code of Ordinances' requirements and using data from their Office of Buildings' business licenses, among other sources. Such a database would provide a foundation for prioritizing the necessary commercial property inspections. The DSSG team developed a predictive risk model based on 1) historical fire incident data from 5 years of fires, and 2) commercial property data collected from their Office of Buildings and a commercial real estate property data set (from the CoStar property group). This model was designed for AFRD's Community Risk Reduction Section to use for human resource allocation and strategic inspection planning, as well as for their front-line fire inspectors to plan their monthly and daily commercial property inspections, based on the risk score assigned by the predictive model.

However, their model was developed in summer 2015, and was not designed to operate on dynamic, temporal data, such as that included in other AFRD fire incident codes (i.e. fire incidents not of code 100 (building fires)), non-fire inspection violations (e.g. noise or sanitation violations), or 311 requests, etc. Due to the highly dynamic nature of much civic data, even purportedly "static" data such as property size, assessed value, property condition, or even property usage data may change over time as properties are bought, sold, renovated, and closed. Therefore, without incorporating temporal data, the Firebird model was unable to be updated on a regular basis, and as such provided merely a snapshot in time of the risk levels of the summer 2015, which are likely to be out of date soon after generation. To address this, our model is deployed on the Pittsburgh Bureau of Fire's server, taking in new data from various sources on a weekly basis and retraining the model, generating new risk scores every week.

Other recent related work includes work from Jonathan Jay, who used open civic data to build a predictive model of residential fire risk in Baton Rouge[1] and conducted analyses of fire incidents in Boston[2]. Others, like the Code for San Francisco's data science brigade, have used the City of San Francisco's open data about fire incidents to conduct analyses of fire data and visualize it using CartoDB[3]. The author of this report has been in discussions with researchers from both of those projects over the duration of this work, discussing best practices for model construction and evaluation. We are encouraged that researchers in other municipalities are developing predictive models for fire risk to improve public safety. However, both of these prior works have not had the benefit of a partnership with their municipal fire safety department, as we have in Pittsburgh. Without this partnership, their work provides insight into fire risk factors, but may not enable their work to have the impact on municipal public safety efforts that it might otherwise have.

[1] https://scholar.harvard.edu/jonjay/blog/how-we-predicted-building-fires-baton-rouge-la-working-version
[2] http://rpubs.com/jonjay/FPBH$_JJ$
[3] http://codeforsanfrancisco.org/projects/SF-Fire-Risk-Project



3. Methodology

3.1 Data Sets

We start by acquiring data sets from a variety of sources that contain data we hypothesize (based on prior work) to be relevant in predicting fire risk in non-residential properties. We start with fire incident and fire inspection data, provided by the Pittsburgh Bureau of Fire, from 2009-2017, updated on a weekly basis, of which we include all fire incident codes that have an associated address. From the Allegheny County Office of Property Assessments, we use the property assessment data, updated on a monthly basis. From the Pittsburgh Department of Permits, Licenses, and Inspections (PLI), we use their record of non-fire inspections and violations (e.g. noise or sanitation violations). Finally, from Allegheny County, we use the parcel database, which contains information about every parcel in the City of Pittsburgh. More detail about the data sets can be found in Table 3.1, with the number of records, features, and dates up-to-date as of the time of this writing.

3.2 Data Pre-Processing

To obtain the most complete set of information about the non-residential properties, we joined the above data sets for each non-residential address in our dataset. Each address may align with a single parcel, or one address may contain multiple parcels, as in, for example, a condominium, or university with multiple buildings at the same address.

Table 3.1: List of data sources used in building the risk model

Data Set	Source	Features	Records	Dates	Updated
Fire Incidents	Bureau of Fire	27	387,264	2009-2017	Weekly
Fire Inspections	Bureau of Fire	24	1483	2015-2017	Weekly
Non-Fire Violations	PLI	20	13,892	2015-2016	Daily
Parcels	Allegheny County	46	579,474	2017	Weekly
Non-Residential Properties	Allegheny County	80	37,268	2017	Monthly

Although the most granular unit of analysis would be the parcel, the fire incidents were logged at the address level, and thus, to predict fire risk, all of the other data first needed to be aggregated across the individual parcels in each address. See Figure 3.1 for a visualization of the unit of analysis associated with each data set.

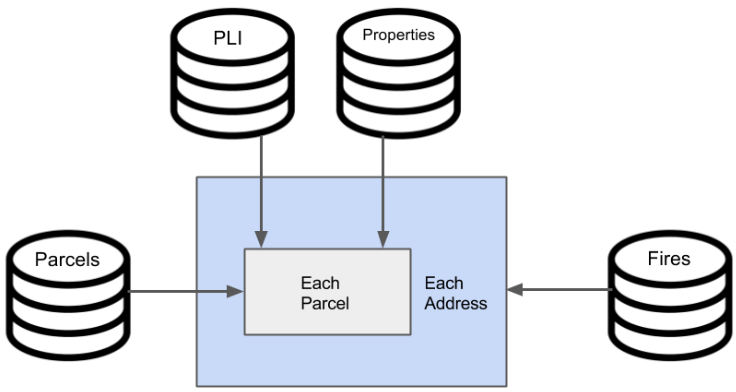


Figure 3.1: Data aggregation levels

To do this, we start with the Allegheny County property assessment data set, which we obtained from the Western Pennsylvania Regional Data Center (WPRDC)[1]. We first drop all of the data listed as Residential class properties, since we focus in this work on non-residential property inspections. We then merge the PLI non-fire inspection violations dataset with the non-residential properties (by which we mean Commercial, Industrial, Governmental, or Utility), at the parcel level. After minor cleaning such as stripping white spaces from text values, dropping duplicate columns, and dropping rows with significant (85%) missing values for data, we aggregate the parcel data at the address level. We aggregate by taking the mean of the following features:

- Lot area
- Sale price
- Fair market land value
- Fair market building value

Then, because many features are categorical in nature, we aggregated the following features by using the most frequently occurring category within each address:

- Class description (e.g. commercial, industrial, etc)
- Property usage description (e.g. restaurant, church, etc)
- Neighborhood
- School district
- Municipality
- Owner description
- Tax subcode description
- Non-fire inspection result (from PLI)

[1] www.wprdc.org

We then merged the resulting aggregated address-level dataframe with the fire incidents dataframe. We performed minor cleaning of the fire incident data (e.g. stripping white space, standardizing hyphens, standardizing street abbreviations, etc). We then removed some columns and some fire incident types not relevant for property risk analysis (e.g. overheated motor, authorized controlled burning, etc).

3.3 Model Construction

After joining all of our data sets together at the address level, we were left with a single table which we used to train and test the risk model. For this analysis, we used all fire incidents of a 100-level code (i.e. building fires) as the outcome to predict, and all other features were used as predictive features in the model. Because some of the features were events (e.g. fire incidents and non-fire inspection violations), we restructured the data so that only events that occurred prior to the fire in question were used as a predictive feature. That is, each row of the dataframe was an address, which may have had a fire incident in any given year in the 8 years of our data. Each fire incident had an associated year, with some addresses having multiple fire incidents across the 8 years, and thus, multiple rows in the dataframe. For each address-incident, we only included the PLI violations and non-100 incidents (e.g. smoke alarm activation, electrical wiring issues, etc) as an entry in that address-incident row if the datetime of that event occurred prior to the datetime of the code-100 fire incident to be predicted. We then one-hot encoded all of the categorical features as dummy variables, and divided our data into a training set (7.5 years of data) for cross-validation, and a test set (the final 0.5 years of data).

3.4 Model Comparison

After training, the model returns an evaluation of its performance on the held-out 6 months of the test set, using the following metrics:

- Kappa score
- AUC score
- Recall
- Precision

We evaluated multiple model types (described in Table 3.2) along the metrics described above. Because of the substantial class imbalance in our data (44 fire incidents out of 20,806 addresses for the test set, in July 2017), accuracy is not a useful measure of model performance. That is, because if the model simply assigns "no fire" to every instance, it will be correct 99.8% of the time. Therefore, the kappa score, which is, intuitively, the percent agreement accounting for the class imbalance, is a more useful metric. We also use the "recall" to evaluate the performance of our classifier.

Before discussing the model evaluations, we will say a bit about how to evaluate a predictive classification model such as this. There are a certain number of actual events (fires), and a certain number of predicted events (predicted fires). A perfect classifier would predict all actual events (in the test set), while not erroneously predicting any other events (here, fires). In Figure 3.2, the (perhaps aptly named) "confusion matrix" shows the relationship of the actual to predicted events.

Ideally, a predictive model will maximize the number of true positives (addresses that the model predicted would have a fire, and actually had a fire in the test set) and true negatives, while minimizing the false positives and false negatives. However, in some prediction cases, like fire prediction, for

		Predicted	
		No Fire	Fire
Actual	No Fire	True Negatives	False Positives
	Fire	False Negatives	True Positives

Figure 3.2: Example confusion matrix

instance, false negatives (where the model predicts there will not be a fire, and there actually was) are worse than false positives (model predicted fire, but there was not a fire in the 6 months of the test set). Therefore, we want to evaluate our model not just by how many true positives and true negatives it correctly classifies, but by how *few* false negatives it returns.

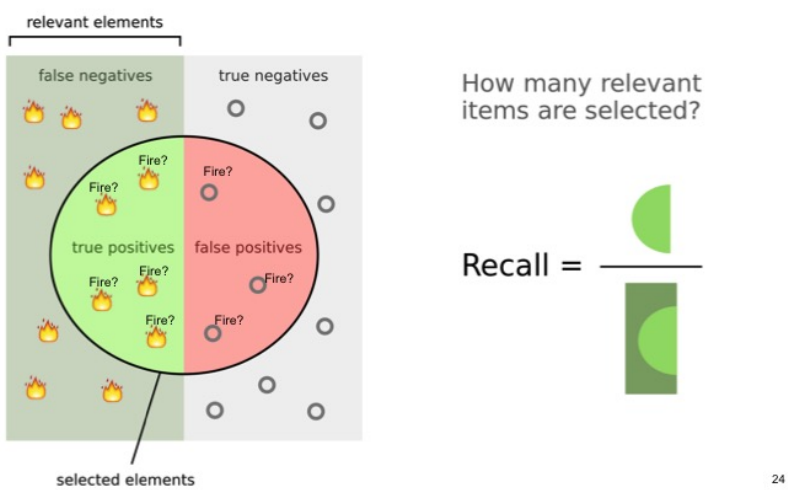


Figure 3.3: Recall: a metric to evaluate classifications

One metric for this is called the "recall." Intuitively, the recall can be thought of as the ratio of the true positives the model predicted to all of the positive instances. See Figure 3.3 for a visualization of how to think about the "recall." This evaluation measure penalizes a model for having too many false negatives. Because for fire prediction, we care more about correctly classifying more of the positive class than minimizing the false positives, it is a more useful evaluation measure for this application than the "precision" (which penalizes the model for having more *false* positives). For thoroughness, we show four evaluation measures (Kappa, AUC (or, area under the curve), recall, and precision. However, we ultimately use the kappa (percent agreement, accounting for chance), and recall as our two main evaluation measures.

Due to the significantly larger kappa score out of all the models we evaluated, we decided to use the XG Boost model as our final model type, which is a variant on the random forest model type. From manual tuning, we determined the optimal hyperparameters were a learning rate of 0.13, 1500

Table 3.2: Evaluation of 4 different model types using a 7:1 train-test split, as of July, 2017

Model Type	**Kappa**	**AUC**	**Recall**	**Precision**
Logistic Regression	0.004	0.53	0.11	0.004
Ada Boost	0.15	0.6	0.2	0.13
Random Forest	0.21	0.76	0.52	0.13
XG Boost	**0.37**	**0.77**	**0.55**	**0.14**

estimators, max depth of 5, 27 seeds, 4 threads, and a binary logistic objective. Future work will use a grid search to identify the optimal hyperparameters.

We then output the prediction probabilities for every address in our data, which is the probability that that address will be the positive class (i.e. fire incident of code 100) in the final 6-month window. A larger probability means that it is more likely that that property will have a code 100 fire incident. In order to have a more interpretable fire risk score for the Bureau of Fire, we convert these probabilities into integers from 1-10. In the following chapter, we discuss how this model compares to other prior work, discuss some preliminary findings from analysis of these risk scores, and discuss some predictive features that were discovered to be most important in the risk model.

4. Findings

4.1 Overview

Using data joined from fire incidents, non-fire inspection violations, and property assessments, we used an XG Boost classification model to predict the likelihood of a fire occurring in a given address. We will first report how well this model predicts property fires compared to other previous approaches, such as [10]. We will then describe some of the predictive features (e.g. property usage type, lot size, etc) that were ranked as highly important in the model's ability to predict fires. Finally, we will describe some initial results of a post-hoc analysis of the high-risk properties.

4.2 Model Performance

In Section 3.4, we described how four different models all trained on the data from Pittsburgh Bureau of Fire compared with each other. Here, we describe how the best performing model ("Metro21 model") compares with the previous state-of-the-art in fire risk prediction, ("Firebird model"), in partnership with the Atlanta Fire Rescue Department [10]. Results are reported in Table 4.1.

Although the Firebird model has a larger recall, and a somewhat larger AUC than the Metro21 model, it is important to note that the Firebird model was only applicable to a subset of properties in the city of Atlanta. Due to high levels of missingness in their data sets, they only generated risk scores for 8,223 non-residential properties, of which only 2,187 were in their test set. While this enabled them to improve their model performance by only generating risk predictions for properties with the most complete data, it did not allow them to apply their fire risk scores to every non-residential property in the city, as reported in [10].

Table 4.1: Model performance of Metro21 risk model and state of the art risk prediction

Model	Kappa	AUC	Recall	Precision	Fires in test set	Properties in test set
Metro21 Fire Risk	0.37	0.77	0.55	0.14	44	20,806
Firebird	0.17	0.8	0.72	0.18	142	2,187

Table 4.2: Most important predictive features and their importance scores

Feature	**Importance Score**
Fire Code 745 - Alarm system activation, no fire	0.12
Fire Code 743 - Smoke detector activation, no fire	0.11
Fire Code 5001 - Smoke Detector Activation, No Fire	0.08
Fair Market Building Value	0.06
Fair Market Land Value	0.06
Lot Area	0.06
Sale Price	0.05
Fire Code 531 - Smoke or odor removal	0.03
Fire Code 522 - Water or steam leak	0.02
Fire Code 651 - Smoke scare, odor of smoke	0.01
Fire Code 412 - Gas leak (natural gas or LPG)	0.01
Fire Code 440 - Electrical wiring/equipment problem, Other	0.008
Owned by Metro Housing Authority 0.008	
Fire Code 445 - Arcing, shorted electrical equipment	0.008
Fire Code 400 - Hazardous condition, Other	0.008
Independent Living (Seniors) 0.007	
Apartment: 40+ Units	0.006
Fire Code 411 - Gasoline or other flammable liquid spill	0.006

Here, our approach uses the most complete data sets we have available on all of the non-residential properties in the city, and while that means that our recall is not quite as large as the Firebird model's, we *are* able to generate a risk score for all ~20,000 non-residential properties in the city. Thus our kappa score (traditionally used as the primary evaluation measure for classification models such as this) is significantly higher. This may be because, while the Firebird work may have labeled more true positives, the likelihood that those 142 fires could have been discovered by chance was greater, due to the smaller data set (10% the number of non-residential properties used in our model). In addition to the significantly higher kappa score, our 0.55 recall for our model is still quite good considering the *actual* distribution of fires in our data set. That is, for any given 6 month period, our model is accurately predicting just over half (55%) of the fires that occurred (in the test set), while if one were to guess randomly, you would guess correctly 0.21% of the time.

4.3 Feature Importance

After running the predictive model, we obtain (1) the model performance measures (described above), (2) the actual risk scores for the ~20,000 properties in our data (described in more detail in the next section), as well as (3) a set of "features" (or, columns in the data set), that were ranked according to their predictive importance in determining the risk level. See Table 4.2 for the top features returned by the model, and their respective "feature importance score."

The most important predictive features our model discovered were when the address in question had an alarm or smoke detector activation prior to the fire incident. Other important predictive features were measures of the address's lot area, assessed value, and sale price. Other significant predictors were other non-code 100 fire incidents, such as smoke removal or smoke scare, steam leak or gas leak, electrical wiring and shorting, hazardous conditions, and gasoline spills, all occurring in the 6-month window prior to the fire incident in question. In addition, several property usage types were important predictive features, such as, properties owned by the Metro Housing Authority, independent living properties, or apartments with 40+ units.

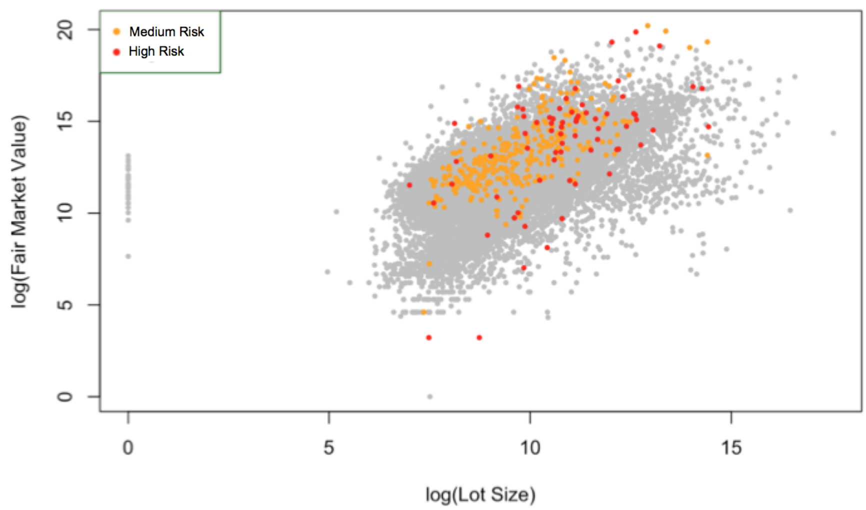


Figure 4.1: Plot of property size and value of medium and high risk properties

To understand in more depth what might be going on with the lot area and sale price, we visualized the interaction of the two features, following Jonathan Jay's[1] finding in his analysis of residential property fire risk in Baton Rouge that there was an interaction between property value and size. Figure 4.1 shows the interaction of these two features plotted (using a logarithmic transformation, due to the large scales). In this visualization, the high-risk properties (red) were those with a risk score of 7-10, medium-risk (orange) those with a risk score of 4-6, and the grey dots low risk (1-3).

4.4 Analysis of High-Risk Properties

Now that we have evaluated how well the model predicts fire risk, compared it to the state of the art, and described the most important features in that predictive model, we will now discuss some preliminary analyses of the properties with high fire risk scores, as determined by the model output as of this writing. See Table 4.3 for the breakdown of how many addresses were assigned each risk score, and Figure 4.2 for a histogram visualization of the risk scores from 2 to 10 (removing scores of 1 due to the significantly larger number of them).

There are significantly more properties with a risk score of 1 (19,509) than any other risk score. Interestingly, there is also a larger number of properties with risk scores of 5 and 6 than of other types. As these were properties with prediction probabilities between 0.4 and 0.6, this may reflect properties with a roughly equal likelihood of being a fire (probability of 1) or not (0).

[1] https://scholar.harvard.edu/jonjay/blog/how-we-predicted-building-fires-baton-rouge-la-working-version

Risk Score	Risk Level	Address Count
10	High	46
9	High	26
8	High	35
7	High	57
6	Medium	152
5	Medium	300
4	Medium	87
3	Low	39
2	Low	450
1	Low	19509

Table 4.3: Property count for each risk score

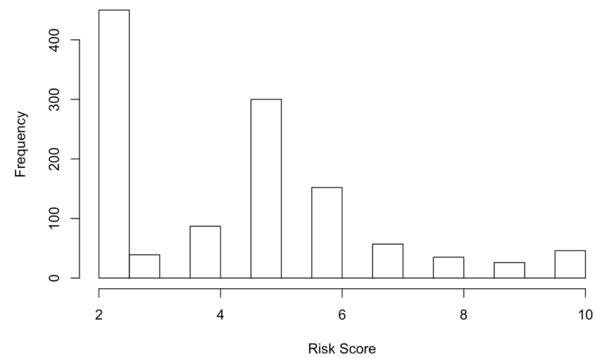


Figure 4.2: Histogram of risk scores (2-10)

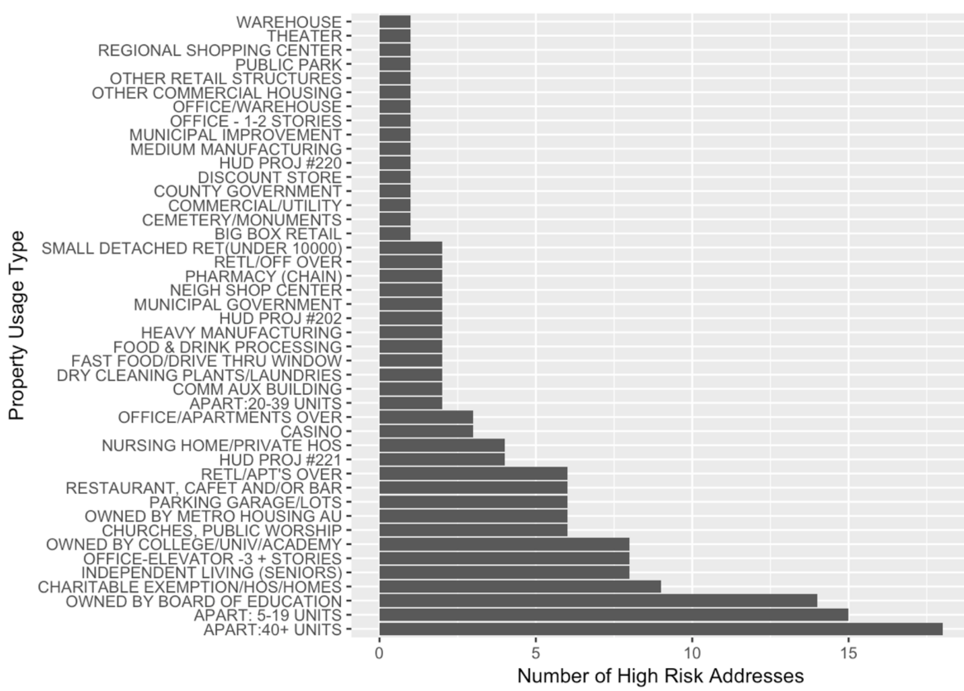


Figure 4.3: High-Risk Property Types

In Figure 4.3, we show the property types which contain the most high-risk properties. However, please note that these are the raw counts of properties with high risk scores (7-10), and thus they have not been adjusted by the total number of those property types in the city. Therefore, while the property types with the most high-risk properties in the city are apartments of 40+ and 5-19 units, there are also many more of those addresses in the city than, for instance, HUD Project #221. Thus, the overall percentage of each property type may be a more useful measure of the relative riskiness of that property type.

As may be expected, the property types which contain the most high-risk properties also showed up as highly important predictive features in Table 4.2. Specifically, the property types of apartments of 40+ units, Independent living (seniors), and Owned by Metro Housing Authority were highly predictive features for fire risk. Those, and other high-risk property types, such as those Owned by Board of Education, Churches, and Restaurants, align with existing initiatives from Pittsburgh Bureau of Fire's inspection program. This serves as a useful measure of "face validity" for the risk scores, such that some of the highly predictive features align with existing expert knowledge and Bureau of Fire risk reduction efforts, in addition to the more traditional risk model performance metrics described above.

5. Deliverables

After developing the fire risk model and evaluating its performance, we provided the Pittsburgh Bureau of Fire with a tangible product they could use to integrate those fire risk scores into their existing inspection decision-making practices. This took the form of a data dashboard and an interactive map, using a platform already incorporated into PBF fire inspectors' existing workflow, called "Burgh's Eye View."

5.1 Deployed Risk Model

Because fire risk levels will change over time as new data about fire incidents and property features are acquired, we first needed to deploy the risk model on the Bureau of Fire's servers, to intake new data on a regular basis, re-"train" the model, and generate the new fire risk scores for each address. We currently have a set of Python scripts in a Github repository[1] that are scheduled to run every weekend, using the time-based job scheduling software utility "cron" to:

- Scrape WPRDC for new versions of 3 datasets
- Join 3 datasets from WPRDC with updated Fire Incident and Fire Inspection data from PBF
- Retrain risk model and generate fire risk score for each address
- Add that risk score to the joined dataset of the addresses and their most complete set of features

As part of our ongoing audit to evaluate the performance of the risk model over time, each time it retrains, it evaluates its performance against the held-out test set of the final 6 months of data. The performance metrics (as in Section 4.2) are logged to evaluate the performance of the model.

5.2 Data Visualizations

To help the Bureau of Fire make use of these risk scores, we provided them with two visualization tools to incorporate the risk scores into their inspection planning and strategic decision-making processes. These take the form of an interactive data dashboard and an interactive map to spatially visualize the properties and their associated fire risk.

[1] https://github.com/CityofPittsburgh/fire_risk_analysis

5.2.1 Data Dashboard

The data dashboard, created using R Shiny, allows inspectors or fire chiefs at the Bureau of Fire to easily display and compare risk levels based on property type, neighborhood, and fire district. They can subset each of these by any of the others, allowing them to compare the high-risk properties of a certain usage type (e.g. restaurants) by neighborhood, to identify the neighborhoods that contain the most high-risk restaurants. See Figure 5.1 for an example of the data dashboard displaying the average risk score by property usage type (with the actual usage types greyed out here).

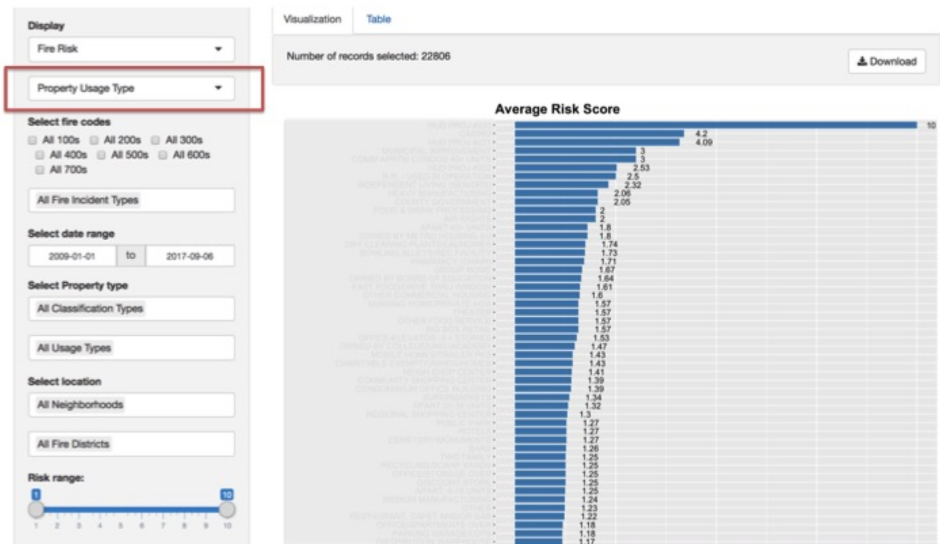


Figure 5.1: Data dashboard displaying property usage types according to their average fire risk score

This same data dashboard also contains data from the historical fire incidents and fire inspections, and can be used for the Bureau of Fire chiefs to reflect on the actual fire incidents of various types (e.g. code 100, 200, etc), across the 8 years of data stored in digital form, using the other filters to visualize the fire risk scores as well. PBF staff can then download the visualization as an image or download the table of property details as a .csv file. See Figure 5.2 for an example of the dashboard displaying historical fire incidents of code 100, according to their property usage type.

5.2.2 Interactive Map

While the bar chart visualizations are helpful for easily comparing multiple types of properties or locations, they don't reflect the spatial nature of this data. Therefore, we plotted the properties onto an interactive map, so fire inspectors and fire chiefs could quickly and easily see the location of high-risk properties of various types. PBF staff already use a version of this interactive map, named "Burgh's Eye View," to visualize their current property inspections. Thus, displaying the risk scores as an additional filterable "layer" here fits within their existing workflow. Fire inspectors can use this map to identify the high-risk properties nearby their assigned inspections, providing better inspection coverage for the riskiest properties.

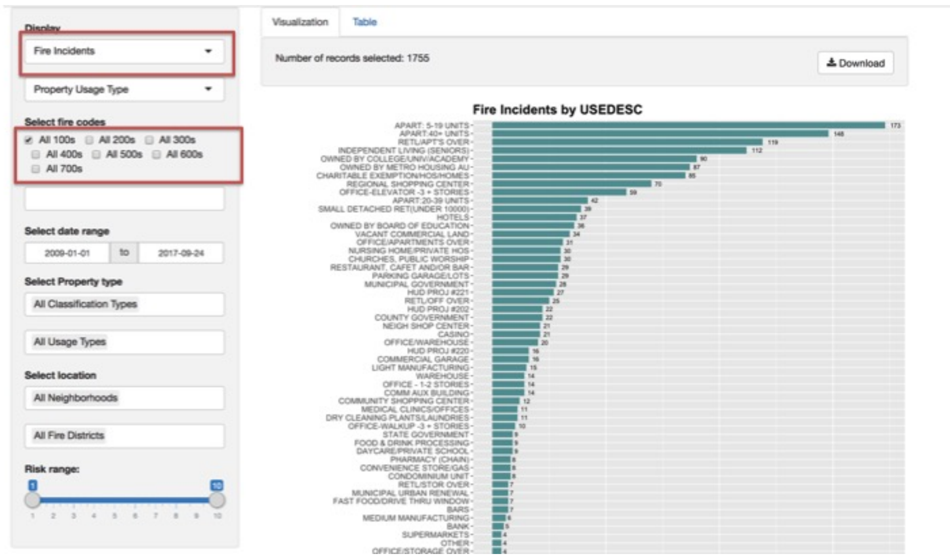


Figure 5.2: Data dashboard displaying property usage types ranked by their Code 100 fire incidents

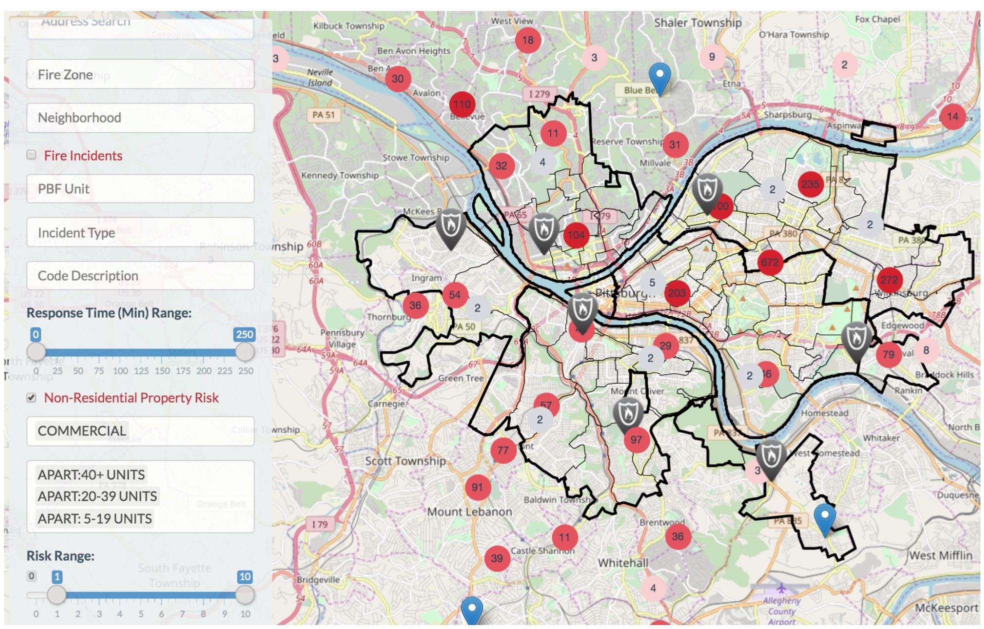


Figure 5.3: Interactive map displaying high-risk properties, filtered by property type

6. Next Steps

Although the first phase of this project is completed, and the model is currently deployed on the Bureau of Fire's servers generating risk scores and updating their data visualizations, members of our team are still hard at work improving and extending the risk model in several ways.

6.1 Improvements to Model Performance

We are in the process of identifying new data sources that prior research [10, 6] has identified as potentially useful in predicting structure fires; acquiring and cleaning those data; joining them to the existing data set; performing feature selection to remove unnecessary features; and re-training the model. These additional data sets may include sources such as unpaid tax lien data, U.S. Census or the American Community Survey (ACS) or American Housing Survey (AHS), 311 service requests, Google Place business data, and other potential data sources.

Additionally, our team is working on improving the existing risk model through experimenting with novel model types. For instance, ensemble models of multiple classifiers may work better than single models on their own. Alternatively, our team is experimenting with models that encode temporality themselves, without needing to hard-code the temporal features in the way we've done here. For instance, sequential neural network models, such as recurrent neural networks (RNN), or their variant, long-short term memory (LSTM) networks, have been shown to outperform other model types on sequential prediction problems (e.g. language generation, translation, etc). Some prior work [16] has explored using a "point-RNN" for spatio-temporal event prediction, work which may inform some of our future directions in this vein. Finally, as our model is set up for a batch re-training every week, we are planning on experimenting with an active learning or reinforcement learning paradigm, where the model's *actual* performance in predicting fire incidents in the subsequent time window (not just measured against the held-out test set) can be used as a "reward" (in the reinforcement learning parlance) to improve the model when it more accurately predicts fire incidents.

6.2 Extensions to Residential Properties

In addition to improving the predictive performance of the model, we plan on extending this work from predicting fire risk in non-residential properties (to prioritize fire inspections), to predicting fire risk in *residential* properties, in order to inform the Bureau of Fire's Community Risk Reduction and fire safety educational efforts. As described in Section 2, other researchers have adopted the risk model framework from Firebird [10] to model fire risk in residential properties in Baton Rouge, though to the best of our knowledge, they are not partnering with the Fire Department there to integrate the results of the model into the fire department's community risk reduction efforts.

Therefore, a key component of our future work on residential fire risk prediction will not only be to generate the risk scores, but to identify the appropriate level of aggregation for analysis and visualization. Because the Pittsburgh Bureau of Fire, like most municipal fire departments, does not conduct inspections of single family homes, the generation of a risk score at an address-level is not useful in informing their subsequent strategic planning. More useful, perhaps, might be the identification of a high-risk census block, census tract, or neighborhood, where the Bureau of Fire could target Community Risk Reduction efforts like fire safety classes or demonstrations. Or, looking beyond existing municipal subdivisions, one might adopt a clustering approach to identify key centroids of the largest clusters of high-risk residential addresses, to pinpoint the optimal location where a community fire safety event might have the best chance at reducing the fire risk of that area.

7. Discussion

In this report, we have described our process for modeling structural fire risk for non-residential properties, based on data about historical fire incidents, non-fire inspections and violations, property assessments, property characteristics (size, sale price), and parcel data. We described how we used these features to train a predictive model, evaluated that model against other model types, achieving the current state-of-the-art predictive performance, and deployed that model at the Bureau of Fire, to update on a weekly basis and display the risk scores on a data dashboard and an interactive map. In this final section, we will describe some of the implications of this work for other civic agencies that may use machine learning and risk modeling as part of their workflow.

7.0.1 Risk

In our approach, we have utilized one method for risk modeling - that of identifying the likelihood of an adverse event (i.e. fire incident) occurring. However, estimating the likelihood of an event is not the only way to model risk. One might also model the severity of the event, were it to occur. One example of this is type of work is the work from Xu et al. [17], where they predict the likelihood of freeway crashes at several levels of severity. Our fire risk model is being used to inform the prioritization of fire inspections; thus, to mitigate loss of life and property damage in the event of a fire, they may want to incorporate aspects of risk that reflect the severity of the fire, in addition to just its likelihood of occurrence. This may be measured by the number of "alarms" the fire had, or data on the loss of life or amount of property damage. It is along these lines of multi-faceted risk analysis, that the Bureau of Fire has targeted inspection efforts at Independent Living (senior) properties, due to the limited mobility of the residents, and thus the potential for severe loss of life in the event of a fire. See [2] for a larger discussion about considering external factors in developing and applying predictive modeling for public policy.

7.0.2 Trust

We thus present these results with a word of caution. It is not our intent for the risk scores to entirely replace the decision-making of the Bureau of Fire chiefs and fire inspectors. Rather, we intend for them to augment existing decision-making practices, with a model that predicts (with some uncertainty!) the likelihood of a fire incident. In meetings with the fire chiefs at Pittsburgh's Bureau of Fire, we have discussed with them the uncertainties involved in predictive modeling, in part through explanations about the meaning of the false positives and false negatives in the model.

An example here may be instructive. Two municipal health and human services agencies, one in Allegheny County, PA[1], and another in the city of Chicago[2], have developed risk models to identify children in their welfare system at greatest risk of various outcomes (i.e. child abuse, or even death, in the case of Chicago). At the Allegheny County Department of Health and Human Services, their staff incorporated the risk scores as one more factor in their caseload decision-making process, not the only factor. In Chicago, on the other hand, it seems that decisions were made about which children to assist based entirely on the scores from the risk model, unfortunately leading to interventions not being taken with other children not identified as high-risk. When several children died that were not marked as high-risk by the model, this led to questions about the validity and reliability of their model, the approach, and its appropriateness in the department of health and human services. See [4, 7, 14] for a review of child welfare predictive analytics, and the ways that such models are incorporated into municipal decision-making.

These are valid questions to ask. No predictive model will be perfect, and it is important to communicate the relative rates of false positives (cases where the model identified something as at-risk when they weren't) and false negatives (cases where the model identified something as *not* at-risk, when they were). It is also important to appropriately stress that the results of the model should not be the sole determinant in civic decision-making. They should instead be integrated into a process that respects decision-makers' expertise and experience, as well as other factors not incorporated into the model (e.g. the mobility of the seniors in assisted living centers).

7.1 Conclusion

We intend for this work to be useful for municipal Fire Departments, for Code for America brigades or analogous data science for social good organizations, or, more broadly, for any civic agency looking to incorporate data-driven predictive modeling into their decision-making processes. For non-government organizations, we want to stress again the importance of working closely with municipal partners. We have benefited from a fruitful, productive partnership with the Pittsburgh Bureau of Fire, as well as with the Pittsburgh Department of Innovation and Performance, without whose support this work would not have been possible. We highly recommend you establish such partnerships at the outset, to access and understand the data, design the model or tool to augment existing decision-making processes, and communicate the various uncertainties to your partners. For representatives of municipal departments without in-house data science capacity, you may consider reaching out to researchers from local universities or student data science organizations such as Students for Urban Data Systems at CMU[3], or others, such as University of Michigan's MDST[4].

[1] https://chronicleofsocialchange.org/news-2/pennsylvania-county-leads-globe-uses-big-data-stem-child-abuse-not-without-probing-ethics-first/18290

[2] http://www.chicagotribune.com/news/watchdog/ct-dcfs-eckerd-met-20171206-story.html

[3] suds-cmu.org

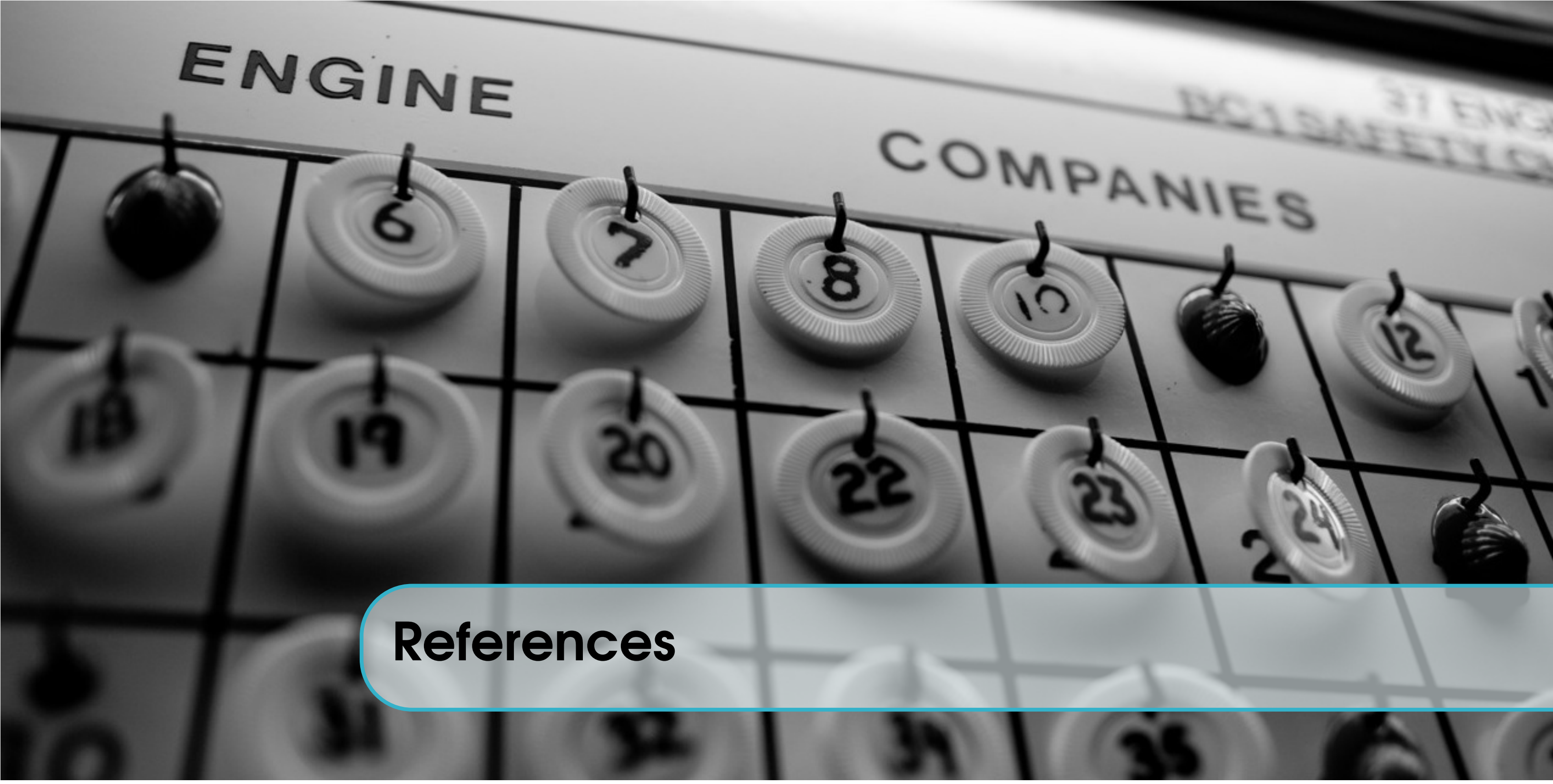
[4] http://midas.umich.edu/mdst/

In the spirit of openness and transparency, and to make it more likely that this approach could be adopted and re-used in other cities, we have provided all of our code in an open-source repository[5]. Other data scientists and municipal agencies can "fork" this repository to make use of this model for their cities, using their own data. While we would have liked to share the data as well as the code, not all of the data used in our model is publicly available. The Western Pennsylvania Regional Data Center (WPRDC) has worked closely with Allegheny County and the City of Pittsburgh to make many civic data sets available to the public (e.g. parcel data, property assessments, etc). Some data sets, however, such as the fire incident data, have been modified to remove individual addresses from the data and aggregate the incidents at the block level.

We are in the process of working with the Pittsburgh Bureau of Fire to determine the most appropriate level of granularity at which to make the risk scores available to the public. While we believe that transparency of government processes is important, we also want to protect property owners from unintended adverse effects that might arise if fire risk scores were made public for individual properties. These adverse effects might take the form of repercussions from members of the public who may be less likely to frequent a particular business (e.g. restaurant or bar), if they had access to the fire risk score, which contains uncertainty, and which may change over time. An additional concern is if private companies involved with property data, such as Zillow for residential properties, or property insurance companies, were to gain access to the risk scores at an individual property level. This is an issue that warrants further discussion about the ethical and legal considerations of predictive risk modeling on open civic data, a discussion we are eager to engage in. With that in mind, we made the Metro21 risk model itself (i.e. the source code) open-source, rather than the input data (fire incidents) or output risk scores, to allow the risk model to be adapted to benefit other municipalities that may not have the technical capacity to develop it on their own.

We hope that this work, and this report, will prove useful to other municipalities, civic agencies, and civil society organizations interested in using data to improve public safety and the provision of social services such as fire risk reduction. We invite you to adopt, adapt, extend, and improve the fire risk model described in this report to improve public safety in your municipality. We also invite a larger public discussion about the ways to mitigate the risks and tradeoffs of inherently uncertain predictive models and integrate their output into civic processes while not jeopardizing the privacy and data security of property owners and residents that may be at risk through public disclosure of data. By integrating a data-driven approach to fire risk modeling into existing legacy approaches to community risk reduction, we intend to contribute to the safety and security of municipal residents, particularly those who may have been overlooked by historical inspection practices. Ultimately, we intend to contribute to the larger body of work on incorporating data science and machine learning into improving civic processes in a democratic, transparent way.

[5] https://github.com/CityofPittsburgh/fire_risk_analysis



References

[1] National Fire Protection Association. Oct. 2017. URL: `http://www.nfpa.org/News-and-Research/Fire-statistics-and-reports/Fire-statistics/Fires-in-the-US`.

[2] Susan Athey. "Beyond prediction: Using big data for policy problems". In: *Science* 355.6324 (2017), pp. 483–485.

[3] John Murray Brown. *Grenfell fire tragedy prompts inspection of 4,000 UK tower blocks*. Financial Times. June 2017. URL: `https://www.ft.com/content/a6d17890-526b-11e7-a1f2-db19572361bb`.

[4] Christopher E Church and Amanda J Fairchild. "In Search of a Silver Bullet: Child Welfare's Embrace of Predictive Analytics". In: *Juvenile and Family Court Journal* 68.1 (2017), pp. 67–81.

[5] Joseph Clare, Len Garis, Darryl Plecas, and Charles Jennings. "Reduced frequency and severity of residential fires following delivery of fire prevention education by on-duty fire fighters: Cluster randomized controlled study". In: *Journal of safety research* 43.2 (2012), pp. 123–128.

[6] Eddie Copeland. "BIG DATA IN THE BIG APPLE". In: *Capital City Foundation.* (2015). URL: `http://capitalcityfoundation.london/wp-content/uploads/2015/06/Big-Data-in-the-Big-Apple.pdf`.

[7] Stephanie Cuccaro-Alamin, Regan Foust, Rhema Vaithianathan, and Emily Putnam-Hornstein. "Risk assessment and decision making in child protective services: Predictive risk modeling in context". In: *Children and Youth Services Review* (2017).

[8] Marc DaCosta, Jeremy Krinsley, and Brian Abelson. "Optimizing Local Smoke Alarm Inspections with Federal Data". In: *Bloomberg Data for Good Exchange* (2015).

[9] Len Garis and Joseph Clare. "A Dynamic Risk–Based Framework for Redesigning the Scheduling of Fire Safety Inspections". In: *University of Fraser Valley Center for Public Safety and Criminal Justice* (2014).

[10] Michael Madaio, Shang-Tse Chen, Oliver L Haimson, Wenwen Zhang, Xiang Cheng, Matthew Hinds-Aldrich, Duen Horng Chau, and Bistra Dilkina. "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*. ACM. 2016, pp. 185–194.

[11] Michael Madaio, Oliver L Haimson, Wenwen Zhang, Xiang Cheng, Matthew Hinds-Aldrich, Bistra Dilkina, and Duen Horng Polo Chau. "Identifying and Prioritizing Fire Inspections: A Case Study of Predicting Fire Risk in Atlanta". In: *Bloomberg Data for Good Exchange, New York, NY, USA* (2015).

[12] City of Pittsburgh. *Fire Prevention Code of Ordinances*. URL: `https://www.municode.com/library/pa/pittsburgh/codes/code_of_ordinances`.

[13] Paul Pringle and Ben Welsh. *Thousands of large L.A. buildings are long overdue for fire inspections*. June 2015. URL: `http://www.latimes.com/local/california/la-me-lafd-inspections-20150616-story.html`.

[14] Ravi Shroff. "Predictive Analytics for City Agencies: Lessons from Children's Services". In: *Big data* 5.3 (2017), pp. 189–196.

[15] The New York Times. *The Oakland Fire: What We Know About Fire Inspections and the Neighborhood*. Dec. 2016. URL: `https://www.nytimes.com/2016/12/14/us/the-oakland-fire-what-we-know-about-building-inspections-and-the-neighborhood.html`.

[16] Shuai Xiao, Junchi Yan, Xiaokang Yang, Hongyuan Zha, and Stephen M Chu. "Modeling the Intensity Function of Point Process Via Recurrent Neural Networks." In: *AAAI*. 2017, pp. 1597–1603.

[17] Chengcheng Xu, Andrew P Tarko, Wei Wang, and Pan Liu. "Predicting crash likelihood and severity on freeways with real-time loop detector data". In: *Accident Analysis & Prevention* 57 (2013), pp. 30–39.