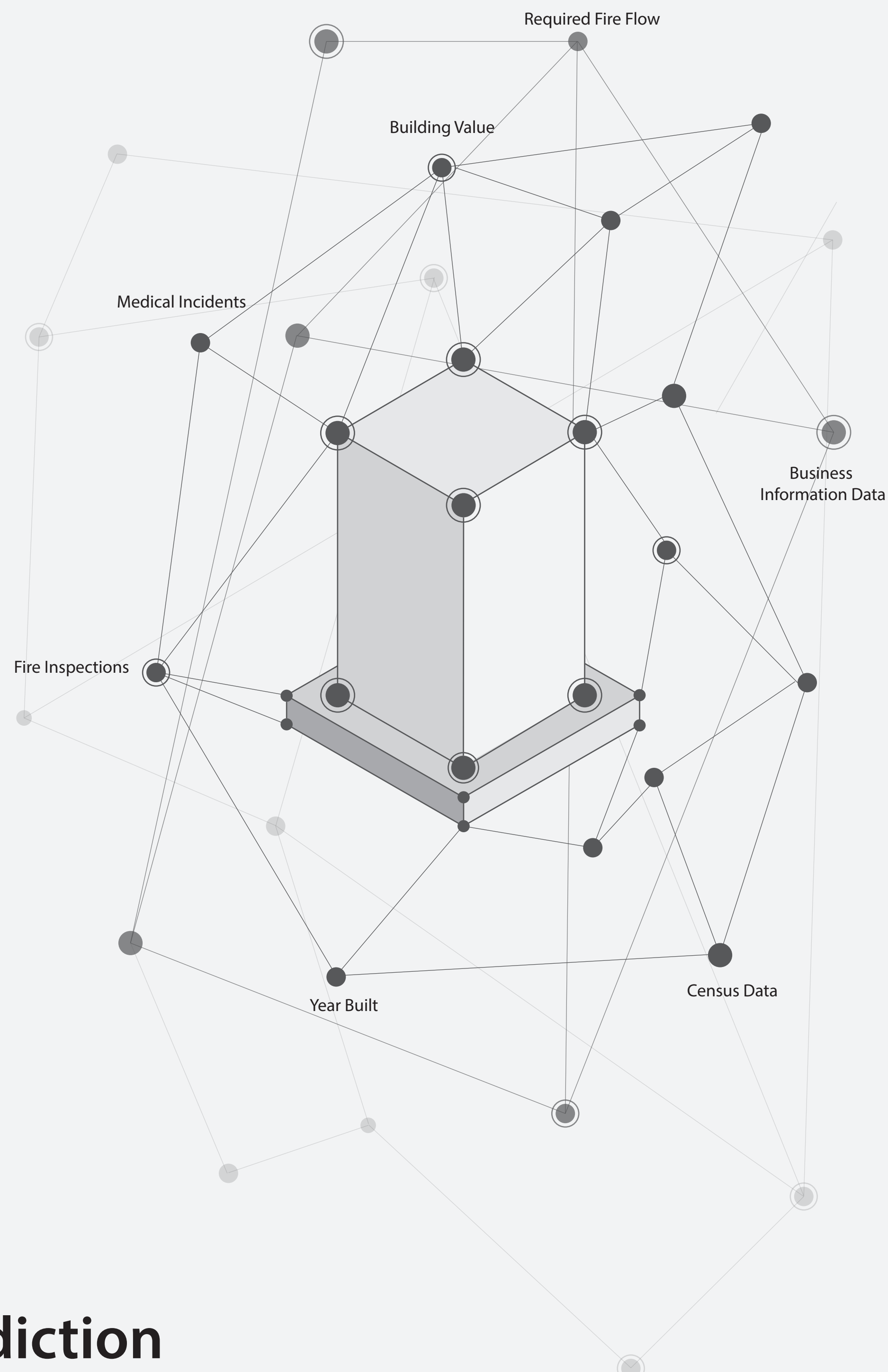


A Building Fire Risk Prediction Validation Project

A **Fire Underwriters Survey**/**Opta Information Intelligence** project
in partnership with
Vancouver Fire Rescue Service and New Westminster Fire Rescue Service

April, 2019

A Building Fire Risk Prediction Validation Project

SUGGEST CITATION: FUS Building Fire Risk Prediction Validation Project

DATE OF PUBLICATION: December 12th, 2018 (updated April 18th, 2019)

Fire Underwriters Survey/Opta Information Intelligence Project team

Robert McGuinness P.Eng. – Project Lead (robert.mcguinness@scm.ca)
Greg Gilbert – Chief Data Scientist (greg.gilbert@scm.ca)
Zakary Whittamore – Senior Data Scientist (zakary.whittamore@scm.ca)
Leon Zhong – GIS Analyst (leon.zhong@scm.ca)
Ruan Van Zyl – Data Scientist (ruan.van.zyl@scm.ca)
Jeeyoung Lee – Data Scientist (jeeyoung.lee@scm.ca)

Fire Department and Municipal Partners

We would like to thank Vancouver Fire Rescue Service (VFRS), New Westminster Fire Rescue Service (NWFRS), the City of Vancouver and the City of New Westminster for partnering with us on this project. Without their generous participation and support this validation project would not have been possible. Of particular note we would like to thank the following from VFRS: Fire Chief Darrell Reid, Deputy Fire Chief Kiran Marohn and Deputy Fire Chief Rob Renning; from NWFRS: Fire Chief Tim Armstrong and Fire Protection Engineer Isabella Li P.Eng.

Image credits

p.3 – Percentage of dangerous buildings identified in first 25% of inspections from "Big Data in the Big Apple" [1]

p.3 – GIS imagery by Christopher Yee. Base aerial imagery source: ESRI, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DC, USDA, USGS, AeroGRID, IGN, and the GIS User Community.

Contents

Summary

"…the percentage of **severe violations discovered increases from 15% to 44%** in the first 25% of inspections…"

In Canada, similar to the US, fire codes are enforced through fire prevention inspections conducted by fire department personnel. In certain provinces changes are happening with fire related legislation where a municipality will be tasked with completing a risk-based approach to providing fire protection services. Where a regular system of inspections is required under current Acts, fire departments generally do not have a defined risk-based framework in place for prioritizing inspections, however some targeting of higher risks is common.

This project was informed by recent work in predictive modeling of building fire risk completed in New York City [1], the City of Atlanta [2] and the City of Pittsburgh [3]. In each of these municipalities the results of the projects led to a change in how building fire inspections are targeted. A data-driven risk-based approach proved to be better at targeting risks than previous practice.

Building fire risk prediction work with the City of Pittsburgh was completed as a *Metro21: Smart Cities Initiative* research project. The following provides a description of the *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."* [3]

In 2018 Infrastructure Canada launched the Smart Cities Challenge with prizes of up to $75 million indicating the importance of smart cities solutions. The competition is described as follows:

> *"The Challenge encourages communities to adopt a smart cities approach to improve the lives of their residents through innovation, data and connected technology."*

We wanted to validate the results of these smart-cities solutions to building fire risk assessment by completing similar risk assessment-based prediction work for two municipalities in Canada. We invited two Cities to participate based on population size. Our aim was to see if machine learning methodologies would prove useful for a larger city, where there would be more data available, as well as a city with a lower population. The City of Vancouver (population of 631,486 in 2016), with approximately 21,000 inspectable properties, and the City of New Westminster (population of 70,996 in 2016), with approximately 2,200 inspectable properties, partnered with us on the project. Could a data-driven building-based risk assessment better prioritize inspections in both these Municipalities?

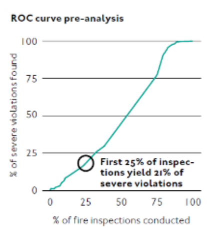
The models were trained on data from 2013 to 2016. We then completed an unbiased test on 2017 fire incident data. The models showed good predictions with approximately 70% of fires

Risk Based Inspection System (RBIS), FDNY:

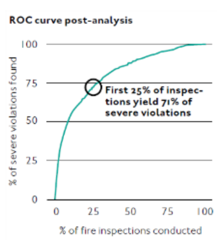
"Where previous versions of the City's fire risk model had weighted criteria based on focus group discussions with fire fighters, MODA tested them against data from actual fires to calculate their relative importance.

Using this information, MODA created a data-driven model that could predict which buildings were most at risk of having serious fires with far greater accuracy.

The contrast is striking. Whereas the old model failed to identify high-risk zones in areas such as Harlem, Downtown Manhattan and the Rockaways, the new model very closely reflected reality." [3]



Prior to model first quarter of inspections for year yields 21% of severe violations.



With MODA model first quarter of inspections for year yields 71% of severe violations.

identified for the 2017 period (with a false positive rate of approximately 25%). We assign risk scores to all properties in both Cities and then look at the impact of identifying violations by using a risk-based approach. In both Cities the identification of violations improves (The percentage of severe violations discovered increases from 15% to 44% in the first 25% of inspections for the City of Vancouver).

Further to the prediction rates produced by the models, this project also demonstrates how various municipal datasets, merged together, can rapidly produce a risk assessment model for a municipality. Once a base model is established and measured, further datasets and fire knowledge can be integrated to improve the model. Furthermore, building a risk assessment model using existing datasets and pipelines allows for regular and dynamic updates.

Introduction

Overview

Municipal fire departments across Canada are tasked with enforcing national/provincial fire codes. In 2014 there were 38,844 fire incidents and a total of 109 fire-related deaths [2]. Fire codes are enforced through fire prevention inspections conducted by fire department personnel. While certain provinces mandate that inspections be completed on a "complaint/request" basis only, other provinces mandate a regular system of inspections for all Part 3 occupancies of the National Building Code. Part 3 of the Code relates to "Fire Protection, Occupant Safety, and Accessibility" in buildings other than those listed in Part 9 "Housing and Small Buildings". Even in provinces not mandating a regular system of inspections, some municipalities will complete inspections on Part 3 occupancy buildings in order to reduce the frequency and severity of fires as well as to improve safety for responding fire fighters. Generally, the scheduling of a regular system of inspections is simply based on a frequency. Some municipalities aim to inspect all inspectable[1] properties (Part 3 occupancies) on an annual basis. There may be some risk-based targeting for certain properties, such as inspecting care homes on a semi-annual frequency; however, inspection targeting does not generally follow a risk-based approach. How can a municipal fire department ensure they are prioritizing the inspection of buildings at most risk of fire?

The following comment on the topic is from the New York Mayors Office of Data Analytics (MODA):

> *"…see if data could complement and strengthen fire fighters' natural intuition in identifying dangerous buildings. Could accessing datasets held by organisations outside the Fire Department be useful to their work? Could the factors underpinning fire fighters' gut instincts (the age of the building, the type of business, etc.) be quantified more precisely with better data"* [1]

Upcoming Legislation

Updates to fire related legislation are happening in both British Columbia and Ontario. The following is proposed for the BC Fire Safety Act:

> *"A monitoring entity must implement a risk-based compliance monitoring system for public buildings"* [5]

In Ontario the following is proposed:

"Every municipality, and every fire department in a territory without municipal organization, must:

a) complete and review a community risk assessment as provided by this Regulation and

[1] An inspectable property is an occupancy as listed in Part 3 of the National Building Code. Buildings under Part 9 of the National building Code are not required to be inspected.

b) *use its community risk assessment to inform decision about the provision of fire protection services"* [6]

Updates to the Ontario Act also specify profiles for the risk assessment which include: geographic, building stock, critical infrastructure, demographic, hazard, public safety response, community services, economic, and past loss and event history.

Related Work

Risk prediction models related to urban property-level fires mostly began with the work of the Mayor's Office of Data Analytics (MODA) in New York City in 2013 [1]. Prior to the fire risk prediction model developed by MODA, New York City's Fire Department (FDNY) used a fire risk model that weighted the risk criteria based on focus group discussions with FDNY personnel. MODA developed a model that used various municipal departmental datasets and determined the feature relative importance using statistical learning techniques. Prior to the MODA data-driven model, the first 25% of inspections resulted in 21% of the most severe violations being identified; using the MODA prediction model 25% of inspections resulted in more than 70% being discovered [1]. The model, called FireCast 2.0, is used to target inspections in FDNY's Risk Based Inspection System (RBIS).

> *"Within a month of the first FireCast model being deployed, the number of violations issued across the city increased by nearly 20%. Inspections can never eliminate the risk of fire because they are primarily caused by human error, but as many as a quarter of buildings that suffer fires will now have been inspected within the previous three months. It is thought that this will save civilian and firefighter lives, as fire crews will be more familiar with buildings they enter."* [7]

Firebird: Predicting Fire Risk and Prioritizing Fire Inspections in Atlanta.

"…predict **71.36%** of the fires in 2014-2015" [6]

In 2015, a group at Georgia Tech worked with the Atlanta Fire Rescue Department (AFRD) on a model called Firebird which involved predicting fire risk and prioritizing fire inspections [2]. The Firebird model used various datasets (parcel, business licence, crime, demographic, socioeconomic, etc.) and historical fire incident data for the period July 2011 to March 2014 as training data. Fires were predicted for the period April 2014 to March 2015 with a prediction rate (True Positive Rate, TPR) of 71.36% (model predicted 71.36% of actual fires) and with a false positive rate (FPR) of 20% (20% of properties predicted as higher risk did not have a fire). The project and results were featured in the National Fire Protection Association Journal in June 2016 [8].

Metro21: Predictive Modeling of Building Fire Risk:

"…For any 6-month window, we were able to accurately detect over half (**57%**) of the fire incidents that occurred" [9]

In 2017/2018 one of the members of the Georgia Tech project moved to Carnegie Mellon University and worked with the City of Pittsburgh Bureau of Fire on a Smart Cities (Metro21 [3]) Research project on "Predictive Modeling of Building Fire Risk". The approach was similar to that of the Atlanta project except that the model type was different. For a 6-month window the model was able to predict 57% of the incidents that occurred. One of the other differences from the Atlanta model is that the Pittsburgh model is deployed on the Bureau of Fire's server and re-calculates risk scores every week as updated datasets are available. This is similar to the Risk Based Inspection System (RBIS) used by FDNY. The author of the project also notes that the risk scoring approach should be used to complement risk assessment and strategic planning for prevention as there may be factors unknown to the model, such as mobility of residents, which may be critical in informing inspection procedures.

The "Predictive Modeling of Building Fire Risk" paper references work completed in British Columbia through the University of the Fraser Valley:

> *"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 these 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 creators of the handcrafted weights and rules, and is not likely to be flexible and improve over time with new data." [3]

One of the major advantages of a prediction modeling approach, such as FDNY, Atlanta and Pittsburgh, is that risks can be categorized using existing datasets leading to a quicker assessment without the need to build new datasets. As more datasets become available these can be added to strengthen the model. Additionally, these models can learn which variables/features are more determinant of fire in a specific municipality over time. The features, and weights of these features, that predict a fire in one municipality are different from those in a neighbouring municipality. This is clearly shown in this report where we compare two municipalities.

London Fire Brigade in the UK has adopted a similar data-driven approach to fire risk prediction for home safety and smoke alarm targeting [10].

Methodology and Model Results

Partnering Municipalities

Two municipalities in the Lower Mainland were selected to participate in the project. While we expected there to be large datasets available for a relatively larger municipality, we also wanted to see how a prediction model would perform for a relatively smaller municipality. The City of Vancouver, with a population of 631,486 (2016), and the City of New Westminster, with a population of 70,996 (2016), agreed to participate. New Westminster Fire and Rescue Services (NWFRS) is responsible for inspecting "inspectable" properties in the City of New Westminster. While Vancouver Fire and Rescue Services (VFRS) provides response to the University of British Columbia (UBC) and the University Endowment Lands (LEL), VFRS is not responsible for inspections in this area. As such the project areas are limited to the municipal boundaries of the Cities as shown in figure 1.

Figure 1 Project Areas

Datasets

Based on the work completed by MODA [1]; Georgia Tech and Atlanta Fire Rescue Department (AFRD) [2]; and Carnegie Mellon University and Pittsburgh Bureau of Fire [3]; we proposed to include fire incident data, fire inspection data, property assessment data, 2016 census data (demographic and socioeconomic), parcel data, 311 data (including sanitation data), crime data, as well as other available property data.

We met with both VFRS and NWFRS staff to discuss the datasets and other available building level information. We realized during these early meetings that there would likely be limited ability to extract certain municipal datasets beyond what was readily available as this was simply a proof-of-concept project. We had to find a balance between allocating time and resources to data requests and completing the project in a reasonable timeframe. Some of the data was only available at the address block level and we had to decide whether this would likely be influential or not and whether we should process the layer for modeling. Some other datasets had the issue of missing data and we did not have the ability to spend time exploring this further with the municipal departments. This project further demonstrates the value of statistical methods for fire prediction as well as structuring municipal data for easier access by other departments in the city. Municipalities implementing a fire prediction model should spend more time on extracting and populating datasets.

Both Departments discussed the issue of abandoned homes and how they are particularly prone to fire. As this project concerned inspectable properties, this data was not included; however, this would be valuable in a fire prediction model applied to all buildings (such as smoke alarm targeting [10]).

Table 1 Main Datasets

	Vancouver		**New Westminster**		
	Records	*Fields*	*Records*	*Fields*	*Description*
Assessment data	206,480	28	9,212	10	Assessment value data for both land and building. Year of construction also included.
Parcel data	100,846	4	9,210	2	GIS based data
Fire inspection data	127,472	9	10,164	26	Vancouver data provided each violation as a record. New Westminster data provided each inspection as a record with the total violations for the address as a field. New Westminster data also provided all incidents prior to an inspection as a field.
Incident data	47,589	15	768	75	New Westminster provided solely fire incident data (with all other incidents kept as a single field in the inspection data). Vancouver provided "fire" and "other alarm" data but no medical/rescue call data.
Census Dissemination[2] Area data	993	47	90	47	Municipal census data at a dissemination area level. Area sizes vary. Includes data related to age, income, marital status, etc.
Building footprint data	154,124	51	8,973	12	Building footprint GIS data.

NWFRS also provided aggregated building level data which included the number of stories above and below ground, occupancy/use data, and whether the building had a sprinkler system. Using this data we were able to include a "Required Fire Flow" calculation for each building [11]. We also included any relevant data from Vancouver and New Westminster open data portals which included care homes, non-market and rental housing, zoning and land use, address and crime data. The main datasets used are listed in Table 1. Data was provided for the years 2012/2013-2017.

Finally, Opta data was also used in the model. Opta is one of the largest property/business level data aggregators in Canada.

Data Cleaning and Joining

After reviewing the raw datasets provided by both municipalities, we had to choose an aggregation level, i.e. building level, address level, or parcel level. In both municipalities multiple addresses are possible in one parcel. Furthermore, in the City of Vancouver it is possible to have multiple parcels connected to a single address. Fire inspection data is reported at the address level and in some cases at the unit number level. Fire incident data is reported at the address level. We used GIS tools to join all datasets at the parcel ID level. The parcel level ID for the City of Vancouver was the LAND_COORDINATE; the parcel level ID for the City of New Westminster was the NwID. In cases where neighbouring parcels have the same ID the parcel area value is the sum of both parcels. The building footprint field is the sum of all footprints in the parcel. Also, the number of units/addresses in the parcel is summed and included as a field. Land value, building value and tax levies were also summed at the parcel level.

[2] A dissemination area (DA) is a small, relatively stable geographic unit composed of one or more adjacent dissemination blocks with an average population of 400 to 700 persons based on data from the previous Census of Population Program. It is the smallest standard geographic area for which all census data are disseminated. DAs cover all the territory of Canada.

We used municipal address data to build a geo-coding service. A complete set of all addresses for the City of Vancouver was not readily available. Addresses in both the fire incident and fire inspection data required a significant amount of cleaning before using GIS tools to join the parcel ID to the incident/inspection record. In both municipalities there were addresses in the incident/inspection data that could not be joined or did not exist in the municipal address dataset. We expected this as it is common to have consistency issues in municipal departmental address datasets. Unmatched address records were not used in this project. For any municipality completing a future prediction model it would be recommended to spend more time resolving address issues. "Databridge" concepts provided by MODA would also be recommended [1] for better municipal data infrastructure.

Other open GIS data layers were included for both municipalities. Finally, population and other census data at the dissemination area level were also included. Crime data for the City of New Westminster was not readily available as a dataset. Crime data for the City of Vancouver was available as a GIS layer and mapped at the municipal block level. This was aggregated at the dissemination area level and included for the City of Vancouver.

311 data could not readily be processed and was not included. Again, any future prediction modeling work should consider this data, especially as it relates to sanitation complaints. Opta business level data was also included in the datasets. Feature engineering was also completed. The results of this project informed on our features and led to suggestions for future work.

Prediction Model Results

The performance of the model is initially described using three metrics (further metrics are provided later in this report which addresses model performance considering class imbalance):

- **True Positive Rate (TPR)** – the ratio of correctly identified fire incidents divided by the total of all fire incidents
- **False Positive Rate (FPR)** – the ratio of falsely predicted fires divided by the total of all non-fire incidents
- **Area Under Curve (AUC)** – the area under the ROC curve (Receiver Operating Characteristic) where the ROC curve is a plot of the TPR vs. FPR.

Plotting the FPR versus the TPR gives the ROC curve. The area under the curve (AUC) is a measure of how well the model performs.

The City of Vancouver model used data for the period 2013-2017. The data was initially split into a training dataset (2014-2016) and a test dataset (2017). 2013 data was used for feature selection. The target was to predict fire incidents (IncidentType = FIRE - Fire Department Responded Reportable – 1000) for 2017 at the parcel level (LAND_COORDINATE). Model parameters were tuned, and features selected by splitting the training set and using the 2016 fire incidents for validations. The final model was generated using the best parameters and the full training set. Finally, the model was tested against the 2017 fire incident data. The City of New Westminster model used data for the period 2013-2017 (incident data back to 2009 was used). The target was to predict fire incidents for 2017 at the parcel level (NwID). Again, data was initially split into a training dataset (2014-2016) and a test dataset (2017).

"…**70%** of fires were predicted for 2017…"

Visual representations of the models are shown in figure 2. The XGBoost algorithm [12] was used in both cases. The City of Vancouver model showed an AUC of 0.78. Choosing a point on the ROC curve, 70% of fires were accurately predicted for 2017 with an FPR of 28.8%. The City of New Westminster model showed a better AUC of 0.83. Choosing the same point on the ROC curve, 70% of fires were accurately predicted for 2017 with an FPR of 23%. Comparably, the Firebird project in the City of Atlanta showed a similar AUC value and predicted 71.36% of fires with an FPR of 20% (Atlanta used a subset of inspectable properties and not all inspectable properties). Even though approximately 25% of high-risk classified buildings for 2017 did not have a fire, the model still classifies these as risky buildings and should therefore be prioritized for an inspection.

Feature importance for each City is also shown in figure 2. Influential features for the City of Vancouver include land value, building footprint area, previous incidents and inspections, employment income rate, and crime rate. Influential features for the City of New Westminster include building area, predicted loss cost (Opta derived business score), employment income rate, demographic/Statscan data and Required Fire Flow [11] for the building.

Figure 2 Prediction Model and Results

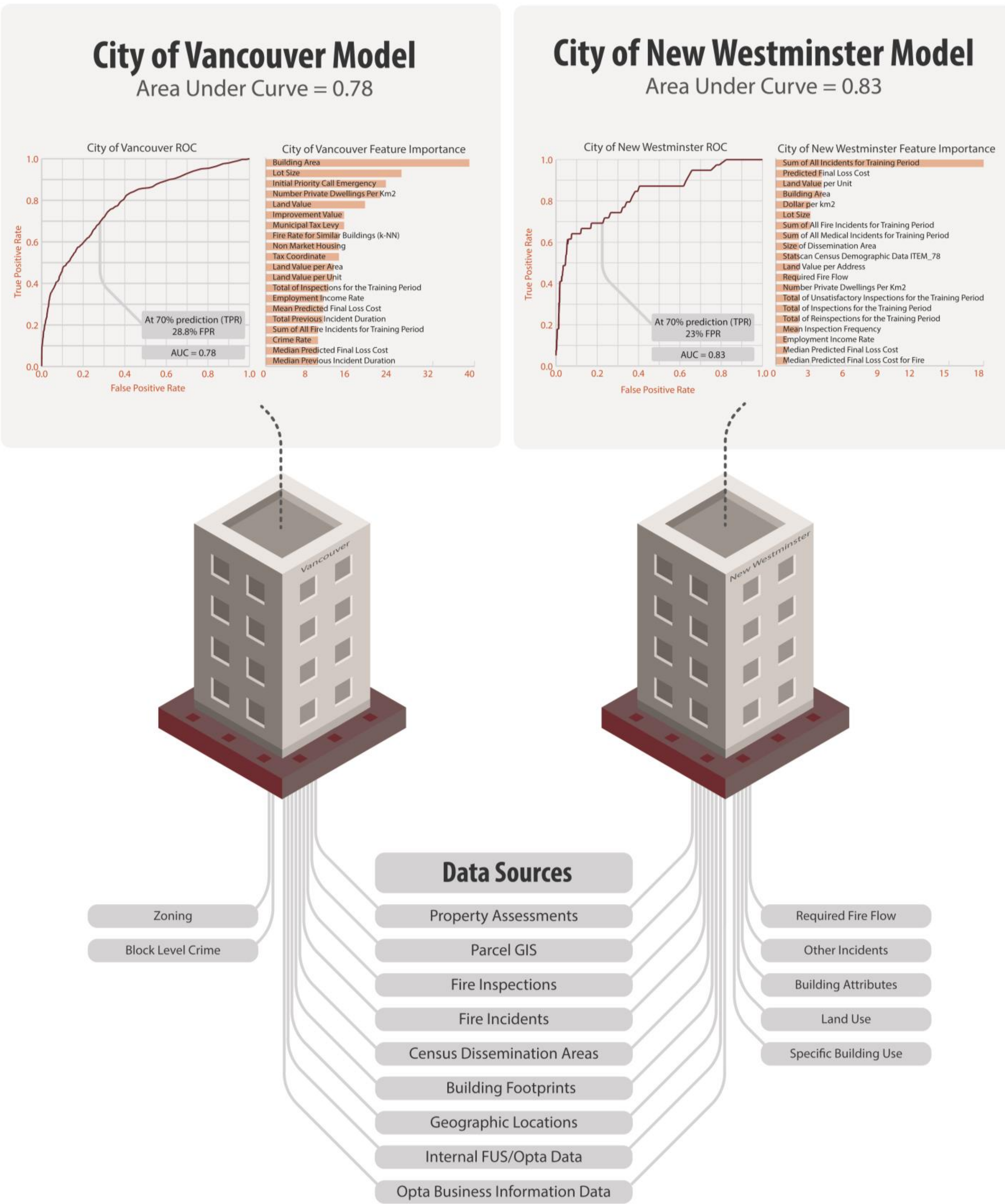
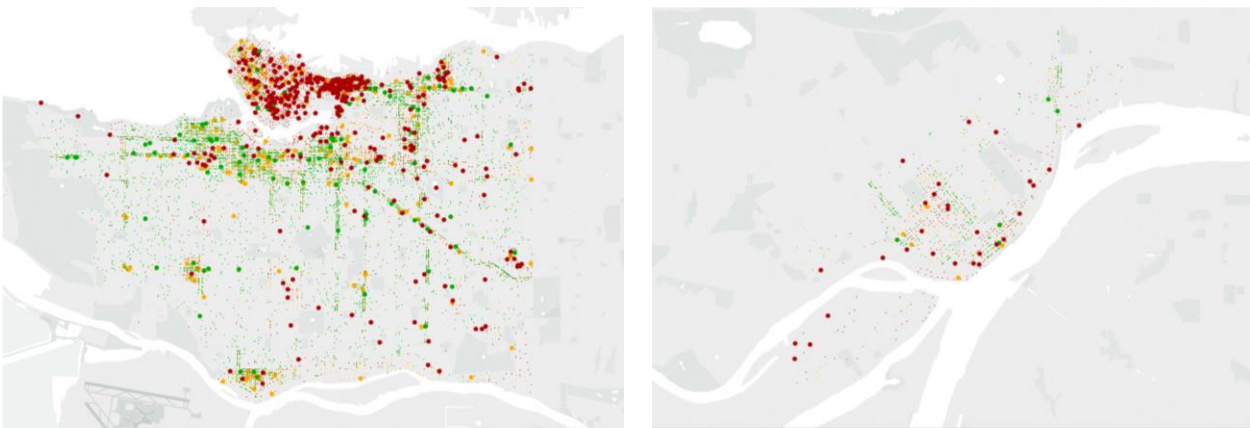


Figure 3 shows maps of the predictions. Inspectable properties are grouped into high (red), medium (orange) and low (green) risks based on the prediction scores. Large circles indicate the actual fires in inspectable properties in 2017 and small circles indicate inspectable properties that did not have fire calls in 2017. Large red circles show where the model predicted a fire in the property in 2017 and where the property had a fire in 2017 (these are the true positives). The City of Vancouver has a grouping of fires and predictions in certain areas. In the City of New Westminster, the actual fires are more dispersed. There are good predictions in both cases.

Figure 3 Fires and Predictions in Inspectable Properties for 2017

City of Vancouver **City of New Westminster**



Map Legend
- High Risk
- Medium Risk
- Low Risk
- Large = fire in 2017
- Small = no fire in 2017

Model Evaluation

(since the initial publication of this report in December 2018, we have elaborated on the methodology and evaluation of the model to provide additional metrics to AUC)

The final product of the model is a “risk score” built by uniformly binning the data on the output probability of the logistic regression for binary classification. As such, we are concerned more with overall performance as opposed to the ability to predict the correct probability, given that a relative probability amongst buildings is sufficient to construct the ranking. Accordingly, we follow the XGBoost recommendation for modelling on an imbalanced dataset with a binary logistic objective, and with a focus on overall performance[3]. Specifically, we employ XGBoost’s built-in *scale_pos_weight* parameter to balance the positive and negative weights in the training set and use Area Under Curve (AUC) as an evaluation metric. The *scale_pos_weight* is derived form the training set as follows:

$$\sum \text{negative instances} / \sum \text{positive instances}$$

In Vancouver, this means *scale_pos_weight* = 18.6, and in New Westminster *scale_pos_weight* = 16.8 (where IncidentType = FIRE - Fire Department Responded Reportable – 1000 is used as a target for high risk buildings; termed “fire”).

For an evaluation of the model performance beyond AUC, the following metrics are calculated:

- **Kappa statistic**
- **Precision**
- **Recall**

In our case, precision is the fraction of actual fires among the set of buildings predicted to have a fire, whereas recall is the fraction of relevant buildings predicted to have a fire over the total number of relevant buildings. The Kappa statistic measures the consistency between actual and

[3] Notes on Parameter Tuning, Handle Imbalanced Dataset, https://xgboost.readthedocs.io/en/latest/tutorials/param_tuning.html

predicted values, taking into account the agreement occurring by chance. Herein, precision and recall values are measured at the probability threshold that maximizes Kappa.

A summary of the performance metrics for both municipalities is given in Table 2. Vancouver achieves a maximum Kappa of 0.3 with a corresponding recall of 0.35, a precision of 0.34, and an AUC of 0.78. New Westminster achieves a Kappa of 0.42 with a corresponding recall of 0.41, a precision of 0.47, and an AUC of 0.83. Comparing Kappa shows that New Westminster's predictions are slightly more significant than those of Vancouver.

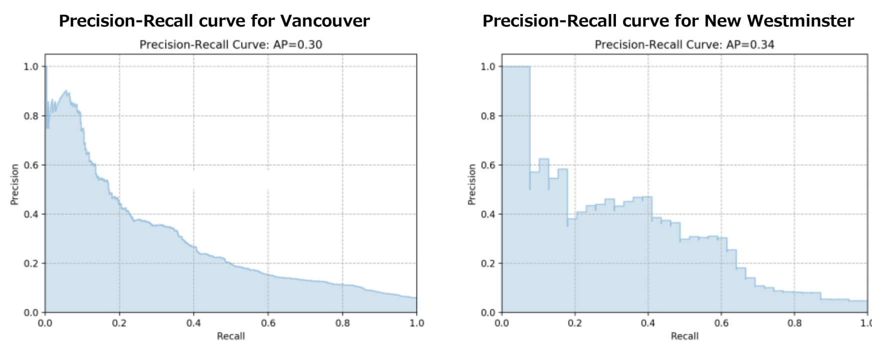
Table 2 Performance Metrics Summary

Municipality	Kappa	Recall	Precision	AUC
Vancouver	0.3	0.35	0.34	0.78
New Westminster	0.42	0.41	0.47	0.83

"…Vancouver experiences a **lift of 12.2** and New Westminster a **lift of 27**."

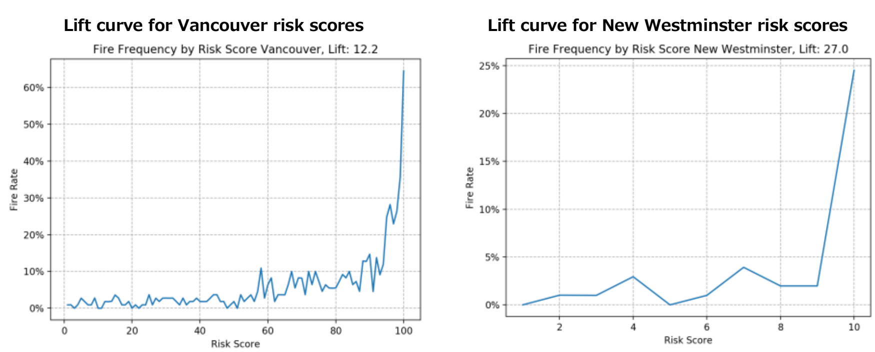
A Precision-Recall curve (PR-curve) is another effective way of evaluating model performance especially in the context of the business use-case. For a given value of recall, we can determine the precision of our model. This will give us insight into how many buildings we might need to inspect in order to achieve the desired coverage of relevant buildings. Figure 4 shows the PR-curves for Vancouver and New Westminster. In Vancouver's case, precision drops relatively smoothly from a value of 0.8 beyond a recall of 0.1. At a recall of 0.5, precision crosses below 0.2. Comparatively, New Westminster's PR-curve has a more dynamic profile. Precision hovers around 0.45 between recall of 0.2 and 0.4, drops to an average of 0.35 between recall of 0.4 and 0.6, and drops again for recall greater than 0.65.

Figure 4 Precision-Recall curves



Finally, we perform a lift analysis on the predicted risk scores. A lift curve can help measure the effectiveness of a predictive model and is calculated by taking the number of relevant events predicted at a given risk score over the total number of events assigned to that score. Figure 5 shows the lift curves for Vancouver and New Westminster.

Figure 5 Lift curves for risk scores



In both cases the frequency of fire events increases significantly in the top quintile of risk scores. The lift here is defined as the fraction of average fire frequency in the top quintile of scores over that in the bottom quintile. The result is that Vancouver experiences a lift of 12.2 and New Westminster a lift of 27.

Both the Vancouver and New Westminster models achieve non-trivial values of Kappa which confidently rules out the possibility that relevant predictions are due to chance. Furthermore, risk scores produced by these models are effective in assessing the risk of fires, as is evident by the ratio between fire rates in the upper and lower quintiles of the risk scores.

Proof of Concept and Implications

Proof of Concept

The City of New York, City of Atlanta, and City of Pittsburgh have completed fire prediction/building risk assessment work using machine learning techniques. Each of these projects showed strong fire predictions that were better than current risk-based models. This project further validates the concepts.

We wanted to do a comparison test in order to better show proof-of-concept and see how the prediction rates compared in two municipalities. We also wanted to see if the work would be beneficial to both a relatively large municipality as well as a relatively small municipality.

Both the City of Vancouver, with a population of 631,486 (2016), and the City of New Westminster, with a population of 70,996 (2016) partnered with us on the project. In both Cities we see very strong prediction rates of 70% of fires predicted (TPR) with an acceptable FPR of approximately 25%. The AUC was 0.78-0.83. Currently VFRS aims to inspect all buildings on an annual basis and NWFRS aims to inspect all buildings on a two-year frequency. The results of this project show that municipal datasets and statistical learning methods can be used to better target high risk buildings for an inspection. Furthermore, these models allow for a fast categorization of building fire risk with currently available data. The feature weight is learnt from actual fire incidents as opposed to a manual categorization using user-biased weights for influential features. Fire department knowledge and experience can be used to further strengthen the outcomes. The model can also be easily updated as soon as new data is available with limited user intervention (in the City of Pittsburgh the model updates weekly). These methods also allow for a municipal-specific fire prediction risk model. This report further validates previous fire prediction work and shows the value of using these methods.

The following comment is from the fire prediction work completed by MODA:

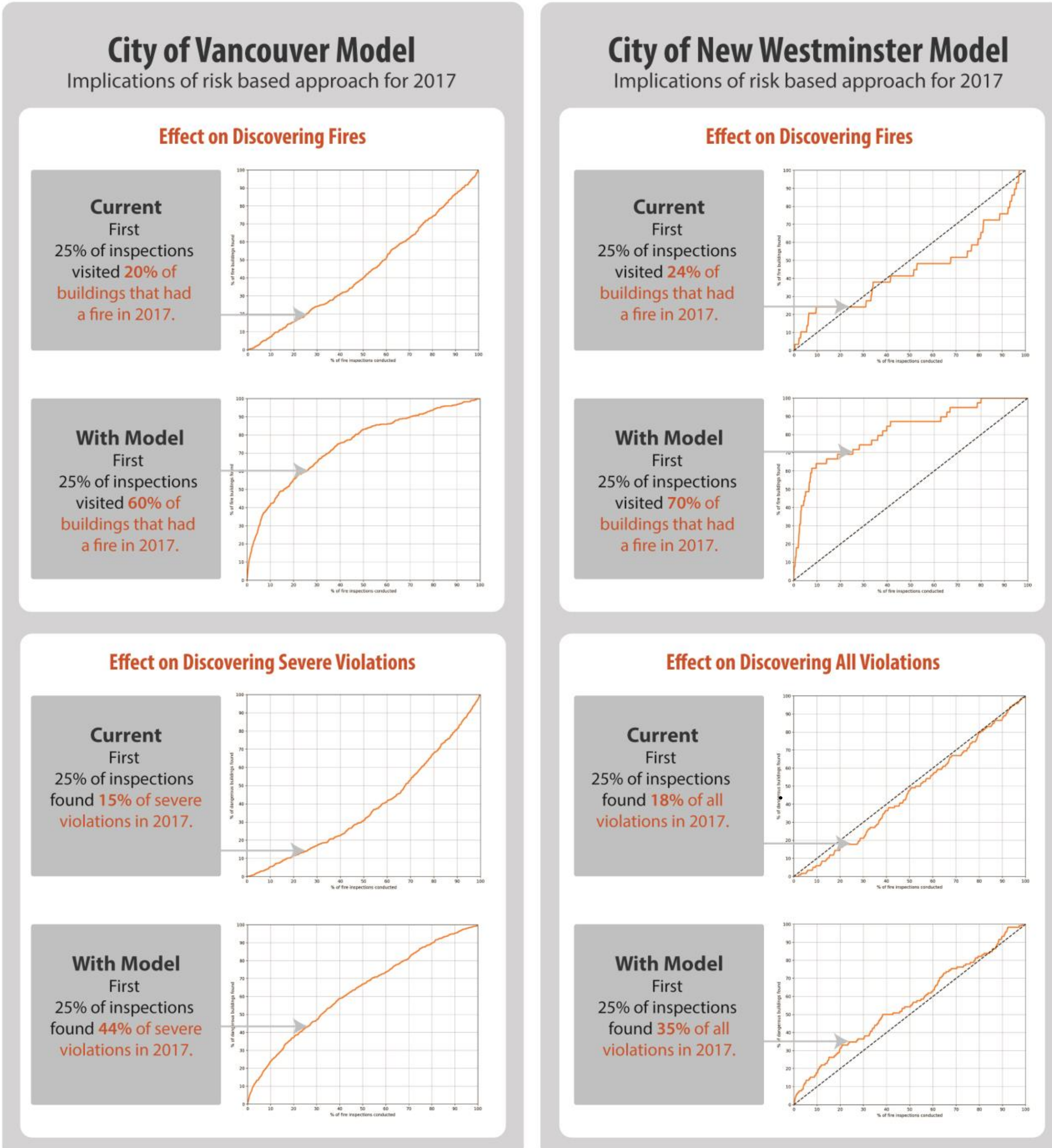
> *"Each morning, fire department building inspectors would still receive a list of properties to investigate that day. The only difference was that now that list was pre-prioritised to focus on the most dangerous buildings first. As a result, the work of MODA won the support of the Fire Department as it led to maximum improvement in their service with almost no disruption to day-to-day business. The front-line staff liked it because it helped them do their jobs even better than before. The City and New Yorkers liked it because it saved lives and made them feel safer."* [1]

Implications for VFRS and NWFRS

Using the model results we assign a Risk Score to each building in the Cities as shown in figure 5. It can be seen that a large proportion of the 2017 fires are grouped in the higher Risk Scores (>90) with the Vancouver model showing a lift of 12.2 and the New Westminster model showing a lift of 27.

Figure 6 shows the implications of using the fire risk prediction scores on the regular inspection programs for both VFRS and NWFRS.

Figure 6 Implications of a Predictive Risk Based Inspection Program



First, we consider the effects on visiting buildings that had a fire incident in 2017. We looked at the first 25% of inspections for the year as it would be preferable to visit higher risk buildings first. Figure 6 shows that under the current VFRS inspection program the first 25% of inspections had visited 20% of buildings that had a fire in 2017. By using a risk-based approach and re-prioritizing inspections, the first 25% of inspections would theoretically visit 60% of inspectable buildings that had a fire incident in 2017. As the model was trained on fire events, we expect an increase in this value; however, the model was not trained to predict severe violations so looking at the improvement in this number further validates the risk-based approach (VFRS provided a list of severe violations which is a subset of all violations). Figure 6 shows an increase in discovering severe violations from 15% to 44% in the first 25% of inspections. This result also shows a good correlation between fire events and severe violations.

For the first 25% of inspections completed in the City of Vancouver, **the percentage of buildings visited that had a fire increases from approximately 20% to 60%** with a risk-based model. The **percentage of severe violations discovered increases from 15% to 44%** in the first 25% of inspections.

Under the current NWFRS inspection program figure 6 shows that the first 25% of inspections had visited 24% of buildings that had a fire in 2017. By assigning risk scores to properties and re-prioritizing inspections the first 25% of inspections would theoretically visit 70% of inspectable buildings that had a fire incident in 2017. A subset list of severe violations was not available for the New Westminster data; however, looking at all violations the increase is from 18% to 35%. We would expect that looking solely at severe violations we would see a greater increase and again a better correlation between severe violations and fires.

Overall, implementing a fire risk prediction/assessment approach shows that, for the year 2017, more violations (and severe violations) would be identified earlier in the inspection year. This is similar to the impact seen by FDNY using their Risk Based Inspection System [1].

Integrating into Operations and Workflow

The following gives a brief overview of the steps to building and integrating a risk-based approach to inspections as covered in this report:

1. Hold working group meetings with fire department and prevention personnel to discuss municipal specific fire issues based on staff experience. Related datasets from both fire department and other municipal departments can be discussed as well.
2. Review the available datasets and how these datasets are populated. It is important to understand any bias in how the data is collected as well as data quality issues. For other municipal departments it is important to understand how well populated the dataset is and the overall intent of populating the data for that specific department (e.g. an address record may be strictly captured in some cases and loosely captured in other cases). It is also important to understand how readily available this data is and how easily it can be extracted.
3. Establish an aggregation level for all the data. In the case of this report we used the parcel level.
4. Use GIS tools to standardize the joining of the data and ensure updated data is easily integrated.
5. Define target, build model, construct features and establish dependent features as well as model validation level. Build tools/pipelines for easy updates.
6. Assign risk scores to inspectable properties.
7. Establish model update interval considering how often new data is available. Re-assign risk scores to properties at established intervals (e.g. every 3 months).

Conclusions and Further Work

"…A smart city is an urban area that uses different types of electronic data collection sensors to supply information which is used to manage assets and resources efficiently…"

This report begins by discussing innovative solutions to municipal work using a Smart Cities type solution. A Smart City can be defined as follows:

> *"A smart city is an urban area that uses different types of electronic data collection sensors to supply information which is used to manage assets and resources efficiently. This includes data collected from citizens, devices, and assets that is processed and analyzed to monitor and manage traffic and transportation systems, power plants, water supply networks, waste management, law enforcement, information systems, schools, libraries, hospitals, and other community services"* [13]

The recent "Smart Cities Challenge" launched by Infrastructure Canada [14], with a top prize of $50 million, demonstrates the importance of cities adopting smart city approaches. The Challenge encourages communities to adopt a smart cities approach to improve the lives of their residents through innovation, data and connected technology. Similar to the work completed by FDNY [1], Atlanta Fire Department [2], Ash Center at Harvard Kennedy School [15], and Pittsburgh Bureau of Fire [3], this project demonstrates a smart cities approach to improving city fire services through data.

The following is a quote from the founding director of the New York City Mayor's Office of Data Analytics:

> *"In 2009 I was given the straightforward mission by New York City Mayor Michael Bloomberg; use data to improve government services to the 8.5 million New Yorkers. His follow up guidance was to do it inexpensively, with minimal staff, and to make it impactful and sustainable."* [1]

This led to the development of FireCast and FDNY's Risk Based Inspection Program in 2013 which led to improvements in the targeting of fire inspections and continues to be used today as part of day to day operations.

A partnership between Georgia Tech and the Atlanta Fire Department in 2016 led to the development of a similar model called Firebird which further validated a fire prediction risk assessment approach to prioritizing fire inspections.

In 2018 a "Metro21: Smart Cities Initiative" partnership between Carnegie Mellon University and Pittsburgh Bureau of Fire showed a similar validation of methodology and led to a model that targets fire inspections and updates weekly.

This report further validates building fire prediction methods in a comparison study of two municipalities/fire departments in Canada. We would recommend this approach to any sizable municipality looking to create a risk-based approach and drive insights from municipal data in order to better target inspections. We demonstrate in this report how machine learning can contribute to improving the provision of public fire services using existing municipal data and data exchange processes.

We are in a position to aid fire departments and municipalities to create these insights and models and look forward to future partnerships on fire prediction and other related data insight work.

References

[1] E. Copeland, "Big Data in the Big Apple," Capital City Foundation, 2015.

[2] Statistics Canada, "Table 35-10-0195-01 Fire-related deaths and persons injured, by type of structure," [Online]. Available: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3510019201.

[3] Minister of Transportation and Infrastructure, "Bill 4 - 2016: Fire Safety Act - Part 6 - Compliance Monitoring," [Online]. Available: https://www.leg.bc.ca/parliamentary-business/legislation-debates-proceedings/40th-parliament/5th-session/bills/third-reading/gov04-3.

[4] Ministry of Community Safety and Correctional Services, "Fire Protection and Prevention Act, 1997 - Proposal," [Online]. Available: https://www.ontariocanada.com/registry/view.do?postingId=26546&language=en.

[5] Apolitical, "New York City is saving lives by predicting where fires will break out," [Online]. Available: https://apolitical.co/solution_article/new-york-city-saving-lives-predicting-fires-will-break/.

[6] M. Madaio, S.-T. Chen, O. L. Haimson, W. Zhang, X. Cheng, M. Hings-Aldrich, D. H. Chau and B. Dilkina, "Firebird: Predicting Fire Risk and Prioritizing Fire Inspections in Atlanta," 2016.

[7] J. Roman, "Embracing Analytics: How communities around the country are utilizing data analytics to predict a host of risk factors and reduce fires.," *NFPA Journal,* no. May June 2016, 2016.

[8] Carnegie Mellon University, "Metro21: Smart Cities Institute," [Online]. Available: https://www.cmu.edu/metro21/.

[9] L. Garis and J. Clare, "A Dynamic Risk-Based Framework for Redesigning the Scheduling of FIre Safety Inspections," University of the Fraser Valley Center for Public Safety and Criminal Justice, 2014.

[10] SAS, "Predicting fire risk and saving lives," [Online]. Available: https://www.sas.com/en_gb/customers/london-fire-brigade.html.

[11] Fire Underwriters Survey, "Fire Underwriters Survey - technical documents," [Online]. Available: http://www.fireunderwriters.ca/docs/FUS_Water%20Supply%20for%20Public%20Fire%20Protection_v2006.pdf.

[12] XGBoost developers, "XGBoost," [Online]. Available: https://xgboost.readthedocs.io/en/latest/.

[13] Wikipedia, "Smart City," [Online]. Available: https://en.wikipedia.org/wiki/Smart_city.

[14] Infrastructure Canada, "Smart Cities Challenge," [Online]. Available: https://www.infrastructure.gc.ca/cities-villes/index-eng.html.

[15] J. Jay, "Can Algorithms Predict House Fires?," Ash Center at Harvard Kennedy School, [Online]. Available: https://datasmart.ash.harvard.edu/news/article/can-algorithms-predict-house-fires-990.

A **Fire Underwriters Survey**/**Opta Information Intelligence** project
www.fireunderwriters.ca/**www.optaintel.ca**