School of
Continuing Studies

YCBS 299 – Data Science Capstone Project

Business Problem Framing Report

Project A:

Predicting High Fire Risk Areas by Month in the City of Montreal

This report aims to define the business problem, explain its relevance, and identify solutions from literature

Team #2

Stéphane Boileau

Kawtar El Bouzidi

Ralph Bou-Fadel

Reem Hoteit

Juliano Pereira

September 17, 2025

Table of Contents

Business Problem

The Montreal Fire Department (SIM) lacks predictive capabilities to anticipate where fires will occur across its 332,000-building territory, forcing the department into a reactive operational stance, responding to incidents only after they happen.

Current Operational Reality

The SIM protects 2 million residents across 500 square kilometers with 67 fire stations and 2,700 staff. Despite recording approximately 1,800 fires annually, the department cannot predict which buildings or neighborhoods face elevated risk. With a capacity for only 7,500 annual inspections (2.3% of buildings), complete coverage would take 44 years.

Inspection Selection Without Data

Building inspections are currently selected by individual fire station chiefs using only personal knowledge of their territories, without access to city-wide data analysis, standardized risk metrics, or systematic prioritization methods. This decentralized approach means:

- No consistent risk assessment methodology across stations
- Twenty years of incident data remains unused for prediction
- Undocumented selection criteria varying by station
- Temporal and seasonal patterns going unconsidered
- Emerging urban development risks remaining undetected
- Neighborhood-level fire patterns across Montreal Island unidentified

Resource Allocation Blindness

Without predictive capabilities, the department cannot position resources proactively, adjust staffing based on predicted risk, schedule inspections during peak periods, or identify deteriorating safety conditions. The department maintains uniform resource distribution regardless of varying risk levels. While achieving approximately 5-minute response times, stations cannot anticipate incident locations for optimal pre-positioning.

Data Collection Without Intelligence

Since 2000, the SIM has collected comprehensive fire incident records, building information, response metrics, and inspection histories. However, this data exists without analytical integration. Station chiefs cannot practically analyze two decades of city-wide data to inform their choices.

Without predictive modeling, Montreal's fire prevention strategy remains reactive, waiting for fires instead of using decades of data to predict them. Both buildings and neighborhoods across the island remain at risk when prevention could be targeted more effectively through data-driven prediction.

Business Problem Rationale

Fires are extremely harmful. A study published in the December 2019 issue of the British Medical Journal estimated nine million new injuries and 121 000 deaths due to fire, heat and hot substances in 20172]. Furthermore, the World Health Organization estimates that around 180 000 deaths annually are caused by burns [3]. Lastly, fire and burns cause more death and disability than any other natural hazard, including floods and earthquake.

The impacts of urban fires go far beyond immediate casualties. They also affect physical and mental health and cause economic and environmental losses. Regardless of the immediate burns and smoke inhalation injuries, exposition to fires increase the risk of cardiovascular problems like heart attacks and strokes at the long-term. People displacement from their homes also lead to leaving the community, thus causing trouble to fulfill jobs[5]. This is without discussing the financial damages a fire can cause as well as the exorbitant cost of reconstruction, which can be quite severe and lead to abandonment of the place. On the environmental side, pollutants generated from home and garage fires contaminate soil and water systems.

Recent events have also shown how quickly fire risks can change with social behaviors. During the COVID-19 pandemic, Canada experienced a reversal of the long-term decline in fire incidents (TheDaily, 2023). Fire-related deaths rose sharply from 148 in 2019 to 199 in 2020, with the majority occurring in residences. Cooking and smoking materials were the leading causes, and seniors were particularly vulnerable. Researchers attribute this increase to more time spent at home during lockdowns, highlighting how lifestyle changes can significantly increase fire risk.

In this context, predicting the likelihood and location of fires can significantly reduce their negative impacts. Furthermore, this ability becomes especially relevant in a dense city such as Montreal, where the risk of urban fires is only going to increase in the future as a lot of people are moving to wildland- urban interface, i.e. areas with flammable vegetation, and staying at home more. Subsequently, the Montreal fire department could use the model provided for a more rapid response and adequate resource allocation as well as an efficient prevention, and deployment.

For example, a machine learning model developed to study firefighting operation time in urban areas found that the number of personnel, rescue equipment, fire engines, and extinguishers were key predictors of shorter response times[6]. In practice, skilled manpower and adequate resource allocation can reduce fire losses significantly. Accordingly, predicting where and when incidents are most likely to occur makes such deployment more efficient.

Moreover, an autoregressive moving average (ARMA) model was also created by Chinese scientists to explore the suppressive effect of six fire prevention campaigns initiated after disastrous fires that resulted in very high numbers of casualties and significant losses[8]. This model utilized monthly fire statistics of 10 cities from January 1997 to December 2001 in the Jiangsu province. These results indicated that fire disasters and serious fires produce suppressive effects on monthly fires for three months after the serious incidents in eight cities and two months after for two cities. These findings suggest that timely prevention actions, guided by predictive analytics, could reduce fire risks and losses.

State-of-the-art Review

Research shows that fire risk prediction has evolved significantly from traditional risk indices and expert-based assessments to advanced data-driven approaches. Early methods often relied on limited building characteristics and historical fire records, which restricted accuracy and adaptability. Recent research demonstrates the power of machine learning and probabilistic modeling in overcoming these limitations. A recent study conducted by Ahn et al. (2024) developed a stacking ensemble that combined 16 machine learning algorithms incorporating 34 variables, including building attributes, land characteristics, and demographic variables (Ahn et al., 2024). This model successfully identified over half of all actual fires within the top 22% of predicted high-risk buildings. Similarly, Huang et al. (2024) advanced urban fire risk prediction by integrating fire incident, meteorological, and Point of Interest (POI) data into a grid-based model (1 km × 1 km), achieving 98.9% accuracy using a soft voting ensemble of deep neural networks, Random Forest, and XGBoost (Huang et al., 2024). These ensemble approaches demonstrate that combining models produces more balanced predictions, capturing both high-risk areas and reducing false alarms. Beyond machine learning ensembles, probabilistic reasoning frameworks such as Bayesian networks (BNs) provide a complementary approach. Chen et al. (2023) demonstrated that Bayesian Networks can effectively assess fire risks in vulnerable buildings by integrating multiple data sources and expert opinions, providing accurate predictions even with limited data (Chen et al., 2023).

Other studies emphasize the critical role of environmental and weather data. Agarwal et al. (2020) merged U.S. fire incident and NOAA weather datasets, applying gradient boosting trees to predict fire losses with high accuracy (R^2 = 0.93–0.97), highlighting the influence of temperature, wind, and precipitation on fire severity (Agarwal et al., 2020). At the property level, Surya (2017) explored the use of artificial neural networks, boosting, and SVMs for fire prediction, with applications ranging from IoT sensor-based anomaly detection to drone-based image recognition for early fire detection (Surya, 2017).

These advances in monthly prediction and grid-based modeling directly address Montreal's need for temporal and spatial fire risk assessment. Atlanta's Firebird framework and Pittsburgh's Metro 21 build on this foundation, exemplifying how municipalities are now applying predictive models to assess fire risk.

The Firebird Framework is a predictive analytics system developed by the Atlanta Fire Rescue Department (AFRD) to improve fire prevention and optimize inspection prioritization. Leveraging machine learning, geocoding, and visualization, it generates fire risk scores for commercial properties and displays them on an interactive map. Predictive models built with Random Forest and Support Vector Machines achieved a 71.36% true positive rate and identified 6,096 at-risk commercial properties. By shifting fire departments from reactive response to proactive risk management, Firebird improves operational efficiency and strengthens safety

The Metro 21 model, developed by Pittsburgh's Bureau of Fire (PBF), prioritizes property fire inspections using historical fire incidents, property assessments, and violation records. Trained on 20,806 commercial properties, it generates weekly updated risk scores (1–10) to estimate fire likelihood within a 6-month window, with results visualized through the Burgh's Eye View platform. The model flagged 164 properties as high risk and 596 as medium risk, with performance evaluated using kappa (37%) and recall (55%).

Leveraging the Metro 21 and Firebird frameworks, Vancouver and New Westminster used XGBoost to model fire risk from municipal data (property, inspections, incidents). Results showed 70% true positive rate, with AUC of 0.78 (Vancouver) and 0.83 (New Westminster). Key drivers included property value, building area, inspection and fire history, crime, and demographics. Risk scoring helped boost severe violation discovery by 29% in Vancouver and overall violation detection by 17% in New Westminster.

References

Include references from literature review

Agarwal, P., et al. (2020). "Big data and predictive analytics in fire risk using weather data." Risk analysis 40(7): 1438-1449.

Ahn, S., et al. (2024). "Comprehensive building fire risk prediction using machine learning and stacking ensemble methods." Fire 7(10): 336.

Chen, J., et al. (2023). "A combined method to build Bayesian network for fire risk assessment of historical buildings." Fire Technology 59(6): 3525-3563.

Frueh, Sarah, The growing challenge of urban wildfires, https://www.nationalacademies.org/news/2025/03/the-growing-challenge-of-urban-wildfires, March 10,2025

Fu, E. Y., et al. (2024). Building Fire Hazard Predictions Using Machine Learning. Intelligent Building Fire Safety and Smart Firefighting, Springer: 387-414.

Fu, E. Y., et al. (2024). Building Fire Hazard Predictions Using Machine Learning. Intelligent Building Fire Safety and Smart Firefighting, Springer: 387-414.

Huang, S., et al. (2024). "Development and validation of a soft voting-based model for urban fire risk prediction." International Journal of Disaster Risk Reduction 101: 104224.

Ku, C.-Y. and C.-Y. Liu (2024). "Predictive modeling of fire incidence using deep neural networks." Fire 7(4): 136.

Madaio, M., Chen, S., Haimson, O. L., Zhang, W., Cheng, X., Hinds-Aldrich, M., Chau, D. H., & Dilkina, B. (2016, August). Firebird: Predicting fire risk and prioritizing fire inspections in Atlanta. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 185–194). ACM. https://doi.org/10.1145/2939672.2939682

Madaio, M. A. (2018). Predictive modeling of building fire risk: Designing and evaluating predictive models of fire risk to prioritize property fire inspections. Metro21: Smart Cities Initiative, Carnegie Mellon University. https://github.com/CityofPittsburgh/fire_risk_analysis

McGuinness, R., Gilbert, G., Whittamore, Z., Zhong, L., Van Zyl, R., & Lee, J. (2019, April). A building fire risk prediction validation project. Fire Underwriters Survey/Opta Information Intelligence, in partnership with Vancouver Fire Rescue Service and New Westminster Fire Rescue Service. https://www.firesurvey.ca

Sahebi, Ali, Sayyadi Hojjat, Havasy Bahram and Veisani Yousef, Predicting firefighting operation time in urban areas using machine learning: identifying key determinants for improved emergency response, https://link.springer.com/article/10.1007/s42452-025-06703-0

Singh Walia, B., et al. (2018). A dynamic pipeline for spatio-temporal fire risk prediction. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.

Song, Lu, Guohui Li, Peng Mei and Zhang Heping, Suppressive effects of fire prevention campaign in China: A time series analysis, March 10th, 2017, https://www.sciencedirect.com/science/article/abs/pii/S0925753516000394?via%3Dihub

Surya, L. (2017). "Risk analysis model that uses machine learning to predict the likelihood of a fire occurring at a given property." International Journal of Creative Research Thoughts (IJCRT), ISSN: 2320-2882.

TheDaily (2023). "Fire incidents increase during the pandemic." Retrieved Sept 14, 2025, from https://www150.statcan.gc.ca/n1/daily-quotidien/230608/dq230608a-eng.htm.

World Health Organization Bulletin, The impact of urban fires on public health, https://pmc.ncbi.nlm.nih.gov/articles/PMC12057281/, May 1 ,2025