A Dynamic Pipeline for Spatio-Temporal Fire Risk Prediction

Bhavkaran Singh Walia
Carnegie Mellon University
Pittsburgh, PA
bbhavkar@alumni.cmu.edu

Qianyi Hu
Carnegie Mellon University
Pittsburgh, PA
qianyih@andrew.cmu.edu

Jeffrey Chen
Carnegie Mellon University
Pittsburgh, PA
jc6@andrew.cmu.edu

Fangyan Chen
Carnegie Mellon University
Pittsburgh, PA
fangyanc@andrew.cmu.edu

Jessica Lee
Carnegie Mellon University
Pittsburgh, PA
jl5@andrew.cmu.edu

Nathan Kuo
Carnegie Mellon University
Pittsburgh, PA
nekuo@andrew.cmu.edu

Palak Narang
Carnegie Mellon University
Pittsburgh, PA
pnarang@andrew.cmu.edu

Lt. Jason Batts
Pittsburgh Bureau of Fire
Pittsburgh, PA
jason.batts@pittsburghpa.gov

Geoffrey Arnold
Dept. of Innovation and Performance
Pittsburgh, PA
geoffrey.arnold@pittsburghpa.gov

Michael Madaio
Carnegie Mellon University
Pittsburgh, PA
mmadaio@cs.cmu.edu

ABSTRACT

Recent high-profile fire incidents in cities around the world have highlighted gaps in fire risk reduction efforts, as cities grapple with fewer resources and more properties to safeguard. To address this resource gap, prior work has developed machine learning frameworks to predict fire risk and prioritize fire inspections. However, existing approaches were limited by not including time-varying data, never deploying in real-time, and only predicting risk for a small subset of commercial properties in their city. Here, we have developed a predictive risk framework for all 20,636 commercial properties in Pittsburgh, based on time-varying data from a variety of municipal agencies. We have deployed our fire risk model on Pittsburgh Bureau of Fire's (PBF), and we have developed preliminary risk models for residential property fire risk prediction. Our commercial risk model outperforms the prior state of the art with a kappa of 0.33 compared to their 0.17, and is able to be applied to nearly 4 times as many properties as the prior model. In the 5 weeks since our model was first deployed, 58% of our predicted high-risk properties had a fire incident of any kind, while 23% of the *building* fire incidents that occurred took place in our predicted high or medium risk properties. The risk scores from our commercial model are visualized on an interactive dashboard and map to assist the PBF with planning their fire risk reduction initiatives. This work is already helping to improve fire risk reduction in Pittsburgh and is beginning to be adopted by other cities.

KDD '18, August 19–23, 2018, London, United Kingdom

ACM ISBN 978-1-4503-5552-0/18/08…$15.00
https://doi.org/10.1145/3219819.3219913

KEYWORDS

Fire risk, predictive modeling, spatio-temporal risk prediction, civic risk modeling

ACM Reference Format:
Bhavkaran Singh Walia, Qianyi Hu, Jeffrey Chen, Fangyan Chen, Jessica Lee, Nathan Kuo, Palak Narang, Lt. Jason Batts, Geoffrey Arnold, and Michael Madaio. 2018. A Dynamic Pipeline for Spatio-Temporal Fire Risk Prediction. In *Proceedings of The 24th ACM SIGKDD International Conference on Knowledge Discovery Data Mining (KDD '18).* ACM, New York, NY, USA, 10 pages.

1 INTRODUCTION

Fire departments, like many municipal agencies, face the challenge of improving public safety, despite often armed with limited resources. With 475,500 structure fires in the United States in 2016 alone, causing 3,390 civilian deaths and 10.6$ billion of property damage, fire departments, as the "Authority Having Jurisdiction" [16] perform community risk reduction efforts to improve public safety. These include regular fire safety inspections of commercial properties as well as fire safety education initiatives to reduce the risk of residential property fires (which they cannot inspect) [16].

However, such risk reduction initiatives often 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 [15], and are not influenced by a data-driven fire risk assessment of individual properties or communities. As a result, some properties can slip through the cracks, resulting in such high-profile fire incidents as the Grenfell Towers fire in London in summer of 2017, or the Bronx apartment fire in New York City in December, 2017. These properties were later found

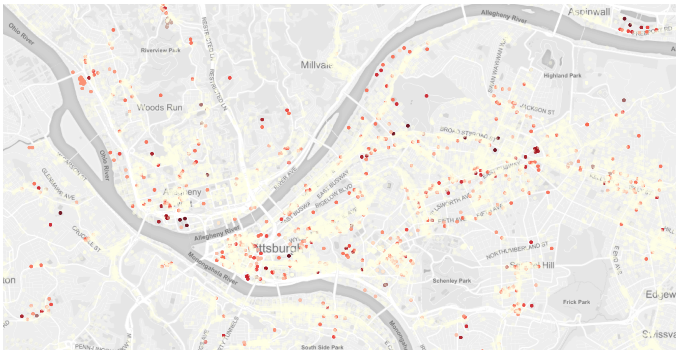

Figure 1: Heatmap of commercial property risk scores in Pittsburgh, with darker colors representing higher risk

to contain several fire safety violations that, if caught, could have helped reduce the fire risk to those properties [2].

To address this, several municipal fire departments, most notably New York and Atlanta, have developed risk-based inspection procedures based on historical fire incidents [6, 13]. However, these prior approaches have either been proprietary, in the case of NYC, and thus the research community is unable to build off of their work or use it as a benchmark, or in the case of Firebird in Atlanta, have relied on static data to create a single set of risk scores for a subset of their commercial properties, without the ability to take in new data and retrain the model as the city changes and evolves over time.

We contribute to the state of the art in fire risk prediction in the following ways:

Commercial Fire Risk Model: We have developed a fire risk prediction model that draws on dynamically updated (i.e. temporal) data about (1) property features, (2) non-fire property inspections and violations (i.e. sanitation, noise, etc, which have been identified in prior work to be significant in informing fire risk levels), and (3) historical emergency calls routed to the fire department (excluding actual building fires). This model generates a risk score (i.e. likelihood of fire) for all of the 20,636 non-residential (i.e. commercial, governmental, industrial, etc) properties in the city of Pittsburgh, with an AUC of 0.74%, and a kappa of 0.33, significantly outperforming the 0.17 kappa of the 2016 Firebird model (the current state of the art) [13].

Deployed Risk Model Pipeline: Our commercial risk model has been deployed for the last five weeks on the Bureau of Fire's server, scraping new data, re-training the model, and generating risk scores on a weekly basis. These risk scores are visualized on an interactive map and data dashboard we developed in partnership with the city of Pittsburgh's Department of Innovation and Performance, to augment Bureau of Fire's planning for commercial property inspection. We have conducted a post-hoc evaluation of our model performance since its deployment, finding that 58% of the predicted high-risk properties had at least one fire incident in the five weeks since our model was deployed.

The Bureau of Fire inspectors and operations chiefs are currently using the fire risk scores to inform their prioritization of commercial properties to inspect, both at the tactical-level of the fire inspectors making daily inspection planning decisions, and for the operations chiefs making strategic-level decisions about their inspection planning and personnel resource allocation.

Residential risk model: We also introduce a novel risk model to generate fire risk scores for residential census blocks in Pittsburgh, achieving 0.74% AUC and 0.45% kappa. Instead of predicting risk for individual properties, as in the commercial model, our model predicts risk for census blocks. By predicting at a broader level, the model can help municipal fire departments target high-risk residential areas for their fire safety education or other risk reduction efforts.

National impact: The commercial risk model deployed at the Pittsburgh Bureau of Fire has been recognized by the GovTech and MetroLab Network as part of their "Innovation of the Month" series for other cities in the MetroLab Network to learn from. The code used in the model and pipeline has been made public to benefit other municipal agencies[1], along with an extensive technical report provided to explain the process[2]. In the weeks since the GovTech publication, the chief data officers for the cities of Baltimore and Syracuse have been in consultation with the project lead for this work to deploy the commercial risk model in their cities.

2 RELATED WORK

Municipal governments around the world are facing the challenge of how to do more with fewer resources and tighter budgets. To address this resource gap, many cities are turning to predictive analytics to prioritize resource allocation and improve outcomes across a variety of domains. Thanks to an increase in the amount of data available about many aspects of civic life, better inter-agency data sharing practices, and an increase in the data analytic capacity of municipal governments, predictive risk modeling is contributing to improved municipal decision-making and civic outcomes. We describe here recent work using predictive risk models for municipal resource prioritization, as well as the prior work in fire risk prediction.

2.1 Civic Risk Modeling

In New Orleans, [9] used neighborhood and administrative data to assess the extent of distressed properties and their impact on the neighborhood. Other recent work incorporated regularly updated temporal data as well as spatial data to predict water contamination at a household-level [3]. Both of those projects, however, were not designed to address a resource prioritization challenge in quite the same way as our fire risk prediction work. More closely related, then, is work such as Potash et. al, who used blood tests, home inspections, and other data to predict lead poisoning at a child-level [17] to inform the Chicago Department of Public Health's home inspection prioritization. Along similar conceptual lines, Salomon et al. developed a model to predict incarceration using data from behavioral health services and the criminal justice system to prioritize mental health interventions [19]. However, neither of those models had (at the time of writing) been deployed in the civic agency with which they were working. For risk prediction models that have been deployed, Lavi et al. used weather, temporal, and spatial data

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

[2] https://www.cmu.edu/metro21/projects/fire-risk-analysis.html

to predict which medical dispatches are most likely to require medical transport [12], while Cuccaro-Alamin et al. describe predictive risk modeling used to inform child welfare interventions, deployed in Allegheny County, PA [7]. We discuss implications of deployed risk models in Section 6.

2.2 Fire Risk Predictive Analytics

In a systematic review of fire risk assessment, Moshashaei et al. found that the vast majority of work on data-driven fire risk prediction has targeted woodland and forest fires, which use weather and topography data, rather than the infrastructure and behavioral data that is at the crux of urban fires (see [18] for the state of the art in using a context-based online learning approach to forest fire risk prediction) [15]. As [15] point out, significantly less work has been done on urban fire risk assessment, and in fact, in their systematic review, 16 of the 17 prior papers on urban fire risk used handcrafted rules for risk based on the consequences of fire (i.e. loss of life) rather than statistical models of likelihood of fire.

Prior work in data-driven urban fire risk analysis, such as [5, 8], has often operated at a less granular level than may be useful, and has not been used to inform prioritization in fire safety inspections. For instance, [5] drew ellipses around the areas of densest concentration of fire incidents to determine the risk level of residential communities, while DaCosta et al. took a more statistical approach to optimizing smoke-alarm inspections, joining 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 [8].

The most relevant precedents for our fire risk prediction 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) [13, 14].

In New York City, in response to high-profile fire incidents such as the Deutsche Bank fire, the Mayor's Office of Data Analytics developed a "risk-based inspection system" using data from structural features and behavioural indicators to predict the fire risk of a building and prioritize the property inspections accordingly [6]. 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 other researchers to build off of it and 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 [13, 14]. 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 purchased from the CoStar property group. This model was designed for AFRD"s Community Risk Reduction Section to help prioritize their commercial property inspections, based on the risk score assigned by the predictive model.

Data	Source	Features	Records	Date Range
Fire Incidents	Bureau of Fire	27	387,264	2009-2017
Violations	PLI	20	13,892	2015-2016
Parcels	OPA	46	579,474	2017
Property Data	OPA	80	578,149	2017
Tax Liens	DCR	11	552,193	2009-2017
ACS	Census Bureau	42	1,100	2016

Table 1: Data Sources

However, their model was developed in summer 2015, and was not designed to operate on dynamic, temporal data, such as data included in other AFRD incident codes (i.e. incidents not included in code 100 (building fires), such as electrical shortages, gas leaks, etc), 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. Thus, 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 summer 2015, which became out of date soon after they were generated.

To address these gaps, our model incorporates dynamically updated, temporal data from the non-building-fire incidents (e.g. gas leak, electrical shortage), non-fire code violations (e.g. noise and sanitation, etc), and property assessment data (i.e. building and land value updated every month), to predict the fire risk for commercial properties in the city of Pittsburgh. While the Firebird model relied heavily on purchasing expensive commercial real estate data, our model uses open data from city agencies, and is thus able to be applied to every commercial property in the city, instead of the small subset the Firebird model was applied to. Our commercial risk model has been deployed on the Pittsburgh Bureau of Fire's server since January 5th, 2018, scraping new data from various sources on a weekly basis and retraining the model every week, generating updated risk scores. Finally, we go beyond previous approaches by developing a preliminary risk model for residential properties at the census block level, to inform Community Risk Reduction efforts such as neighborhood fire safety education.

3 METHODS

3.1 Data Processing

3.1.1 Data description. 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 historical fire incident data, provided by the Pittsburgh Bureau of Fire, from 2009-2017, updated on a weekly basis, from which we include all fire incident codes that have an associated address. From the Allegheny County Office of Property Assessments (OPA), we use property assessment data, updated on a monthly basis, as well as a parcel dataset, which contains information about every parcel in the City of Pittsburgh. Finally, 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). To build the residential risk model, we use two additional datasets. From Allegheny County Department of Court Records (DCR), we use the tax lien data, updated on a monthly basis. We also used the 2012-2016 5-year estimate American Community Survey (ACS) data from the US Census Bureau. We chose the datasets that are likely correlated with fires based on prior work [10], specifically data about income, occupancy, year built, and the year resident moved in. This data is provided at the census block level, which corresponds to the census block level at which we are making our predictions. More detail about the data sets can be found in Table 1, with the number of records, features, and dates up-to-date as of the time of this writing.

3.1.2 Data pre-processing. For the commercial property risk model, we joined the above data sets for each non-residential *address* in our dataset. Although the most granular unit of analysis would be the parcel (instead of address), 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. To do this, we start with the Allegheny County property assessment data set, which we obtained from the Western Pennsylvania Regional Data Center (WPRDC)[3]. We merge the PLI 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 by taking the mean of numeric features and using the most frequent category for categorical features.

For the residential property risk model, we followed the same procedure as above, except that we further aggregated the data to the census block group level by taking the sum for numeric features, which included tax lien, and PLI data, and the most common for categorical features, which included some property data. We took the sum as opposed to the mean to capture the total number of violations and inspections in each block, as well as the total amount of unpaid taxes.

For both models, we then merged the resulting aggregated data frame with the fire incidents dataframe (for both the commercial and residential model). We performed minor cleaning of the fire incident data (e.g. stripping white space, standardizing hyphens, standardizing street abbreviations, etc).

3.2 Model construction

3.2.1 Commercial Risk Model. For the commercial risk model, after joining all of our data sets together at the address level, we were left with a single data set used to train, evaluate, and test the risk model. For this analysis, we used all fire incidents of a 100-level code (i.e. all building fire incidents) as the outcome to predict, and all other features were used as input 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. This is similar to the walk-forward time-partition approach in [13], however, as they point out in their discussion, there was a time-window mismatch between their training set (4 years) and their test set (1 year). They describe that they attempted to unfold their properties into property-years, as we do, but because they did not incorporate time-dependent data, as we do, this approach did not prove useful for them.

In other words, each row of our dataframe is an address-year, which had a 0 or 1 to indicate whether it contained at least one building fire incident in any given year in the 8 years of our data. For each address-year, we included the time-varying features (i.e. PLI violations and non-100 fire incidents [e.g. smoke alarm activation, electrical wiring issues, etc]) as features in that address-year row only if the datetime of that event occurred in the window prior to 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 (6 years of data), a validation set for feature selection (1 year of data), and a test set (the final 1 years of data).

3.2.2 Residential Risk Model. For the residential risk model, after data pre-processing, all of the data was aggregated at the census block group level. Since block groups are much larger than individual addresses, fires were much more frequent, with most blocks having multiple fires over the span of the entire time period of the data. We unfolded each census block into a block-year instance, much like in the commercial model. Each row represented a single census block during a single year, with features that included PLI violations, non-100 incidents, and tax lien data from that year. The target to predict was whether there was a code-100 fire within that block in that time period. After dividing up the time-dependent data, we merged it with the time-invariant data that corresponded with the census blocks, like property data and ACS data. Similar to the commercial model, we then one-hot encoded the features and used time-series cross-validation, and we similarly split the data into a 6-fold cross-validation on a training set of 6 years, a 1-year validation set, and a 1-year test set.

3.3 Model tuning

We used a 6-month validation set for feature selection for the commercial model. We computed the feature importance of each feature on our initial commercial model. Using feature importance as a threshold, we repeatedly pruned the model of features. After deciding the best subset by the F1 metric, the dimensions of the commercial model were reduced from 830 features to 227 features, which allowed for faster run-time and removed irrelevant data. We also ran a 5-fold cross-validated grid search to identify the optimal hyperparameters for our best performing models. For the XGBoost commercial risk model, we searched max_depth, min_child_weight, subsample, colsample_by_tree and tuned the rest manually. For the Random Forest residential risk model, we searched n_estimators, max_depth, and max_features using a 1-year validation set.

4 RESULTS

Using data joined from fire incidents, non-fire inspection violations, and property assessments, we evaluate multiple model types on their ability to predict the likelihood of a fire occurring in a given spatial window (address for commercial, and census block for residential) for the 1-year window of our test set. We first report

[3]www.wprdc.org

	Model Performance			
Model	Kappa	AUC	Recall	Precision
Logistic Regression	0.004	0.53	0.16	0.006
Ada Boost	0.12	0.63	0.27	0.085
Random Forest	0.09	0.70	0.43	0.056
XG Boost	**0.33**	**0.75**	**0.50**	**0.26**
KDD16 Firebird	0.17	0.8	0.72	0.18

Table 2: Evaluation of Multiple Commercial Risk Models

	Model Performance			
Model Type	Kappa	AUC	Recall	Precision
Ada Boost	0.40	0.72	0.46	0.98
Random Forest	**0.45**	**0.74**	**0.58**	**0.89**
XG Boost	0.43	0.73	0.49	0.97

Table 3: Evaluation of Multiple Residential Risk Models

evaluations of multiple model types, comparing our results to the current state of the art for commercial risk prediction [13]. We then describe some of the predictive features that were ranked as highly important in the models' ability to predict fires, for both commercial and residential models.

4.1 Commercial model evaluation

Because for the use case of fire prediction we want to prioritize correctly classifying more of the positive class (i.e. fire) over minimizing false positives (which may result in more inspections, but would be less likely to lead to missed incidents), we decided to use kappa and recall as our two main evaluation measures. For our commercial risk model, we selected the XG Boost model for its better performance across all measures, as seen in Table 2.

Although the Firebird model has a larger recall and AUC[4], it is important to note that the Firebird model was only applicable to a small subset of properties in the city of Atlanta. Because their dataset was highly incomplete, in order to increase model performance, they created a model for only 5,022 non-residential properties (out of the more than 20,000 properties in the city). This prevents their model from creating a fire risk score for every non-residential property in the city, as reported in [13].

By using open data provided by the city government, our approach, in contrast, can be applied to all 20,000 non-residential properties in the city. Our recall of 50% is significant considering the extreme class imbalance in our commercial property data. For example, for any given 1-year period, our model accurately predicts nearly half of the fires, 70 times more effective than random guess, which would be correct 0.71% of the time, given the distribution of fire incidents.

[4]While only the AUC was reported in the Firebird KDD16 paper, the other metrics were provided through correspondence with an author of that paper

4.2 Residential risk model evaluation

For the residential risk model, we used the same metrics of evaluation, namely kappa, AUC, recall, and precision. Instead of generating risk scores for individual properties, we generated a risk score for each of the approximately 350 census blocks in Pittsburgh, as that is the level at which the Fire Department may conduct Community Risk Reduction fire safety education efforts, instead of inspecting the individual commercial properties. We found Random Forest to be the best performing model (n_estimators=500, max_depth=10, max_features='log2'), with a mean AUC of 0.82 on the validation set (Figure 3) and an AUC of 0.74 on the 2017 test set.

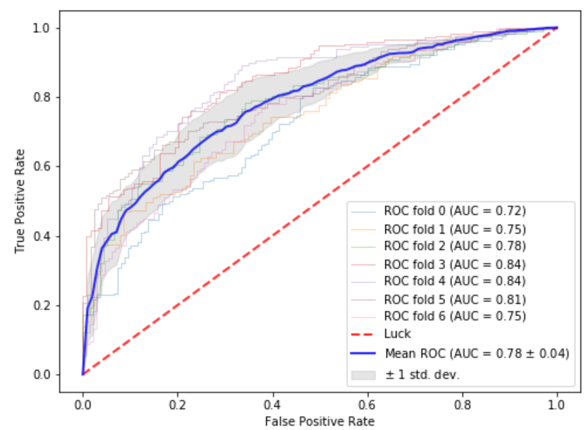

Figure 2: ROC Curve for XGBoost on Residential Property Census Blocks

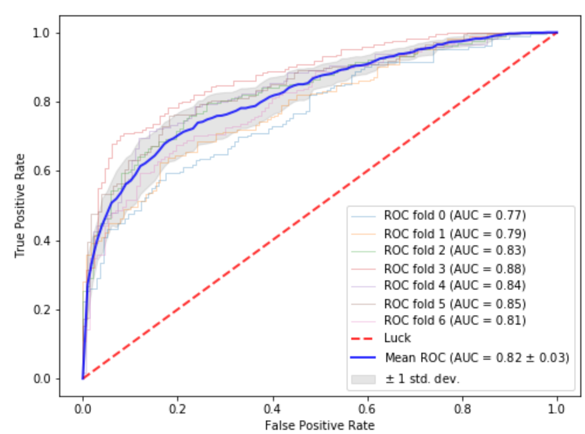

Figure 3: ROC Curve for Random Forest on Residential Property Census Blocks

	Commercial	Residential
1	lot area	EMS call
2	fair market building	tax lien amount
3	fair market land	Allegheny County tax lien
4	sale price	medical assist
5	activated smoke detector, no fire	dispatched and cancelled
6	unintent. activation of smoke detector	city & school tax lien
7	tax exempt	owner occupied
8	unintent. activation of alarm	good intent call
9	apart 5-19 units	fair market land
10	individual owner	smoke detector activation

Table 4: Features ranked by feature importance

Risk Level	Commercial Addresses	Residential Census Blocks
High (7-10)	103	23
Medium (4-6)	550	207
Low (1-3)	19983	113

Table 5: Addresses and census blocks at each risk level

The results are better than the commercial model results, most likely because of the better class balance (53% of the total census-block-years contained fires). Interestingly, for all of the trained models, precision was significantly higher than recall. Future work will continue to investigate why this is the case, to help improve the recall. While some prior work *has* been conducted for residential property risk prediction, for instance with Jon Jay's work predicting residential building fires in Baton Rouge, LA [11], Jay's model predicted fires at the building-level, instead of the census-block level, which we believe through discussion with the Bureau of Fire to be the most useful level for the fire department to intervene through community risk reduction. Thus, the two approaches may not be directly comparable.

4.3 Feature importance ranking

The most important predictive features in the commercial model were lot area, appraised building and land value, and sale price. Other significant predictors include smoke detector activation, densely populated apartment complexes, tax exempt properties (e.g. low-income housing, religious institutions, etc.), and whether or not a building is owned by an individual owner, compared to a corporation. The most important features for residential property census blocks were EMS call, unpaid tax liens on properties in that census block, owner-occupied (which may be an indication that the property is better kept up), as well as fairmarket land values.

4.4 Analysis of High-Risk Properties

Our final commercial risk model outputs the prediction probabilities for each address in our data, which is the probability that the address will be a positive class (i.e. fire incident of code 100) in the final 1-year window. A larger probability means that it is more likely that the property will have a code 100 fire incident. In order to have risks scores that are easier to understand for the Bureau of Fire, we convert these probabilities into a risk score integer from 1-10.

Table 5 shows the distribution of addresses and census blocks at each of the three risk categories associated with the risk scores. The low-risk category contains risk scores 1-3, medium-risk scores 4-6, and high-risk 7-10. We arrived at these three risk categories through discussion with our partners at the Bureau of Fire, to best fit their inspection planning needs, though they can continue to be tuned. There are significantly more commercial properties with a risk score of 1 (18,364) than any other risk score, while the residential census blocks have more blocks in the medium risk category. Future work should continue to optimize the discretization of risk integers into risk categories.

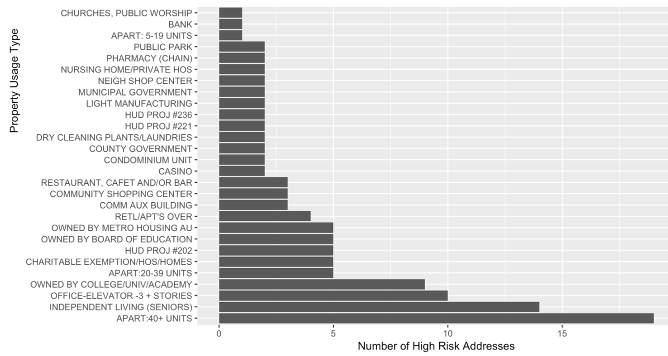

Figure 4: High-Risk Property Types

In Figure 4, we show the property types which contain the most high-risk properties. The property types with the most high-risk properties in the city are apartments of 40+ and independent living for seniors, although it should be noted that there are many more of those addresses in the city than, for instance, HUD Project #221. The charitable exemption property type contained several high-risk properties, while also showing up as a highly predictive feature. Interestingly, while they were not in the top 10 features, the property types of apartments of 40+ units, Independent living (seniors), and Owned by Metro Housing Authority were highly predictive features for commercial property fire risk. Those, and other high-risk property types, such as those owned by Board of Education, Apartments, 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 DEPLOYMENT

5.1 Integration into existing practices

As part of the Pittsburgh Bureau of Fire's risk reduction efforts, they conduct regular property inspections for commercial properties. However, because of the large disparity between the number of possible properties to inspect and the capacity of PBF's inspectors, prioritization based on fire risk is necessary, as pointed out by [13]. After developing the commercial fire risk model and evaluating its performance, we provided the Pittsburgh Bureau of Fire with a tool they could use to integrate those fire risk scores into their existing inspection decision-making practices. Through conversations with PBF fire chiefs and fire inspectors, we decided on an interactive data dashboard and an interactive map, using a platform already incorporated into PBF fire inspectors' existing workflow, called "Burgh's Eye View" (BEV).

5.2 Fire Risk Visualization Tools

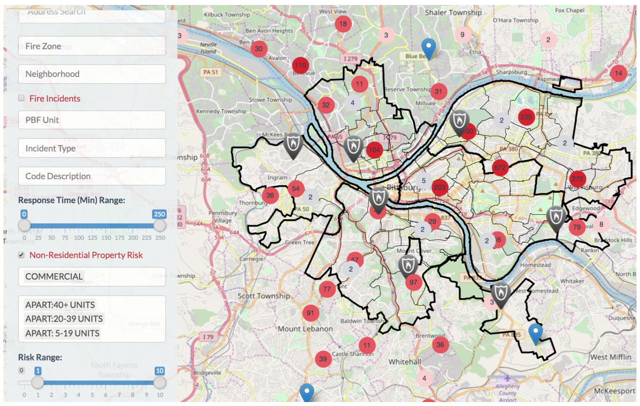

Figure 5: Interactive map displaying high-risk properties, filtered by property type

5.2.1 Interactive Map. Inspectors currently use BEV to find information on the properties their risk reduction operational planning has determined are due for inspection on any given day. Thus, we displayed the risk scores as an additional filterable "layer" here to best fit 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. Additionally, operations chiefs can filter by property usage type in addition to risk level to support their strategic planning around risk reduction initiatives to target, for instance, senior centers or high-rise apartment complexes, and prioritize by the highest-risk of those.

5.2.2 Data Dashboard. While the interactive map is helpful for spatial visualizations of the risk scores, we also wanted to support strategic planning from the fire chiefs by allowing them to view the risk scores in the aggregate, and compare the relative risk levels of multiple types of properties or locations. We thus created an interactive data dashboard, using R Shiny, to allow fire chiefs (or inspectors) 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, for privacy). PBF staff can then download the visualization as an image or download the table of property information as a .csv file.

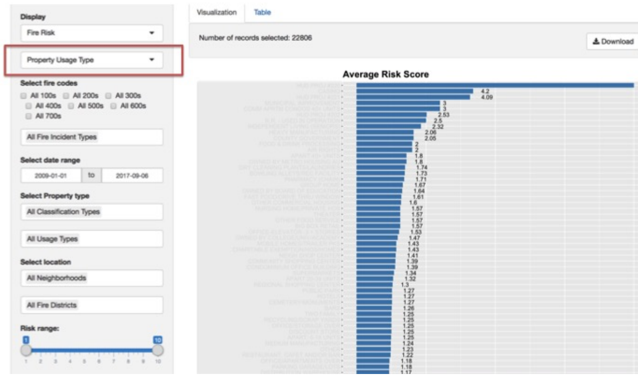

Figure 6: Data dashboard displaying property usage types according to their average fire risk score

5.3 Model Pipeline

Since data on fire incident and property features are updated weekly, our model must be re-trained periodically. Therefore, our model is deployed on the Bureau of Fire's servers and triggered by a "cron" script to run every Saturday. The program deployed scrapes the data source, retrains the model and updates risks scores on the map and dashboard.

Because fire risk levels will change over time as new data about fire incidents and property features are acquired, in order for our model to continue to be useful for the Bureau of Fire, we 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 new fire risk scores for each address. Our model is currently deployed as a set of Python scripts that are triggered by a "cron" script on their server to run every Saturday. It scrapes the data sources for the latest dataset, retrains the model, and updates those risk scores on the map and dashboard.

5.4 Post-hoc analysis

The commercial risk model has been deployed on the Bureau of Fire's server since January 5th, 2018, and has retrained five times in the subsequent five weeks as of this writing. Here, we discuss post-hoc analysis of the first model's performance on predicting fires that occurred since then, as well as stability of the model's performance across the five iterations. As of this writing, we use the results from the model with the test set ending on December 31st, 2017. Of the 13 fire incidents between 1/1/2018 and 2/10/2018 (when this analysis was conducted), 3 of them occurred in high or medium risk properties. Additionally, of those 103 high risk properties, 1.94% of them had code 100s incident, compared to 0.72% of medium risk, and 0.06% of low risk properties. See Table 6 for more detail on the number of fires that *actually* occurred in low, medium, and high risk properties since our first model ran.

To understand the stability of our model's performance through the 5 iterations, we report our model's mean kappa, AUC and F1 in Table 7. The low standard deviation of each metric (less than 0.01) shows that the performance is stable.

Risk Level	All Incidents	All Code-100 Incidents
High Risk	58.25%	1.94%
Medium Risk	22.18%	0.72%
Low Risk	2.48%	0.06%

Table 6: Percent of properties in each risk category that had an incident since the model first ran

	Kappa	AUC	F1
Mean	0.287	0.715	0.293
Std Dev	0.0073	0.0081	0.0072

Table 7: Model stability over 5 weeks

6 DISCUSSION

In this report, we have described our process for modeling structural fire risk, based on data about historical fire-related incidents, property inspections and violations, property assessments, property characteristics (size, sale price), and parcel data, as well as tax lien and American Community Survey data. We described how we used these features to train two predictive models, one for commercial properties and the other for residential property census blocks, and discuss results from evaluations of those models. We described how we deployed the commercial property risk model at the Pittsburgh Bureau of Fire, where it has been updating on a weekly basis for the past five weeks, displaying the risk scores on a data dashboard and an interactive map.

In inspecting the property types that had a high frequency of high-risk properties, as well as some of our important features, we find that, similar to the Firebird model, high-rise apartment complexes have a high number of high-risk properties, perhaps due to the high density. However, we also find that Housing and Urban Development (HUD) properties and properties owned by the Metro Housing Authority are also some of the most common high-risk properties. This suggests that the Pittsburgh Bureau of Fire may benefit from strengthening their fire risk reduction initiatives around such properties, particularly the commercial property fire inspections.

We go beyond the previous state of the art in structural fire risk prediction in the Firebird model by 1) providing a risk model that incorporates time-varying data about inspections and non-fire incidents at those locations, 2) allowing us to deploy the model on the Bureau of Fire server where it is retraining every week. We also go beyond prior work by providing 3) a post-hoc analysis of the performance of this model in the five weeks since its deployment. Finally, we extend the Firebird work by 4) developing a risk model for residential property census blocks, to inform community fire safety education efforts.

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 the work from Xu et al. [20], 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, PBF 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 potential loss of life or amount of property damage in the event of a fire. 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 [1, 4, 7] for a larger discussion about integrating predictive modeling into municipal decision-making.

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 the likelihood of a fire incident.

6.1 Limitations

While the most granular level of prediction would have been to predict fires at the parcel level, fire incidents in Pitsburgh were only recorded at the address level. While we would have liked to include the fire inspection and violations in addition to the noise and sanitation violations from PLI, those data were quite sparse for Pittsburgh, and were thus not useful. Further, while this model has been deployed for 5 weeks, we have no longitudinal evaluation over several months or years. Once the model has been in production long enough, we'd like to analyze the hit rate stability and decay over time. Since it has been only six weeks since model deployment, it may not yet be useful to conduct that sort of analysis.

6.2 Future Work

As the commercial risk model is deployed on the Bureau of Fire's servers, we intend to continue to monitor the stability and efficacy of the model's performance over time. Future work should experiment with an active learning or reinforcement learning paradigm, where the model's *actual* performance in predicting fire incidents in some subsequent time window (not just measured against the held-out test set) can be used as a "reward" (in the reinforcement learning sense) to improve the model when it more accurately predicts fire incidents. Future work can also consider additional data sets such as 311 data, or novel model types, particularly models that take time into account, such as recurrent neural networks or long-short term memory models.

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. We have targeted the census block level for our initial experiments with residential property risk prediction, but future work should evaluate different levels of aggregation. As municipal fire departments conduct Community Risk Reduction efforts like fire safety classes or demonstrations instead of property-level inspections, a deployed residential risk model should provide a prioritized location for the fire department to conduct its fire safety education efforts. Or, looking beyond existing municipal subdivisions (e.g.

census block, neighborhood), future work 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 CONCLUSION

We intend for this work to be useful for municipal Fire Departments, or for any municipal agency or civic-minded organization (e.g. Code for America) that wants to incorporate data-driven predictive risk modeling into their fire risk reduction decision-making processes.

In the spirit of openness and transparency, and to make it more likely that this approach can 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. We are currently in discussion with the chief data officers of the cities of Syracuse and Baltimore to help them adapt this approach for their purposes, given their data.

While we would have liked to share our 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. In addition, 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. 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. We invite a larger public discussion about the ways to mitigate the risks and tradeoffs of integrating inherently uncertain predictive models into civic decision-making while not jeopardizing the privacy and data security of stakeholders 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 existing inspection practices. We hope that this work 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. 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

ACKNOWLEDGMENTS

We thank our partners at Pittsburgh Bureau of Fire, including Chief Cook, Chief Auvil, and Chief Jones for being willing to examine and improve their processes, as well as Director Hissrich of the Department of Public Safety. We also extend our heartfelt thanks to our partners at the Department of Innovation and Performance for their assistance with data acquisition and model deployment, including Laura Meixell and Maxwell Cercone. We are indebted to many at the Metro21 Smart Cities Initiative for the stimulating conversation as well as the generous financial support provided by the Hillman Foundation. Finally, we thank others for the rich conversations, from Dr. Hinds-Aldrich at the NFPA, to the other members of the Firebird team.

REFERENCES

[1] Susan Athey. 2017. Beyond prediction: Using big data for policy problems. *Science* 355, 6324 (2017), 483–485.
[2] John Murray Brown. 2017. Grenfell fire tragedy prompts inspection of 4,000 UK tower blocks. Financial Times. (jun 2017). https://www.ft.com/content/a6d17890-526b-11e7-a1f2-db19572361bb
[3] Alex Chojnacki, Chengyu Dai, Arya Farahi, Guangsha Shi, Jared Webb, Daniel T Zhang, Jacob Abernethy, and Eric Schwartz. 2017. A Data Science Approach to Understanding Residential Water Contamination in Flint. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 1407–1416.
[4] Christopher E Church and Amanda J Fairchild. 2017. In Search of a Silver Bullet: Child Welfare's Embrace of Predictive Analytics. *Juvenile and Family Court Journal* 68, 1 (2017), 67–81.
[5] Joseph Clare, Len Garis, Darryl Plecas, and Charles Jennings. 2012. Reduced frequency and severity of residential fires following delivery of fire prevention education by on-duty fire fighters: Cluster randomized controlled study. *Journal of safety research* 43, 2 (2012), 123–128.
[6] Eddie Copeland. 2015. BIG DATA IN THE BIG APPLE. *Capital City Foundation.* (2015). 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. 2017. Risk assessment and decision making in child protective services: Predictive risk modeling in context. *Children and Youth Services Review* (2017).
[8] Marc DaCosta, Jeremy Krinsley, and Brian Abelson. 2015. Optimizing Local Smoke Alarm Inspections with Federal Data. *Bloomberg Data for Good Exchange* (2015).
[9] Ben Green, Alejandra Caro, Matthew Conway, Robert Manduca, Tom Plagge, and Abby Miller. 2015. Mining administrative data to spur urban revitalization. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 1829–1838.
[10] Katherine Hillenbrand. 2016. Predicting Fire Risk: From New Orleans to a Nationwide Tool. *Data-Smart City Solutions* (2016). https://datasmart.ash.harvard.edu/news/article/predicting-fire-risk-from-new-orleans-to-a-nationwide-tool-846
[11] Jonathan Jay. 2017. How we predicted building fires in Baton Rouge, LA – working version. *Urban Analytics Blog* (2017). https://scholar.harvard.edu/jonjay/blog/how-we-predicted-building-fires-baton-rouge-la-working-version
[12] Karen Lavi, Ritvik Kharkar, Mathew Kiang, Christoph Hartmann, Paul Van Der Boor, Adolfo De Unanue, Leigh Tami, Anson Turley, Cedric Robinson, Brandon Crowley, Eric Potash, and Rayid Ghani. 2017. Using Machine Learning to Improve Emergency Medical Dispatch Decisions. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM.
[13] Michael Madaio, Shang-Tse Chen, Oliver L Haimson, Wenwen Zhang, Xiang Cheng, Matthew Hinds-Aldrich, Duen Horng Chau, and Bistra Dilkina. 2016. 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, 185–194.
[14] Michael Madaio, Oliver L Haimson, Wenwen Zhang, Xiang Cheng, Matthew Hinds-Aldrich, Bistra Dilkina, and Duen Horng Polo Chau. 2015. Identifying and Prioritizing Fire Inspections: A Case Study of Predicting Fire Risk in Atlanta. *Bloomberg Data for Good Exchange, New York, NY, USA* (2015).
[15] Parisa Moshashaei, Seyed Shamseddin Alizadeh, et al. 2017. Fire Risk Assessment: A Systematic Review of the Methodology and Functional Areas. *Iranian Journal of Health, Safety and Environment* 4, 1 (2017), 654–669.
[16] City of Pittsburgh. [n. d.]. Fire Prevention Code of Ordinances. ([n. d.]). https://www.municode.com/library/pa/pittsburgh/codes/code_of_ordinances

[17] Eric Potash, Joe Brew, Alexander Loewi, Subhabrata Majumdar, Andrew Reece, Joe Walsh, Eric Rozier, Emile Jorgenson, Raed Mansour, and Rayid Ghani. 2015. Predictive modeling for public health: Preventing childhood lead poisoning. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2039–2047.

[18] Mahsa Salehi, Laura Irina Rusu, Timothy Lynar, and Anna Phan. 2016. 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*. ACM, 245–254.

[19] Erika Salomon, Tzu-Yun Lin, Kate Boxer, Hareem Naveed, Lauren Haynes, Joe Walsh, Jen Helsby, Steve Yoder, Robert Sullivan, Chris Schneweis, et al. [n. d.]. Reducing Incarceration through Prioritized Interventions. *University of Chicago Data Science for Social Good* ([n. d.]).

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